Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 23 (2021), Issue 9 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Viewpoints

Original Papers

Corrigenda and Addendas

Review

The Effect of Telehealth on Hospital Services Use: Systematic Review and Meta-analysis

Guido M Peters[1,2], MSc; Laura Kooij[2,3,4], MSc; Anke Lenferink[2], PhD; Wim H van Harten[2,4,5], MD, Prof Dr; Carine J M Doggen[1,2], Prof Dr

[1]Department of Clinical Research, Rijnstate Hospital, Arnhem, Netherlands

[2]Department of Health Technology and Services Research, Technical Medical Centre, University of Twente, Enschede, Netherlands

[3]Department of Information and Medical Technology, Rijnstate Hospital, Arnhem, Netherlands

[4]Division of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Amsterdam, Netherlands

[5]Rijnstate Hospital, Arnhem, Netherlands

Corresponding Author:
Carine J M Doggen, Prof Dr
Department of Health Technology and Services Research
Technical Medical Centre
University of Twente
Drienerlolaan 5
Enschede, 7522NB
Netherlands
Phone: 31 534897475
Email: c.j.m.doggen@utwente.nl

Abstract

Background: Telehealth interventions, that is, health care provided over a distance using information and communication technology, are suggested as a solution to rising health care costs by reducing hospital service use. However, the extent to which this is possible is unclear.

Objective: The aim of this study is to evaluate the effect of telehealth on the use of hospital services, that is, (duration of) hospitalizations, and to compare the effects between telehealth types and health conditions.

Methods: We searched PubMed, Scopus, and the Cochrane Library from inception until April 2019. Peer-reviewed randomized controlled trials (RCTs) reporting the effect of telehealth interventions on hospital service use compared with usual care were included. Risk of bias was assessed using the Cochrane Risk of Bias 2 tool and quality of evidence according to the Grading of Recommendations Assessment, Development and Evaluation guidelines.

Results: We included 127 RCTs in the meta-analysis. Of these RCTs, 82.7% (105/127) had a low risk of bias or some concerns overall. High-quality evidence shows that telehealth reduces the risk of all-cause or condition-related hospitalization by 18 (95% CI 0-30) and 37 (95% CI 20-60) per 1000 patients, respectively. We found high-quality evidence that telehealth leads to reductions in the mean all-cause and condition-related hospitalizations, with 50 and 110 fewer hospitalizations per 1000 patients, respectively. Overall, the all-cause hospital days decreased by 1.07 (95% CI −1.76 to −0.39) days per patient. For hospitalized patients, the mean hospital stay for condition-related hospitalizations decreased by 0.89 (95% CI −1.42 to −0.36) days. The effects were similar between telehealth types and health conditions. A trend was observed for studies with longer follow-up periods yielding larger effects.

Conclusions: Small to moderate reductions in hospital service use can be achieved using telehealth. It should be noted that, despite the large number of included studies, uncertainties around the magnitude of effects remain, and not all effects are statistically significant.

(J Med Internet Res 2021;23(9):e25195) doi:10.2196/25195

KEYWORDS

telehealth; systematic review; meta-analysis; hospitalization; health services use; eHealth

Introduction

Many see the COVID-19 crisis as an opportunity to stimulate digital transformation. We can expect digital care and eHealth to receive a boost during this era. Creativity and flexibility are stimulated to formulate an answer to challenges in patients fearing infection in a hospital and to social distancing being necessary within hospital premises. Telehealth, defined as health care provided over a distance using information and communication technology (ICT) to enable interaction between patients and health professionals [1], may offer a solution. However, the efficacy of telehealth is unclear. When the dust has settled, there is a need to properly evaluate experiences and the evidence base underlying various forms of telehealth.

In addition, digital transformation is considered in response to the need to improve patient centeredness and concerns about growing health care expenditures [2,3]. Limiting the need for inpatient care, which is the main driver of hospital costs, may reduce health care expenditures [4,5]. Manufacturers' claims and commercial pilot reports seem to dominate the debate, and policy makers frequently embrace those claims. In the Netherlands, the government presumes that hospital care can return to a very low percentage of annual volume growth in view of the anticipated effects of digital transformation. However, the extent to which telehealth can reduce hospital service use remains unclear. Some reviews have reported on the effect of telehealth on this outcome, finding both reductions and increases in hospital service use [6-8]. A recent systematic overview of telehealth interventions found that the effect on all-cause hospitalizations ranged from a reduction of 13.8% to an increase of 4.7% [6]. No prior review has compared the effects between health conditions, and most have focused on a single telehealth type, limiting generalizability [6-8]. Firm evidence for economic benefits is also limited, as cost-effectiveness studies are sparse and show contradictory results [9,10]. Moreover, telehealth can be implemented in various ways. Telehealth interventions include (1) video consultation, (2) automated device-based monitoring, (3) web-based monitoring, (4) interactive voice response (IVR) systems, (5) mobile telemonitoring, and (6) structured telephone support (STS) [6].

We conducted a systematic literature review of randomized controlled trials (RCTs) aiming to provide an overview of the evidence for the effect of telehealth on hospital services use, that is, all-cause and condition-related hospitalizations, and their duration (per patient and per hospitalization). Furthermore, we evaluated the risk of bias in all studies, as well as the quality of evidence for all outcomes. Finally, we explored which types of telehealth are most effective and which patient groups are the optimal target for reducing hospital service use.

Methods

Overview

This review followed the guidelines of the Cochrane Handbook, with some modifications [11]. Notably, we used reporting of the outcomes of interest as an inclusion criterion, selected studies and extracted data partially in duplicate (20%), and deviated somewhat from the suggested algorithm to judge the risk of bias arising from the randomization process (Multimedia Appendix 1).

Data Sources and Searches

We searched MEDLINE, Scopus (Elsevier), and the Cochrane Central Register of Controlled Trials (Wiley) from inception up to April 2019. The search strategy (Multimedia Appendix 2) was developed by GMP using MeSH (Medical Subject Headings) terms and reference lists of relevant reviews until it encompassed all important keywords, and the search found all pertinent articles included in earlier reviews. WHVH and CJMD critically evaluated the search strategy before implementation.

Eligibility Criteria

RCTs and cluster RCTs reporting the use of telehealth interventions compared with usual care were included. Telehealth was defined as health care interventions provided over a distance using ICT to enable interactions between patients and health professionals or among health professionals. Patients of any age and with any health conditions were considered. Reported outcomes included at least one of the following: all-cause hospitalization, condition-related hospitalization, or length of hospital stay. We considered only published, English, full-text, and peer reviewed articles. We did not apply any restrictions to the setting or date of publication.

This review follows the taxonomy of telehealth interventions developed in another systematic review [8], which differentiates between video consultations, (automated) device-based monitoring, web-based telemonitoring, IVR, mobile telemonitoring, and STS.

Video consultations are defined as any intervention using synchronous, two-way, audio-visual communication between patients and health care providers to perform triage or provide health advice. If measurement devices were provided, measurements were communicated solely during the video consultations.

In device-based monitoring, patients are provided with devices to measure vital signs or to report symptoms essential for detecting changes in health status. Automated alerts triggering actions from health care providers, such as phone calls, are frequently included.

Web-based telemonitoring includes interventions using a web portal to enable patients to report vital signs and symptoms, and to enable health professionals to provide educational material and feedback.

In IVR systems, patients are required to enter vital signs and symptoms through their home or mobile telephone in response to automated questions. These systems are typically combined with automated alerts that trigger actions from health care providers.

With mobile telemonitoring, patients actively submit vital signs and symptoms through their personal mobile devices. Vital signs are measured using external measurement devices.

STS provides patients with a specified number of telephone contacts for a given period of time, during which patients report

their health status and receive health advice, medication adjustments, or referrals to health professionals.

We defined condition-related hospitalizations as hospitalizations due to the targeted health conditions. Studies that explicitly reported only condition-related outcomes are not aggregated with all-cause outcomes, as outcomes resulting from causes other than the condition of interest are unknown in that case, which could bias the results.

For the mean length of hospital stay, the total number of hospital days was divided by the total number of hospital stays. This is in contrast to the number of hospital days, where the total number of hospital days was divided by the total number of patients.

Data Collection and Extraction

GMP screened all titles and abstracts. This screening was independently verified on a sample basis (10%) by LK and AL. Screening of full text articles was performed identically. Disagreements were resolved through discussion, or adjudication by CJMD. Screening was performed using the Covidence systematic review software [12].

Using a standardized data extraction form, GMP extracted the following data from all included studies: study characteristics (eg, country and setting), population characteristics (eg, health condition, age, and gender), intervention details (eg, ICT components used and frequency of use), and outcomes (hospitalizations, length of hospital stay, and hospital days; Multimedia Appendix 3). Data extraction was verified by LK on a sample basis.

Assessment of Risk of Bias

We used the Cochrane Risk of Bias 2 (RoB 2) tool to assess the risk of bias for each study [13]. A number of rules were derived from the manual to ensure consistent judgments between reviewers (Multimedia Appendix 1). GMP assessed the risk of bias of all studies. Risk of bias assessment was performed independently and in duplicate for all studies by LK, AL, or CJMD. Disagreements were resolved through discussion or arbitration by a third reviewer, if necessary. The authors of the studies were not contacted for additional information in case of missing data or methodological unclarities.

Data Synthesis and Analysis

Risk differences between telehealth and usual care were calculated for data reported as cumulative incidences. Cumulative incidences reported as percentages were converted to the number of participants with events. For data reported as means, such as the mean number of hospitalizations per patient, the mean differences (MDs) between telehealth and usual care were calculated. Missing SDs were calculated, where possible. All calculations were performed according to Chapter 6 of the Cochrane Handbook [14]. Meta-analyses were conducted with the meta package in R, Version 3.6.3, (R Foundation for Statistical Computing) [15], using Mantel-Haenszel random-effects models. Hartung-Knapp adjustment is used to better reflect the uncertainty in the estimation of between-study heterogeneity in CIs [16,17].

The overall quality of evidence was rated according to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach (Multimedia Appendix 4) [18]. GMP rated the quality of evidence for each outcome (Multimedia Appendix 5). This rating was verified by all other authors, and disagreements were resolved by discussion.

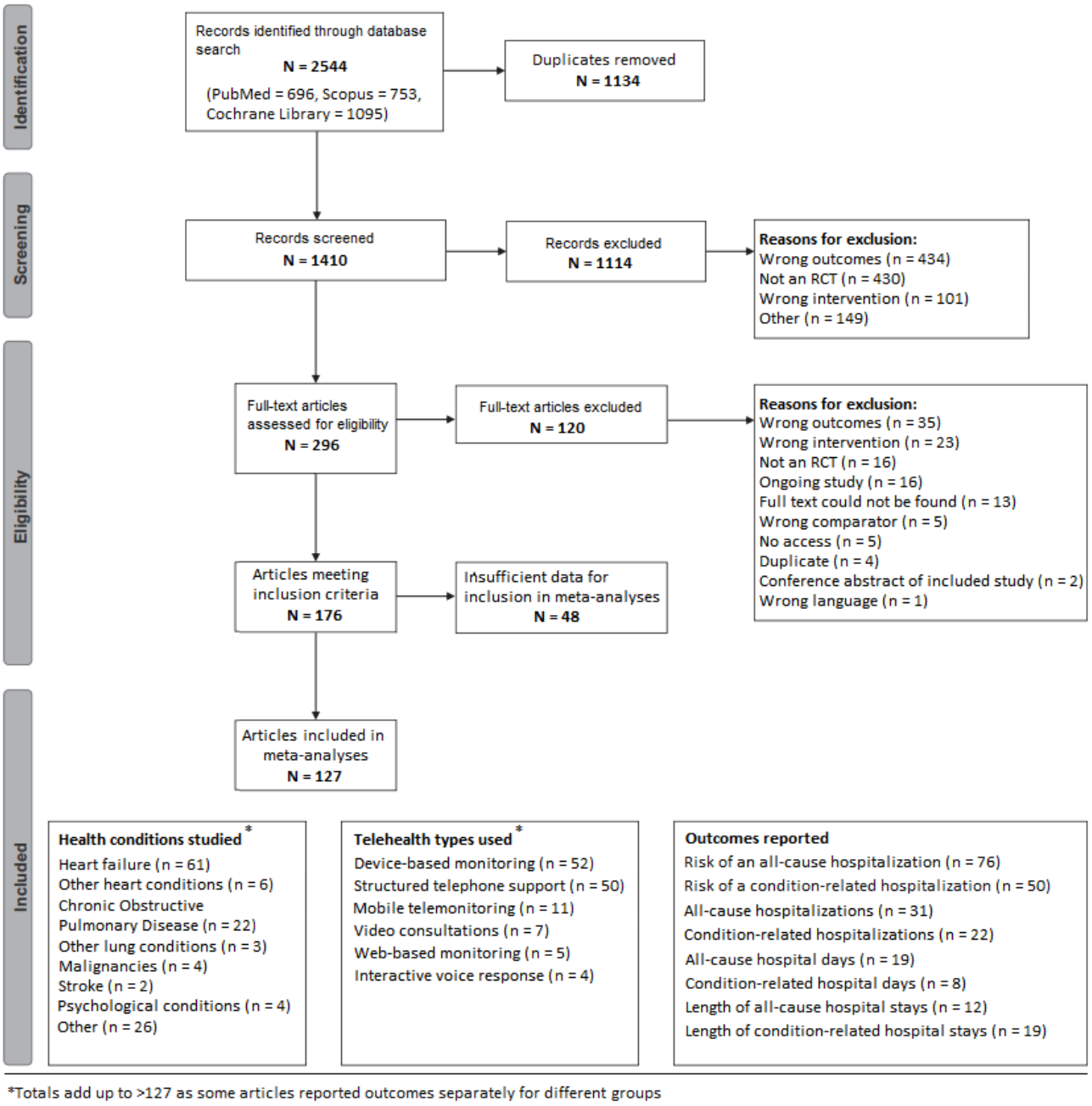
We conducted subgroup analyses for health conditions that were studied in at least two articles, as well as for each type of telehealth, length of follow-up, and risk of bias. These analyses were planned a priori. The risk of bias was analyzed using the robvis package in R [19]. To assess publication bias, we visually inspected funnel plots (using the meta package in R).

Results

Study Selection

The search identified 2544 records. After removing duplicates, 1410 records remained for the screening of titles and abstracts, through which 1114 (79.0%) records were excluded. We assessed 296 full-text articles for eligibility and excluded 120 articles. Of the remaining 176 articles, 127 (72.2%) provided sufficient data for inclusion in the meta-analysis (Multimedia Appendix 6). Figure 1 provides an overview of the study selection process.

Figure 1. Study selection flowchart and study characteristics. RCT: randomized controlled trial.



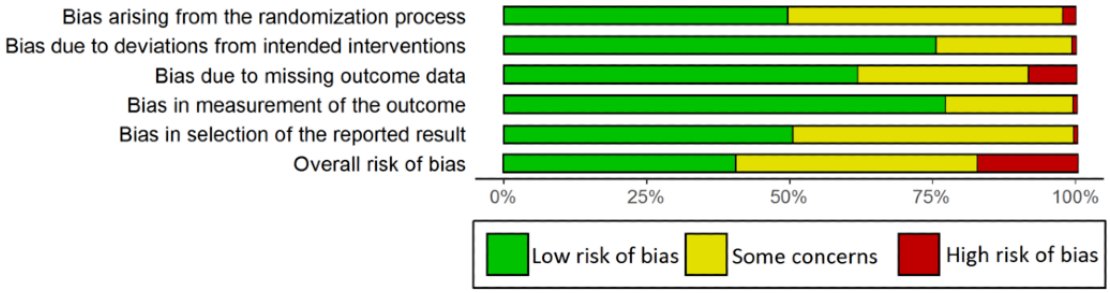
Study Characteristics

An overview of telehealth types, health conditions, and outcomes is provided in Figure 1 (details are provided in Multimedia Appendix 3). Most studies were conducted in Europe (n=55) and North America (n=41).

Risk of Bias

We judged 50 articles to be at low overall risk of bias, 55 to have some concerns, and 22 to be at high risk of bias. Most articles were assessed at low risk of bias for all five domains (64/127, 50.4% to 98/127, 77.2%), except for selection of the reported result (63/127, 49.6%; Figure 2). High risk was found for bias arising from the randomization process in only 3 articles, bias due to deviations from intended interventions in one, due to missing outcome data in 11, bias in measurement of the outcome in one, and in selection of the reported result in 1 out of 127 articles. Weighted risk of bias summaries are provided for each analysis in Multimedia Appendix 5. In the analyses of condition-related hospitalizations and the length of hospital stay due to any cause, studies at high risk of bias in at least one domain cumulatively accounted for approximately 20% of the weight. In all other analyses, this figure was below 10%.

Figure 2. Unweighted risk of bias summary.



Outcomes

The summary of findings table (Table 1) provides a comprehensive overview of the main results for all outcomes.

For each analysis, most RCTs used device-based monitoring or STS and included mainly patients with heart failure or chronic obstructive pulmonary disease (COPD; details Multimedia Appendix 3). Complete analyses are available in Multimedia Appendix 5.

The outcomes are reported as rates in 14 articles. Although these could not be incorporated in the meta-analyses, an overview of these results is provided in Multimedia Appendix 7.

Table 1. Summary of findings table for the effect of telehealth interventions on various outcome measures compared with usual care.

Outcome	Studies (RCTs[a]), n	Participants, n	Follow-up (months)	Usual care estimate	Intervention effect estimate	Effect estimate (95% CI)	GRADE[b] Strength of evidence[c]	Plain language summary
Patients with an all-cause hospitalization (patients hospitalized per 1000 patients)	76	34,423	1-60	373	355	Risk difference: −18 (−30 to −0)	High	The number of patients hospitalized for any cause is reduced by 4.8%[d]
Patients with a condition-related hospitalization (hospitalizations per 1000 patients)	50	20,867	1-60	237	200	Risk difference: −37 (−60 to −20)	High	The number of patients hospitalized for the condition targeted is reduced by 15.6%[d]
Mean all-cause hospitalizations per patient (hospitalizations per 1000 patients)	31	11,191	3-12	880	830	Mean difference: −50 (−140 to +30)	High	All-cause hospitalizations are reduced by 5.7%[d]
Mean condition-related hospitalizations per patient (hospitalizations per 1000 patients)	22	3461	1-60	470	360	Mean difference: −110 (−200 to −10)	High	Condition-related hospitalizations are reduced by 23.4%[d]
All-cause hospital days[e] (hospital days per patient)	19	9735	0-60	6.06	4.99	Mean difference: −1.07 (−1.76 to −0.39)	High	The mean number of days spent in the hospital for any cause per patient is reduced by 17.7%[d]
Condition-related hospital days[e] (hospital days per patient)	8	1216	3-60	2.84	1.71	Mean difference: −1.13 (−1.64 to −0.61)	Moderate[f]	The mean number of days spent in the hospital for the condition targeted is reduced by 39.8%[d]
Length of all-cause hospital stay[g] (days per hospitalization)	12	1964	0-60	8.37	7.89	Mean difference: −0.48 (−1.50 to 0.53)	Low[h]	Hospitalizations for any cause are 5.7%[d] shorter with telehealth
Condition-related hospital length of stay[g] (days per hospital stay)	15	2047	0-24	2.92	2.03	Mean difference: −0.89 (−1.42 to −0.36)	High	Hospitalizations for the condition targeted are 30.5%[d] shorter with telehealth

[a]RCT: randomized controlled trial.

[b]GRADE: Grading of Recommendations Assessment, Development and Evaluation.

[c]High: we are very confident that the true effect lies close to that of the estimate of the effect; moderate: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different; low: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.

[d]Percentages were calculated by dividing the effect estimate by the usual care estimate.

[e]Participants are the unit of analysis.

[f]Downgraded by one level for risk of publication bias.

[g]Hospitalizations are the unit of analysis.

[h]Downgraded by one level for risk of bias and another for imprecision.

Risk of All-Cause Hospitalization

The risk of all-cause hospitalization was reported by 76 RCTs, including 34,423 participants. The analysis provides high-quality evidence for a risk difference of −18 (95% CI −30 to 0) hospitalized patients per 1000 patients (−4.8% of usual care).

Risk of Condition-Related Hospitalization

We found 50 RCTs reporting the risk of condition-related hospitalization, including 20,867 participants. The absolute risk was reduced by 37 per 1000 patients (95% CI 20-60), with high-quality evidence (−5.7% of usual care). When stratified by health condition, only the heart failure group showed a statistically significant effect (risk difference = −0.03), although the subgroup difference was not significant (P=.40).

Mean All-Cause Hospitalizations

We found 31 RCTs reporting the mean number of all-cause hospitalizations per patient, including 11,191 participants. Follow-up varied between 3 and 12 months. The analysis showed high-quality evidence for an MD of −50 (95% CI −140 to +30) hospitalizations per 1000 patients, a 5.7% reduction with regards to the number of hospitalizations in the usual care group. Only the COPD subgroup showed a statistically significant MD between telehealth and usual care of −200 (95% CI −390 to −10) hospitalizations per 1000 patients. No effects

were found for heart failure and other diseases. In addition, an RCT studying malignancies reported an MD of +0.09 hospitalizations per patient compared with usual care but did not report a SD and was therefore excluded from the meta-analysis.

Mean Condition-Related Hospitalizations

The mean number of condition-related hospitalizations per patient was reported in 22 RCTs, including 3461 participants. Follow-up varied between 1 and 60 months. The analysis showed high-quality evidence for an MD of −110 (−200 to −10; −23.4% of usual care) hospitalizations per 1000 patients with telehealth compared with usual care. Differences between outcomes appeared to depend on the length of follow-up (P<.01). The difference increased gradually with a longer follow-up from an MD of −90 between 3 and 6 months up to a reduction of 1190 hospitalizations per 1000 patients for outcomes reported after more than 12 months. When stratified by health condition, only heart failure showed a statistically significant effect (MD −120; −200 to −40 hospitalizations per 1000 patients).

All-Cause Hospital Days

The mean number of days patients were hospitalized for any cause was reported in 19 RCTs including 9735 participants. Overall, the analysis showed high quality evidence for an MD of −1.07 (95% CI −1.76 to −0.39) hospital days per patient. In addition, 9 RCTs reported the total number of days for which patients were hospitalized, and 2 reported the rate of hospital days. Furthermore, 1 RCT reported an MD of +0.60 hospital days with telehealth compared with usual care but did not report an SD nor the necessary information to calculate one. These 12 RCTs, which included 3144 participants, could not be incorporated in the meta-analysis.

Condition-Related Hospital Days

The mean number of days patients were hospitalized for the condition of interest was reported by 8 RCTs, including a total of 1216 participants. The analysis showed moderate quality evidence of an MD of −1.13 (95% CI −1.64 to −0.61) hospital days per patient. The quality of evidence was downgraded because of risk of publication bias. A statistically significant difference was found for the length of follow-up (P<.01), with longer follow-up resulting in larger reductions in hospital days. It is notable that when stratified by health condition, a statistically significant result was only achieved in heart failure (MD −1.06 hospital days, 95% CI −1.71 to −0.40). For COPD, an MD of −1.75 (95% CI −4.62 to 1.11) was found. In addition, 7 studies reported the total number of days patients were hospitalized, and one reported the rate of hospital days. These studies, including 2492 participants, could not be included in the meta-analysis.

Length of All-Cause Hospital Stay for Hospitalized Patients

A total of 12 RCTs reported length of all-cause hospital stay, including 1964 hospitalized patients. Low-quality evidence was found for an MD of −0.48 (95% CI −1.44 to +0.47 days) hospital days per stay. The quality of evidence was downgraded by one level for risk of bias and by another for imprecision. Subgroup differences were found between different lengths of follow-up (P<.01) and different levels of risk of bias (P≤.01), but no clear trends were found. Three studies reported the length of hospital stay as medians and IQRs, and they could therefore not be included in the meta-analysis.

Length of Condition-Related Hospital Stay for Hospitalized Patients

Fifteen RCTs reported length of condition-related hospital stay, including 2047 hospitalized patients. The analysis showed high-quality evidence for an MD of −0.89 hospital days per stay (95% CI −1.42 to −0.36 days).

Subgroup differences were found in reporting outcomes at different lengths of follow-up (P<.01). An MD of −3.95 hospital days per stay (95% CI −6.06 to −1.84 days) was found for reporting between 7 and 12 months, whereas other MDs ranged from −1.00 to −0.42 days. An additional 3 RCTs reported the length of hospital stay as medians and IQRs and 4 did not report SDs nor any information that could be used to calculate them. These 7 RCTs, including 922 participants, were therefore excluded from the meta-analysis.

Discussion

Principal Findings

Our review indicates that the risk of all-cause hospitalization decreased significantly by 18 hospitalizations per 1000 patients (−4.8%) and 37 (−15.6%) for condition-related hospitalizations. We found high-quality evidence that, compared with usual care, telehealth leads to reductions in mean all-cause (MD −0.05, 95% CI −0.14 to 0.03 hospitalizations per patient; −5.7% of usual care) and condition-related hospitalizations (MD −0.11, 95% −0.20 to −0.01; −23.4%), that is, 50 to 110 fewer mean hospitalizations, respectively, per 1000 patients. Overall, it is evident that all-cause hospital days decreased significantly with a mean of −1.07 (−17.7%) hospital days per patient and condition-related hospital days with −1.13 (−39.8%) days, although evidence for the latter was only moderate. For hospitalized patients, the mean stay for any cause could potentially be reduced (MD −0.48 days, 95% CI −1.50 to 0.53; 5.7%, low-quality evidence), and mean stay for condition-related hospitalizations even more (MD −0.89 days, 95% CI −1.42 to −0.36; 30.5%, high-quality evidence). The effects were similar for various health conditions and types of telehealth. A trend was observed for studies with longer follow-up periods, yielding larger effects. It should, however, be noted that, although this is a systematic review including a large number of studies, uncertainties around the magnitude of effects remain, and not all differences were statistically significant.

The quality of evidence was high for most of the analyses. Downgrading was only necessary for two analyses because of the risk of bias, risk of publication bias, and imprecision because of a small cumulative sample size. Overall, there were approximately as many articles with some concerns as there were articles at low risk of bias. The main culprits were insufficient reporting of the randomization method, lack of available trial registrations or study protocols, and incomplete outcome data (mostly due to deaths). None of these aspects

necessarily indicate issues with the study itself, but rather with the reporting of a study. It is desirable that more information is made available, such as by providing web-based supplementary material.

Comparison With Prior Work

In our review, the most commonly used telehealth types were device-based monitoring and STS. In general, only small differences in effects were found between telehealth types, which did not appear to be relevant. This finding is in line with a Cochrane review including RCTs investigating the effect of either STS or device-based monitoring in the management of heart failure, which also found no difference [20]. It should be explored whether design aspects, such as monitoring frequency or duration, or patient engagement, could explain the differences in effect. Furthermore, patient compliance is often important for the success of telehealth interventions. For example, the patients must consistently take and send measurements, be available for telephone contacts or video consultations, or report symptoms. If these actions are not taken by the patient, telehealth interventions cannot function. Therefore, it is important to consider patient preferences during the design process [21,22].

Studies including patients with heart failure or COPD accounted for the majority of the weight in the meta-analyses of this review, although the effects found for other health conditions seemed similar. No other review has combined the results for multiple health conditions. However, reviews of heart failure and COPD specifically are available for comparison. A systematic review including reviews on telehealth for chronic heart failure patients published between 1996 and 2014 found low-quality evidence for absolute risk reductions in patients with an all-cause hospitalization of 4.7% to 13.8% and of 3.7% to 8.2% for patients with a condition-related hospitalization [6]. Our estimate for patients with all-cause hospitalization was considerably lower (2%) and more precise. This is caused by the larger number of studies (75 in our study vs 8 in the other meta-analysis) and thus participants in our analysis (N=30,937 vs N=2343). Our estimate for patients with condition-related hospitalization was similar (3.8%). A recent review on telehealth for heart failure patients also found a trend toward reduced hospitalizations [23]. Another recent review, on coronary heart disease patients, found a relative risk of 0.56 (95% CI 0.39-0.81), although absolute differences were also small [24].

A systematic overview of reviews including COPD patients found 3 reviews investigating the effect of telehealth on hospitalizations, all of which found a reduction in hospitalizations [7]. Another systematic review reported reduced hospitalizations in 8 out of 11 studies, ranging from −10% to −63%. The findings were similar for all-cause hospitalization and condition-related hospitalizations [25]. Our review confirms the reduction in hospitalizations also found in previous reviews and provides a more realistic estimate of the effect through meta-analyses, which was rarely performed in previous reviews.

In a systematic overview of the use of telehealth for various chronic health conditions, reviews on health conditions other than heart failure or COPD also found only a few articles, except for diabetes [8]. This result is consistent with the findings of our review. As COPD and heart failure only make up a small part of the care provided by hospitals [26], more research is necessary on the effect of telehealth on hospital services use in health conditions other than COPD and heart failure, which are also highly prevalent.

The length of follow-up seems to be an important factor influencing the effect of telehealth in our review. We found subgroup differences in length of hospital stay (both all-cause and condition-related), condition-related hospitalizations, and condition-related hospital days, with larger effect sizes for studies with longer follow-up. A similar trend was observed for all-cause hospital days. One review reported a reduction in mortality at 6 months, with no differences at 1 year [21]. No other reviews assessed differences in effects between the lengths of follow-up.

When telehealth replaces face-to-face contact, it is clear that this can aid in reducing outpatient contacts and supporting social distancing in outpatient departments. In view of the small effects on hospitalizations and moderate effects on hospital inpatient days, it is important to determine whether telehealth actually contributes to cost reduction. Telehealth comes at a cost, for example, because health professionals make phone calls, conduct video consultations, or interpret data. To reduce the costs of interventions, automation of some of these aspects, for example, by developing algorithms to recognize deterioration of patients' health status, should be studied. Although we investigated whether the mechanism by which telehealth is often claimed to reduce costs is indeed present, we did not directly investigate whether costs were reduced. Thorough budget impact and cost-effectiveness studies are needed to reach firm conclusions in this domain.

Limitations

This review has several strengths and limitations. First, the wide scope enabled us to find a large number of articles meeting our inclusion criteria. Furthermore, we quantitatively compared the effects achieved in different health conditions using different types of telehealth and length of follow-up. Another important strength is that we assessed all included articles for risk of bias and graded the strength of evidence for each analysis, providing a comprehensive overview of the evidence on the effect of telehealth on hospital service use.

The wide scope also acts as a double-edged sword in that it makes the participants in the various studies less comparable than in a typical review. This concern is alleviated by the fact that we did not find significant differences between health conditions or types of telehealth, although for some comparisons only a few studies were available. Telehealth interventions often entail many more changes to the health care process, besides the application of technology [27]. The effect of the telehealth type thus becomes entangled with the effects of changes to processes and infrastructure, which requires a more detailed analysis to unravel. Study selection was performed partially in duplicate, which may have caused some articles to have been missed. As we only included peer reviewed articles published in English, it is unknown what evidence exists in other languages. This review is further limited by our scope, which focuses on types of telehealth requiring interaction between

patients and health professionals. Passive forms of digital health care, such as self-management applications or health information provision, were not included. These types of services could reduce hospital service use [28], while potentially being more efficient in terms of resource use because of their passive nature. Furthermore, we did not contact the study authors for details in the case of missing data or methodological unclarities.

Conclusions

Thus, the effects of telehealth are small to moderate and appear to be stronger for condition-related outcomes than for all-cause outcomes. Further research is needed to obtain more insight into the effects of telehealth on other diseases, apart from COPD and heart failure, and into which aspects of telehealth interventions result in positive effects.

Finally, in the context of the COVID-19 crisis, it is important to acknowledge that a great deal of health care can be provided from a distance, eliminating the need for vulnerable individuals to come to a potentially hazardous environment to receive health care and enabling hospitals to continue providing care to all who need it.

Acknowledgments

This study was funded by a nonrestricted grant from Menzis, a Dutch insurance company. The funding source was not involved in the design of the study, data analysis, writing of the manuscript, or the decision to submit for publication.

GMP was involved in the study concept and design, data acquisition (abstract and full-text screening, risk of bias assessment, data extraction), analyses and interpretation of the results, and drafting and revising of the manuscript. CJMD was involved in the study concept and design, interpretation of the results, and critically revised the manuscript for important intellectual content. LK and AL were involved in data acquisition and critically revised the manuscript. WHVH was involved in the study concept, obtained funding, and critically revised the manuscript. CJMD and WHVH both supervised the study equally. All authors approved the final version of the manuscript.

The authors thank Martin Hemels (Rijnstate Hospital) for providing input from a clinical perspective.

Conflicts of Interest

GMP, LK, AL, and CJMD have nothing to declare. WHVH received nonrestricted research grants from Novartis and Agendia BV.

Multimedia Appendix 1
Deviation from and clarification of Cochrane Risk of Bias 2 Tool guidance document.
[DOCX File , 12 KB - jmir_v23i9e25195_app1.docx]

Multimedia Appendix 2
Search syntaxes.
[DOCX File , 15 KB - jmir_v23i9e25195_app2.docx]

Multimedia Appendix 3
Study characteristics.
[DOCX File , 70 KB - jmir_v23i9e25195_app3.docx]

Multimedia Appendix 4
Grading of Recommendations Assessment, Development and Evaluation protocol.
[DOCX File , 12 KB - jmir_v23i9e25195_app4.docx]

Multimedia Appendix 5
Complete analyses and Grading of Recommendations Assessment, Development and Evaluation assessments.
[DOCX File , 1680 KB - jmir_v23i9e25195_app5.docx]

Multimedia Appendix 6
Studies included in the meta-analysis.
[DOCX File , 27 KB - jmir_v23i9e25195_app6.docx]

Multimedia Appendix 7
Hospitalization rates.
[DOCX File , 28 KB - jmir_v23i9e25195_app7.docx]

References

1. Lindberg B, Nilsson C, Zotterman D, Söderberg S, Skär L. Using information and communication technology in home care for communication between patients, family members, and healthcare professionals: a systematic review. Int J Telemed Appl 2013;2013:1-31. [doi: 10.1155/2013/461829] [Medline: 23690763]
2. World Population Ageing 2015 (ST/ESA/SER.A/390). United Nations, Department of Economic and Social Affairs, Population Division. 2015. URL: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2015_Report.pdf [accessed 2021-07-31]
3. National Health Expenditures 2017 Highlights. Centers for Medicare & Medicaid Services. 2017. URL: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/highlights.pdf [accessed 2019-05-13]
4. Healthcare expenditure by provider, 2015 (% of current healthcare expenditure). Eurostat. 2015. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Healthcare_expenditure_by_provider,_2015_(%25_of_current_healthcare_expenditure)_FP18a.png [accessed 2019-05-13]
5. OECD. Health expenditure in hospitals. In: Health at a Glance: Europe 2018: State of Health in the EU Cycle. Paris: OECD Publishing; 2018:138-139.
6. Kitsiou S, Paré G, Jaana M. Effects of home telemonitoring interventions on patients with chronic heart failure: an overview of systematic reviews. J Med Internet Res 2015 Mar 12;17(3):e63 [FREE Full text] [doi: 10.2196/jmir.4174] [Medline: 25768664]
7. Murphy LA, Harrington P, Taylor SJ, Teljeur C, Smith SM, Pinnock H, et al. Clinical-effectiveness of self-management interventions in chronic obstructive pulmonary disease: an overview of reviews. Chron Respir Dis 2017 Aug 24;14(3):276-288 [FREE Full text] [doi: 10.1177/1479972316687208] [Medline: 28774200]
8. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth interventions to support self-management of long-term conditions: a systematic metareview of diabetes, heart failure, asthma, chronic obstructive pulmonary disease, and cancer. J Med Internet Res 2017 May 17;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
9. McLean S, Sheikh A, Cresswell K, Nurmatov U, Mukherjee M, Hemmi A, et al. The impact of telehealthcare on the quality and safety of care: a systematic overview. PLoS One 2013 Aug 19;8(8):e71238 [FREE Full text] [doi: 10.1371/journal.pone.0071238] [Medline: 23977001]
10. Shields GE, Wells A, Doherty P, Heagerty A, Buck D, Davies LM. Cost-effectiveness of cardiac rehabilitation: a systematic review. Heart 2018 Sep 13;104(17):1403-1410 [FREE Full text] [doi: 10.1136/heartjnl-2017-312809] [Medline: 29654096]
11. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions 6.0. Canada: Cochrane; 2019.
12. Covidence systematic review software. Veritas Health Innovation. Melbourne, Australia URL: https://www.covidence.org/ [accessed 2021-07-31]
13. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. Br Med J 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
14. Higgins J, Li T, Deeks JJ. Chapter 6: Choosing effect measures and computing estimates of effect. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 6.0. Canada: Cochrane; 2019.
15. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid Based Ment Health 2019 Dec 28;22(4):153-160. [doi: 10.1136/ebmental-2019-300117] [Medline: 31563865]
16. Hartung J, Knapp G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. Stat Med 2001 Dec 30;20(24):3875-3889. [doi: 10.1002/sim.1009] [Medline: 11782040]
17. Deeks J, Higgins J, Altman D. Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 6.0. Canada: Cochrane; 2019.
18. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. Br Med J 2008 Apr 26;336(7650):924-926 [FREE Full text] [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]
19. McGuinness L. robvis: An R package and web application for visualising risk-of-bias assessments. Github. URL: https://github.com/mcguinlu/robvis [accessed 2021-07-31]
20. Inglis SC, Clark RA, McAlister FA, Stewart S, Cleland JG. Which components of heart failure programmes are effective? A systematic review and meta-analysis of the outcomes of structured telephone support or telemonitoring as the primary component of chronic heart failure management in 8323 patients: abridged Cochrane review. Eur J Heart Fail 2011;13(9):1028-1040 [FREE Full text] [doi: 10.1093/eurjhf/hfr039] [Medline: 21733889]
21. Hashemi A, Nourbakhsh S, Tehrani P, Karimi A. Remote telemonitoring of cardiovascular patients: benefits, barriers, new suggestions. ARTRES 2018;22(C):57. [doi: 10.1016/j.artres.2018.04.001]

22. Shah SG, Robinson I, AlShawi S. Developing medical device technologies from users' perspectives: a theoretical framework for involving users in the development process. Int J Technol Assess Health Care 2009 Oct;25(4):514-521. [doi: 10.1017/S0266462309990328] [Medline: 19845981]
23. Pekmezaris R, Tortez L, Williams M, Patel V, Makaryus A, Zeltser R, et al. Home telemonitoring in heart failure: a systematic review and meta-analysis. Health Aff (Millwood) 2018 Dec;37(12):1983-1989. [doi: 10.1377/hlthaff.2018.05087] [Medline: 30633680]
24. Jin K, Khonsari S, Gallagher R, Gallagher P, Clark AM, Freedman B, et al. Telehealth interventions for the secondary prevention of coronary heart disease: a systematic review and meta-analysis. Eur J Cardiovasc Nurs 2019 Apr;18(4):260-271. [doi: 10.1177/1474515119826510] [Medline: 30667278]
25. Pedone C, Lelli D. Systematic review of telemonitoring in COPD: an update. Adv Respir Med 2015 Nov;83(6):476-484 [FREE Full text] [doi: 10.5603/PiAP.2015.0077] [Medline: 26559801]
26. Torio C, Moore B. National inpatient hospital costs: the most expensive conditions by payer, 2013. HCUP Statistical Brief #204. Rockville, MD: Agency for Healthcare Research and Quality; 2016. URL: http://www.hcup-us.ahrq.gov/reports/statbriefs/sb204-Most-Expensive-Hospital-Conditions.pdf [accessed 2020-07-03]
27. Kooij L, Groen WG, van Harten WH. The effectiveness of information technology-supported shared care for patients with chronic disease: a systematic review. J Med Internet Res 2017 Jun 22;19(6):e221 [FREE Full text] [doi: 10.2196/jmir.7405] [Medline: 28642218]
28. Lenferink A, Brusse-Keizer M, van der Valk PD, Frith PA, Zwerink M, Monninkhof EM, et al. Self-management interventions including action plans for exacerbations versus usual care in patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2017 Aug 04;8:CD011682 [FREE Full text] [doi: 10.1002/14651858.CD011682.pub2] [Medline: 28777450]

Abbreviations

COPD: chronic obstructive pulmonary disease
GRADE: Grading of Recommendations Assessment, Development and Evaluation
ICT: information and communication technology
IVR: interactive voice response
MD: mean difference
MeSH: Medical Subject Headings
RCT: randomized controlled trial
STS: structured telephone support

Edited by R Kukafka; submitted 22.10.20; peer-reviewed by M Behzadifar, H Zihao; comments to author 01.12.20; revised version received 17.12.20; accepted 11.06.21; published 01.09.21.

Please cite as:
Peters GM, Kooij L, Lenferink A, van Harten WH, Doggen CJM
The Effect of Telehealth on Hospital Services Use: Systematic Review and Meta-analysis
J Med Internet Res 2021;23(9):e25195
URL: https://www.jmir.org/2021/9/e25195
doi:10.2196/25195
PMID:34468324

Review

Evolutionary Overview of Consumer Health Informatics: Bibliometric Study on the Web of Science from 1999 to 2019

Wei Ouyang[1,2,3*], PhD; Wenzhao Xie[1,2,3*], PhD; Zirui Xin[1,2,3], PhD; Haiyan He[1,2,3], PhD; Tingxiao Wen[2,3], PhD; Xiaoqing Peng[1,2,3], PhD; Pingping Dai[1,2,3], PhD; Yifeng Yuan[2,3,4], MS; Fei Liu[1,2,3], MS; Yang Chen[1,2,3], MS; Aijing Luo[4], PhD

[1]The Third Xiangya Hospital, Central South University, Changsha, China

[2]School of Life Sciences, Central South University, Changsha, China

[3]Key Laboratory of Medical Information Research, Central South University, College of Hunan Province, Changsha, China

[4]The Second Xiangya Hospital, Central South University, Changsha, China

[*]these authors contributed equally

Corresponding Author:
Aijing Luo, PhD
The Second Xiangya Hospital
Central South University
139 Renmin Middle Road
Changsha, 410011
China
Phone: 86 073188618316
Email: luoaijing@163.com

Abstract

Background: Consumer health informatics (CHI) originated in the 1990s. With the rapid development of computer and information technology for health decision making, an increasing number of consumers have obtained health-related information through the internet, and CHI has also attracted the attention of an increasing number of scholars.

Objective: The aim of this study was to analyze the research themes and evolution characteristics of different study periods and to discuss the dynamic evolution path and research theme rules in a time-series framework from the perspective of a strategy map and a data flow in CHI.

Methods: The Web of Science core collection database of the Institute for Scientific Information was used as the data source to retrieve relevant articles in the field of CHI. SciMAT was used to preprocess the literature data and construct the overlapping map, evolution map, strategic diagram, and cluster network characterized by keywords. Besides, a bibliometric analysis of the general characteristics, the evolutionary characteristics of the theme, and the evolutionary path of the theme was conducted.

Results: A total of 986 articles were obtained after the retrieval, and 931 articles met the document-type requirement. In the past 21 years, the number of articles increased every year, with a remarkable growth after 2015. The research content in 4 different study periods formed the following 38 themes: *patient education*, *medicine*, *needs*, and *bibliographic database* in the 1999-2003 study period; *world wide web*, *patient education*, *eHealth*, *patients*, *medication*, *terminology*, *behavior*, *technology*, and *disease* in the 2004-2008 study period; *websites*, *information seeking*, *physicians*, *attitudes*, *technology*, *risk*, *food labeling*, *patient*, *strategies*, *patient education*, and *eHealth* in the 2009-2014 study period; and *electronic medical records*, *health information seeking*, *attitudes*, *health communication*, *breast cancer*, *health literacy*, *technology*, *natural language processing*, *user-centered design*, *pharmacy*, *academic libraries*, *costs*, *internet utilization*, and *online health information* in the 2015-2019 study period. Besides, these themes formed 10 evolution paths in 3 research directions: patient education and intervention, consumer demand attitude and behavior, and internet information technology application.

Conclusions: Averaging 93 publications every year since 2015, CHI research is in a rapid growth period. The research themes mainly focus on patient education, health information needs, health information search behavior, health behavior intervention, health literacy, health information technology, eHealth, and other aspects. Patient education and intervention research, consumer demand, attitude, and behavior research comprise the main theme evolution path, whose evolution process has been relatively stable. This evolution path will continue to become the research hotspot in this field. Research on the internet and information technology application is a secondary theme evolution path with development potential.

(J Med Internet Res 2021;23(9):e21974) doi:10.2196/21974

KEYWORDS

consumer health informatics; consumer health information; thematic evaluation; co-word analysis; informatics; SciMAT

Introduction

The concept prototype of consumer health informatics (CHI) was first proposed by Kenneth R. Thornton of the School of Health Information Science (University of Victoria) in 1994. Kenneth R. Thornton also elaborated 4 important research directions: network evolution, automation of the patient record, outcome and other quality-related databases, and consumer health education [1]. In 1995, Ferguson [2] of the Harvard University School of Medicine put forward the concept of consumer health informatics (CHI) for the first time in his paper. He pointed out that CHI is a branch of science that studies the application of computer and wireless communication technology in consumer health care [2]. Eysenbach [3] proposed that CHI is the branch of medical information that does the following: analyses the consumers' needs for information; studies and implements methods of making information accessible to consumers; and models and integrates the consumers' preferences into medical information systems. This definition has been cited by the academic community more than 300 times. The American Medical Informatics Association defines CHI as the field devoted to informatics from multiple consumer or patient views and which includes patient-focused informatics, health literacy, and consumer education [4]. CHI is an interdisciplinary subject that includes nursing information, public health, health promotion, health education, library science, and communication science [4].

With the rapid development of computer and information technology, the continuous popularization of the internet, and the continuous strengthening of people's health awareness, consumption awareness, and information literacy for making health decisions, increasingly more people obtain health-related information through the internet. Because of the interdisciplinary nature of CHI, there is no consensus on the definition of CHI. However, the core content of various definitions can be summarized as follows: CHI emphasizes consumers or users as the center and takes consumers or users and computer and information technology as the research object to explore how to use computer and information technology to meet consumers' health or medical information needs. It is a discipline that helps consumers access relevant health or medical information and make decisions about health care and health promotion.

Previous studies have analyzed and summarized the research progress of CHI. For example, Eysenbach [3] collected literature, internet information, and reports related to CHI before 2000 and summarized the research progress of CHI in the context of health care in the information age, medical knowledge delivery to consumers, the accessibility of electronic health records to patients, decision aids to support the consumers' choices, the quality control of health information on the internet, and other aspects. Kokol et al [5] carried out a bibliometric analysis on the literature related to health informatics and electronic health for the period 1984-2015 and discussed the current research status in this field, including trends in literary production, the geographic and journal distribution, and theoretical analyses. However, their research only clusters all research topics and does not reflect the dynamic changes in research themes at different times. Zhao and Zhang [6] reviewed the literature about consumer health information seeking in social media before 2016 and discussed the characteristics of existing research from the following perspectives: the prevalence of health information seeking in social media, discussion topics emerging from health information in social media, seeking health information from online peers, social and emotional support from social media, concerns of accessing consumer health information in social media, and other aspects. However, they did not discuss the dynamic evolution path and the evolution rules of the research themes of CHI. Further, many of these studies are qualitative research. There is also considerable subjectivity in the selection of literature, identification of important topics, and prediction of research frontiers. The combination of qualitative and quantitative analysis methods can increase the objectivity, accuracy, and comprehensiveness of the research results [7]. Bibliometrics is a measurable informatic method [8], and is often used to discover top journals and authors in a field, identify research progress [9], and predict research trends [10].

Many bibliometrics visualization tools are available, such as SciMAT, CiteSpace, UCINET, HistCite, VOSviewer. The full name of SciMAT is the Science Mapping Analysis tool. In 2012, it was developed by Cobo, López-Herrera, Herrera-Viedma, and Herrera at the Department of Computer and Artificial Intelligence in Granada University, Spain. It can be used for data preprocessing, data network analysis documentation, and result visualization. It can also be used to produce 4 kinds of maps: an overlapping map, an evolution map, a strategic diagram, and a cluster network [11]. Compared with other bibliometric visualization tools, SciMAT is unique in expressing the evolution of the theme and performs excellently in longitudinal timing analysis [7].

Overall, there have been few achievements in the field of CHI research based on bibliometrics, and the dynamic evolution of the research theme of CHI has not been explored. Therefore, this article conducted a bibliometric analysis to present the evolutionary overview of CHI by using SciMAT. The main research questions of this paper are as follows:

- What was the development trend of CHI research in 1999-2019?
- What is the main research direction of CHI research?
- How does the theme of CHI research evolve?
- What are the research trends of CHI?

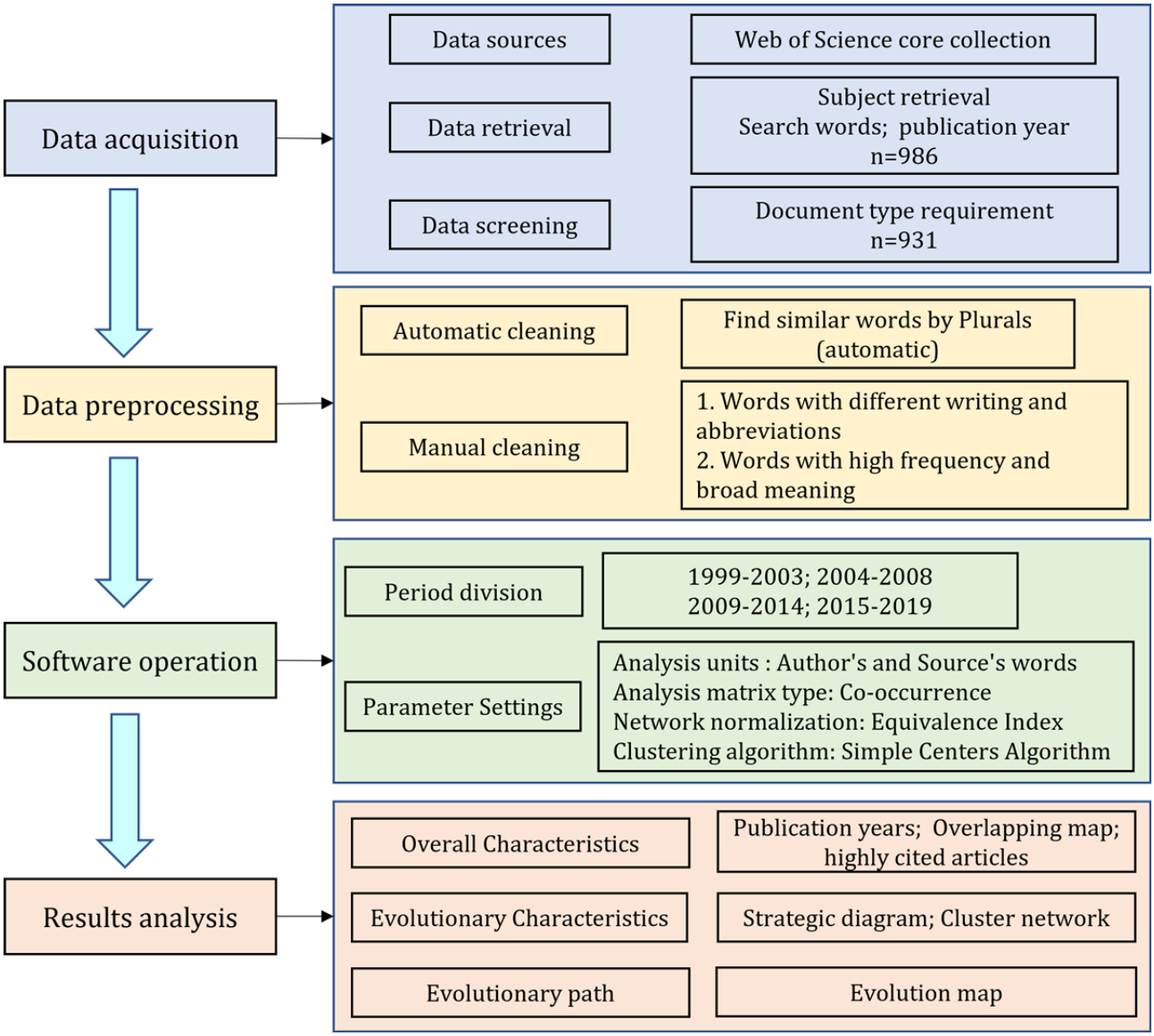
Methods

Overview

In bibliometrics, to obtain the research topic of a certain research field in a specific period, it is common to carry out keyword co-occurrence analysis and cluster analysis on the literature collected and to obtain the development path and state of the research topic through a comparative analysis of the split and fusion of research themes in different periods. This paper uses SciMAT tool to draw the research themes' knowledge map, analyzes the research themes and evolution status of CHI in different study periods using a strategic diagram and data flow, and discusses the dynamic evolution path and evolution law of the research themes of CHI.

The analysis framework of this paper is shown in Figure 1, which is mainly divided into 4 parts: data acquisition, data preprocessing, software operation, and results analysis.

Figure 1. Analysis framework of consumer health informatics evolutionary.



Data Sources and Retrieve Strategies

The Web of Science core collection database of the Institute for Scientific Information is an important database for obtaining global academic information [7]. It consists of the Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, and so on. Users can access core academic literature from the fields of natural sciences, social sciences, biomedicine, engineering, arts, and humanities from the database. Therefore, we selected Web of Science as the data source. In this database, subject retrieval was used, and the search terms were "consumer health informatics", "consumer health information", "consumer medic * informatics", "consumer medic * information". The publication years were limited to 2019 and before. All languages were selected. Selecting all languages provides a more comprehensive coverage of the literature on this topic. Non-English articles generally have English titles, abstracts, and keywords, from which one can understand their main research content. For these articles, we also used translation software to read the full text and obtained the main research content. The literature types were limited to articles, proceeding papers, and reviews. All papers

were retrieved on May 2, 2020. A total of 986 documents were obtained, of which 931 met the document-type requirement. The downloaded data were saved in a text format that SciMAT could read directly.

Data Preprocessing

To accurately obtain the research themes in this field, it was necessary to perform data cleaning. SciMAT was used to clean keyword information. First, SciMAT's cleaning function "find similar words by Plurals (automatic)" automatically merges the singular and plural expressions of keywords, such as "attitude" and "attitudes" into "attitudes". Then, similar words with different forms and abbreviations were merged by hand. For example, "consumer health information technology," "consumer health IT," "consumer health information technologies," and "consumer health information technology (CHIT)" were merged into " consumer health information technology." synonyms, such as "physician patient relations," "physician–patient relationship," "patient–physician relationship," "doctor–patient relationship," "doctor–patient relations," and "doctor–patient relationships" were merged into "doctor–patient relationship." Some words, such as "consumer health information," "consumer health informatics," "association," and "campaign," which had high frequency and broad meaning and that might cover the association between other micro words, were deleted.

Parameter Settings

Words (the author's words and the source's words) were selected as the units of analysis. The data reduction thresholds of the 4 study periods are 1, 1, 1, and 2, and the type of analysis matrix is co-occurrence. The network reduction thresholds are 1, 1, 1, and 1, and the network normalization method is the equivalence index. The clustering algorithm used was the simple centers algorithm. The maximum network value was 15 and the minimum network value was 3, and the scale of the cluster network was limited to a reasonable range. The H-index and sum citations were selected as the measurement indexes of clustering quality. Jaccard's index and the Salton index were selected as similarity measurement methods for the evolution map and the overlapping map, respectively.

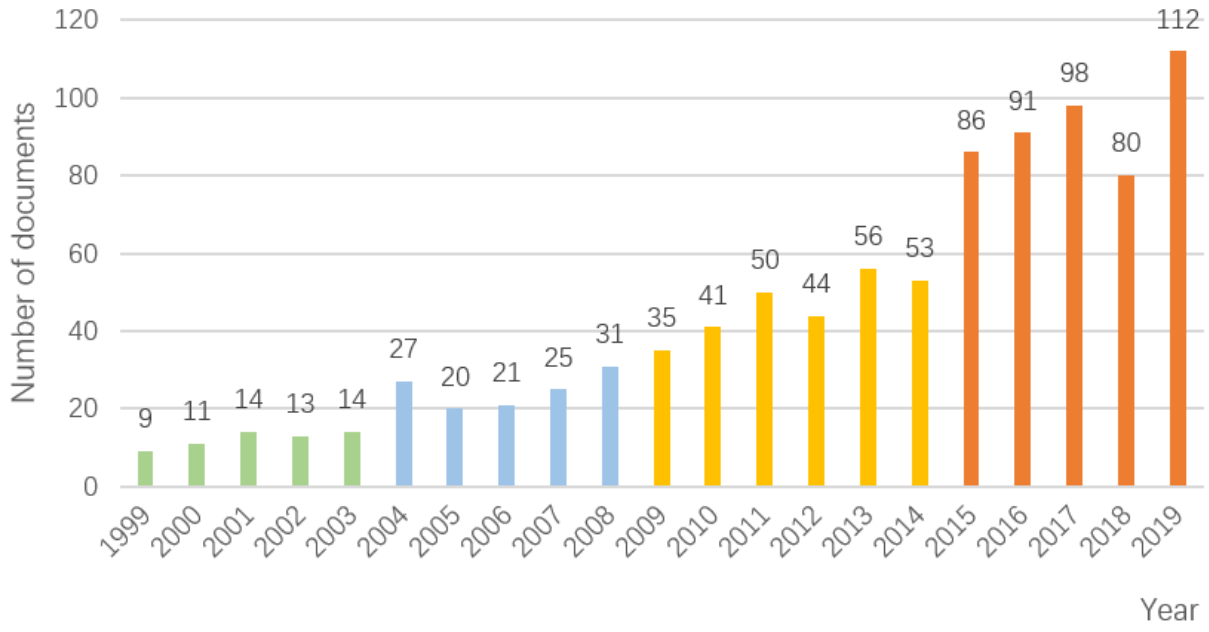
Results

Overall Characteristics Analysis

Trends in the Number of Articles

The number and trend of the articles published each year in a certain field can reflect the scholars' attention to this field. As shown in Figure 2, the amount of literature can be divided into 3 stages: slow growth, stable growth, and rapid growth. The 1999-2003 period was the first stage, with an average of 12.2 papers published each year. The period from 2004 to 2014 was the second stage, with an average of 36.6 papers published each year. The period from 2015 to 2019 was the third stage. After 2015, the number of articles increased substantially, with an average of 93.4 papers published each year. The volume of published articles peaked in 2019 at 112, with an annual total accounting for 12.0% (112/931) of all literature.

Figure 2. Consumer health informatic research documents published from 1999 to 2019.



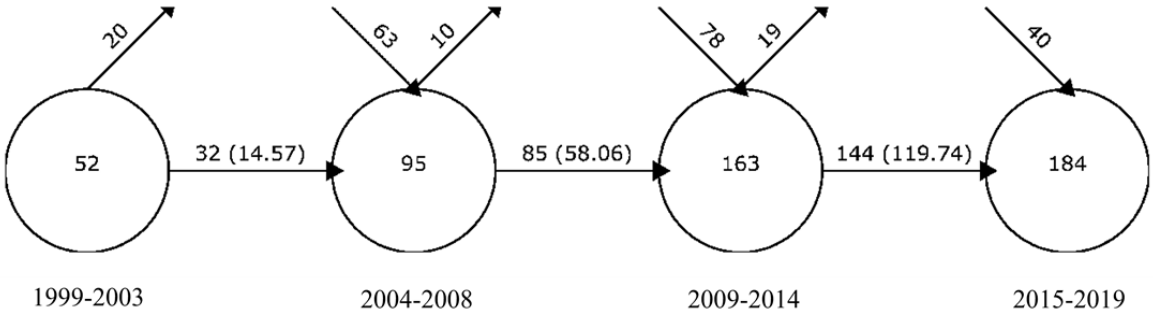
Based on the trend of literature growth, this paper divides the research on CHI into 4 study periods: 1999-2003, the first study period; 2004-2008, the second study period; 2009-2014, the third study period; and 2015-2019, the fourth study period. The number of articles in the 4 study periods is 61, 124, 279, and 467.

Overlapping Map Analysis

An overlapping map uses the number of keywords to represent the number of themes at each period and shows the stability of a research theme in a certain field in the form of a data flow. Figure 3 is an overlapping map of CHI research for the period from 1999 to 2019 and clearly shows the emerging and declining themes in this field. The 4 circles in the figure represent the 4

study periods. Displayed from left to right, they are as follows: the 1999-2003 study period, the 2004-2008 study period, the 2009-2014 study period, and the 2015-2019 study period. The number of themes in the first, second, third, and fourth study periods is 52, 95, 163, and 184, respectively. As can be seen, the number of themes increases rapidly.

Figure 3. Overlapping map of the consumer health informatics research from 1999 to 2019.



From 1999 to 2003, a total of 32 themes were passed to the next study period, and the stability index with the next study period was 14.57. In this study period, the research on CHI remains in a slow-growth stage, which is part of the basic research stage. The number of new themes in the 2004-2008 study period was 63. In the 2009-2014 period, 85 themes were continued from the previous study period, and the stability index between the 2 study periods was 58.06. Reflecting the steady growth of CHI research and the steady increase of research articles in these 2 study periods, many emerging themes and a small number of declining themes were included. There were 40 new themes in the 2015-2019 study period, and the stability index between this and the previous study period was 119.74. At this point, the number of themes is increasing steadily, which reflects that the research area of CHI is gradually expanding, and the research content is becoming increasingly richer.

Analysis of Highly Cited Articles

Six of the top 10 highly cited articles in this field came from the 1999-2003 study period (Table 1), and the cited frequency of the articles is ranked as the first, second, third, fifth, sixth, and tenth. This reflects that the papers published in this period form the basis of the research on CHI and play a fundamental role in the research in this field. The most frequently cited paper (940 times) is an empirical study of the quality evaluation of online consumer health information. The evaluation of health-related websites is quite different due to the differences in research methods, preciseness, quality standards, research population, and subjects. Therefore, the operability of quality standards needs to be defined [12]. The second most frequently cited (804 times) paper is the one by Cline in 2001 [13]. This paper summarizes, from the perspective of communication, the potential benefits and comprehensive quality of online consumer health information searching and identifies and discusses the criteria of online health information evaluation. In 1999, Charnock et al [14] developed a tool named DISCERN, which can be used by health information providers and consumers to judge the quality of written consumer health information about treatment options. This paper was cited 521 times. At present, DISCERN has become a highly useful tool for the quality evaluation of network health information. Many scholars use this tool to evaluate the quality of online health information about different diseases and published in different languages. For example, Cerminara et al [15] used the DISCERN tool to evaluate the reliability, accuracy, and relevance of the top 50 links for childhood epilepsy (online information) displayed by the Google search engine. Alnaim [16] used the DISCERN tool to assess the quality of information published on websites that share breast cancer information online in Arabic.

Four of the top 10 highly cited articles in the field of CHI research came from the 2004-2008 and 2009-2014 study period (Table 1). The paper “eHealth literacy: Essential Skills for Consumer Health in a Networked World” was published in the *Journal of Medical Internet Research* in 2006. In this article, the concept of eHealth literacy was defined for the first time. The ability to find, discover, understand, and evaluate health information from electronic resources, and the ability to apply the acquired knowledge to solve health problems were also explained in this paper [17]. Ranking fourth in the number of citations, the paper was cited 514 times. eHealth literacy has become a mature research topic, and increasingly more scholars are paying attention to it. For example, Kim et al [18] examined the association among eHealth literacy, perceived benefits, self-efficacy, and health-promoting behaviors in patients with type 2 diabetes. Cherid et al [19] investigated the level of mobile technology acceptance, health literacy, and electronic health literacy of 401 patients aged over 50 years with recent fractures.

Table 1. Top 10 cited articles in the consumer health informatics research field.

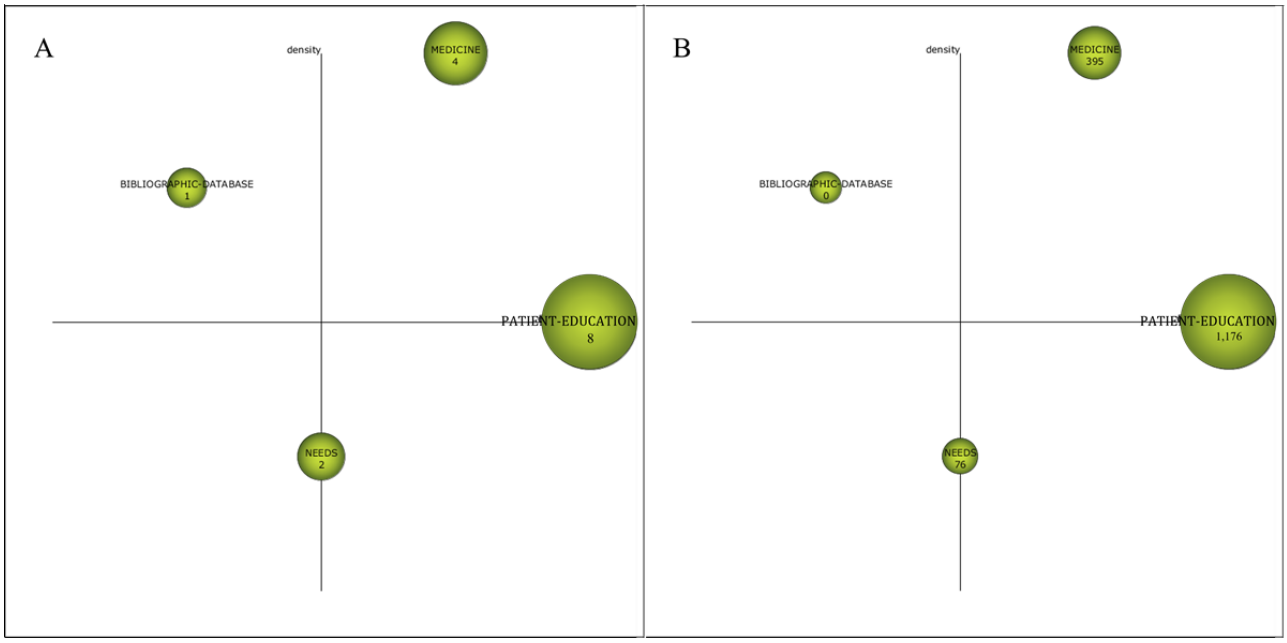
Reference	Journal	Year	Citations
Eysenbach et al [12]	*Journal of the American Medical Association*	2002	940
Cline and Haynes [13]	*Health Education Research*	2001	804
Charnock et al [14]	*Journal of Epidemiology and Community Health*	1999	521
Norman and Skinner [17]	*Journal of Medical Internet Research*	2006	514
Eysenbach [3]	*British Medical Journal*	2000	336
Gustafson et al [20]	*American Journal of Preventive Medicine*	1999	320
Or and Karsh [21]	*Journal of the American Medical Informatics Association*	2009	247
van den Berg et al [22]	*Journal of Medical Internet Research*	2007	236
Dutta-Bergman [23]	*Health Communication*	2004	229
Hibbard and Peters [24]	*Annual Review of Public Health*	2003	227

Evolutionary Characteristics Analysis

Design and Overview

The strategic map is mainly used to describe the relationship between themes and the relationship strength within the theme to reflect the importance of the theme in the development of the whole field and the development of the theme. The node in Figure 4 is a cluster, and the sizes of the nodes indicate the sizes of the theme clusters. The horizontal axis is the centrality, which measures the relevance of the theme to other themes. With higher centrality, the theme is more important in the whole field. The vertical axis is the density, which measures the relational strength of the cluster keywords within a theme. The higher the density of the theme, the more mature the theme. The strategic map is divided into 4 quadrants. Quadrant 1 (Q1) contains motor themes, quadrant 2 (Q2) contains highly developed and isolated themes, quadrant 3 (Q3) contains emerging or declining themes, and quadrant 4 (Q4) contains basic and transversal themes. The strategic map of the 4 study periods, the relevant bibliometrics indicators, and the evolutionary status of the research themes are as follows.

Figure 4. Strategic diagram of themes for the subperiod of 1999-2003 (A: based on the publications, B: based on the citation frequency).



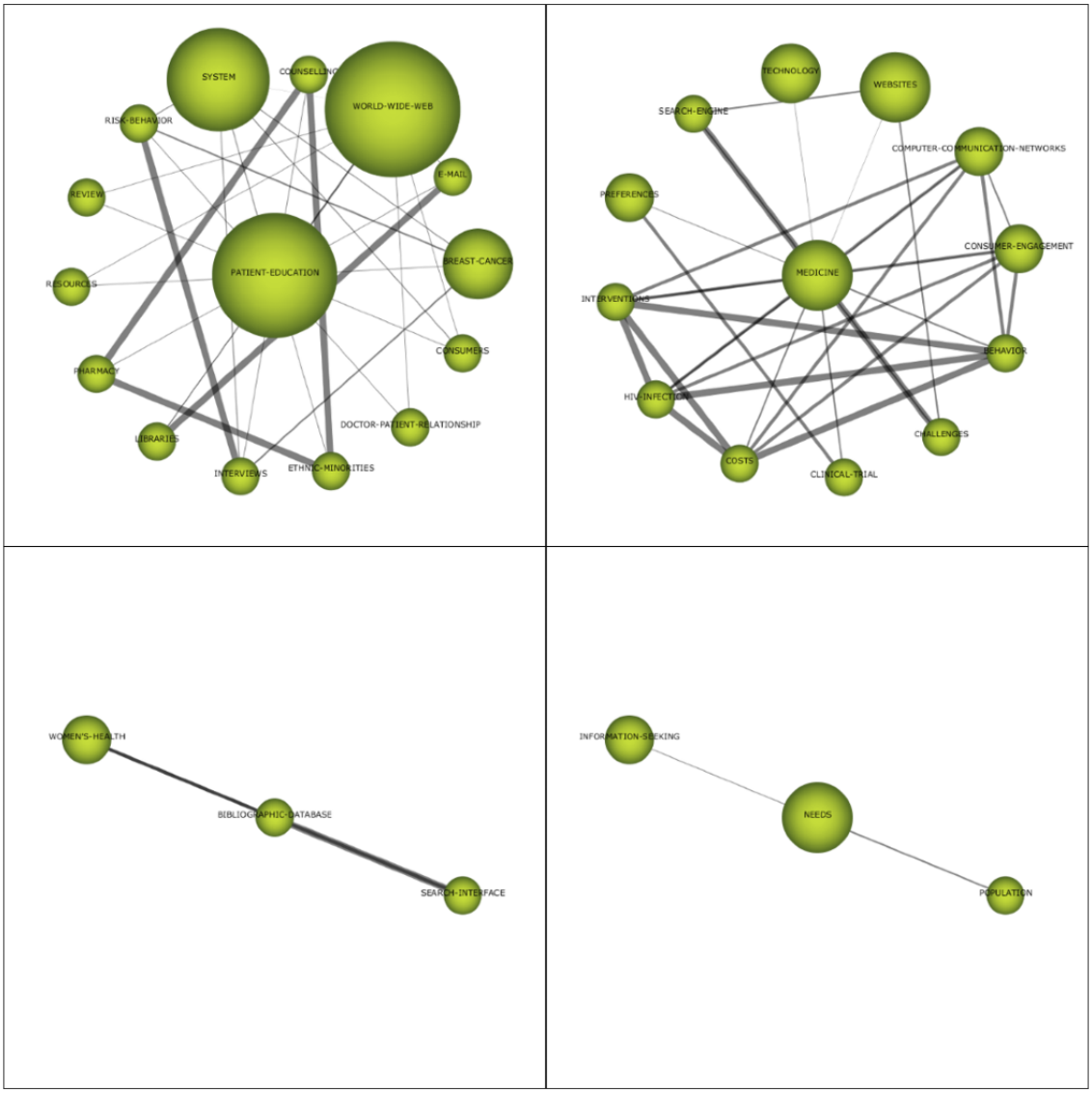
Themes (n=4) in 1999-2003: Patient Education, Medicine, Needs, and Bibliographic Database

The theme of Q1 is *medicine*, which has the highest density value and a higher centrality (Table 2 and Figure 4). The keywords closely connected to *medicine* are *search engine*, *interventions*, and *challenges* (Figure 5). The websites and pages displayed by search engines have become the main source from which people obtain medical and health information [25]. Health information/support systems for health promotion are also emerging, such as the CHESS comprehensive health promotion support system [20].

Table 2. Performance measures for the themes of the subperiod 1999-2003.

Theme	Centrality	Density	Number of documents	H-index	Number of citations
Patient education	45.99	56.26	8	7	1176
Medicine	36.24	111.7	4	3	395
Needs	1.85	12.5	2	2	76
Bibliographic database	1.25	66.67	1	0	0

Figure 5. Cluster network of the themes (1999-2003).



The theme of Q2 is *bibliographic database*, which has centrality and density values of 1.25 and 66.67, respectively. The theme is closely related internally and has low relevance with other themes. The amount of relevant literature is the least, as the theme has not attracted the attention of the academic circle. The keywords closely connected to *bibliographic database* are *search interface* and *women's health*. For the prototype network interface of the women's health bibliography database, Marton et al [26] performed a comparative analysis from 3 aspects: web interface design, overall visual effect, and hypertext navigation and information organization.

Patient education is located on the centrality axis and has the highest centrality. *Patient education* has the largest amount of research literature and has been cited the most times, indicating that it is an area of high concern and has had a great impact on later research. *Patient education* is interconnected with many different keywords, such as world wide web and system, but the connection is not strong. The result of the cluster network shows that the internet and information systems are important tools and approaches in patient education [13] and that individualized customized education can be carried out for patients through computers [27]

Needs is located on the density axis and has centrality and density values of 1.85 and 12.5, respectively. The cluster network shows that *needs* is connected with *population* and *information seeking*. The internal demand for health information drives millions of consumers to search for health information on the internet. If the search terms used by consumers do not match the terms set at the information source, the search results will not meet the needs of consumers. Therefore, consumer health information retrieval needs a full range of terminology support [28].

Themes (n=9) in 2004-2008: World Wide Web, Patient Education, eHealth, Patients, Medication, Terminology, Behavior, Technology, and Disease

The themes of Q1 have higher centrality and density (Table 3 and Figure 6) and are *world wide web*, *patient education*, *eHealth*, and *medication*. *World wide web* has the largest amount of research literature and has been cited 635 times. The cluster network shows that the relationship between the *world wide web* and internal keywords is weak. The keywords that are strongly connected to *patient education* are *internet health information*, *record*, and *challenges* (Figure 7). Patient education is an important part of providing health care services and helps to improve the effect of medical care. Doupi and van der Lei [29] discussed the possibility of integrating electronic medical record data and online health information resources to provide personalized patient education for patients. *eHealth* has the highest density and a relatively high centrality; the cluster network reveals that it is closely related to *health communication*, *tailored intervention*, *cell phone*, and *teleconsultation*, which is an important research topic in this field. Tufano and Karras [30] used teleconsultation for an obesity intervention that was expected to achieve mass customization function, interactive function, and national customized electronic health information. The location of *medication* is far from the centrality axis and close to the density axis. The cluster network shows that it has a strong relationship with *health literacy* and *pharmacist*. In the research on drug treatment in the field of CHI, scholars are concerned about the health literacy of consumers and whether they can understand the accompanying instructions of drugs [31]. Webb et al [32] have proposed that the design of patient-centered consumer medication information can improve the comprehensibility of warning labels.

Table 3. Performance measures for the themes of the subperiod of 2004-2008

Theme	Centrality	Density	Number of documents	H-index	Number of citations
World wide web	77.3	28.66	14	11	635
Patient education	64.37	26.23	10	8	786
eHealth	41.52	87.63	3	3	129
Patients	33.26	16.05	6	5	439
Medication	31.25	32.5	4	4	80
Terminology	3.85	16.67	2	2	45
Behavior	2.38	17.59	3	3	77
Technology	18.49	41.2	2	2	238
Disease	16.39	8.33	2	2	48

Figure 6. Strategic diagram of themes for the subperiod of 2004-2008 (A: based on the publications, B: based on the citation frequency).

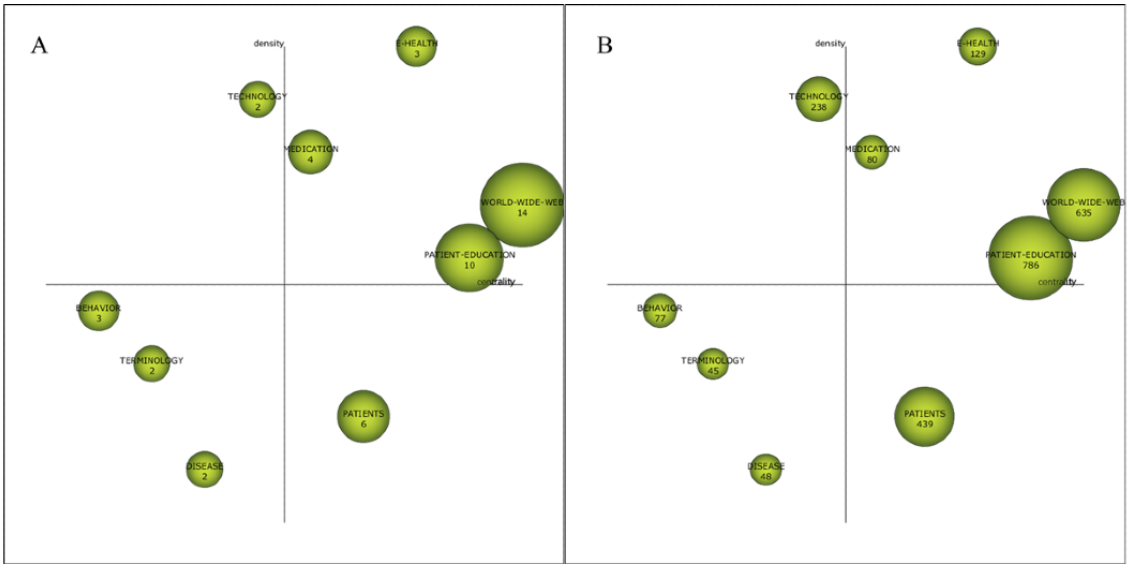
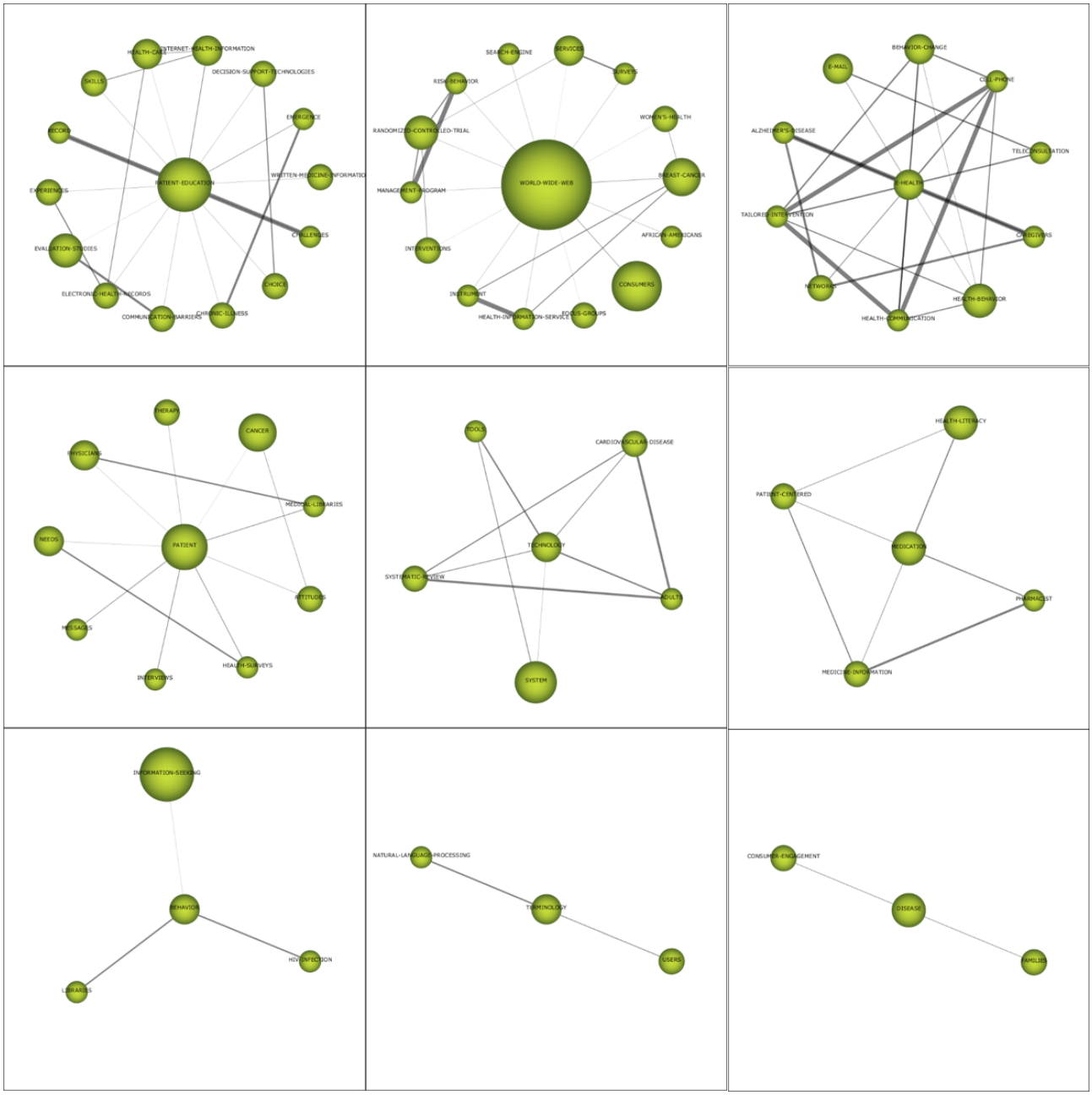


Figure 7. Cluster network of the themes (2004-2008).



The theme of Q2 is *technology*, with centrality and density values of 18.49 and 41.2, respectively. The important nodes that are connected internally are *tools* and *adults*. The literature corresponding to cluster nodes reveals that scholars began using internet technology to intervene in people's health activities [22].

The themes of Q3 are *terminology*, *behavior*, and *disease*, which have centralities of 3.85, 2.38 and 16.39, respectively, and density values of 16.67, 17.59, and 8.33, respectively. *Terminology* is only associated with *natural language processing* and *users*. The literature corresponding to *terminology* reveals that the rapid development of consumer health education websites and other applications has promoted research on consumer health vocabulary and that term *recognition* [33] is one of the important research directions. *Disease* has weak internal relations. *Behavior* is associated with *libraries*, *HIV infection*, and *information seeking*. By studying the literature corresponding to *behavior*, we found that researchers had analyzed health information searching behavior according to different populations. For example, Hesse et al [34] analyzed the 2005 Administration of the Health Information National Trends Survey data to explore the information searching behavior of survivors of cancer and found that their information searching behavior was very common and would not decrease with time.

The theme of Q4 is *patients*, with centrality and density values of 33.26 and 16.05, respectively. The internal correlation of the *patients* theme cluster was not strong. An examination of the literature corresponding to cluster nodes revealed that scholars paid more attention to the research of the patients' attitude, cognition, and intention regarding health information [35].

Themes (n=11) in 2009-2014: Websites, Information Seeking, Physicians, Attitudes, Technology, Risk, Food

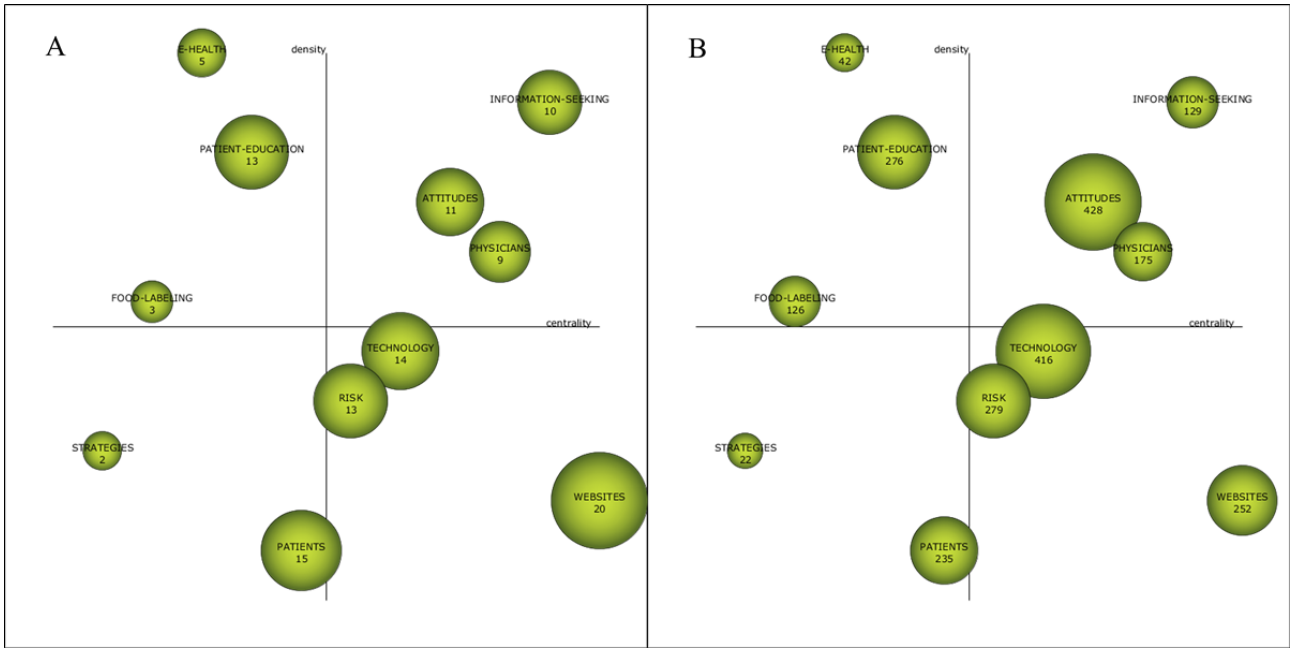
Labeling, Patient, Strategies, Patient Education, and eHealth

The themes of Q1 are *information seeking*, *physicians*, and *attitudes*, which have centralities of 69.64, 67.77, and 64.83, respectively, and density values of 37.42, 19.27, and 28.02, respectively (Table 4 and Figure 8). The cluster network (Multimedia Appendix 1) shows that there are 15 nodes and 25 links in *information seeking* and that the keywords closely related to *information seeking* are *search interface*, *management program*, and *community*. The literature corresponding to *information seeking* revealed that scholars analyzed, from the perspective of consumers, the content of online questions, explored the influencing factors of health information searching, and designed a new connected exploratory navigation interface to improve the effectiveness of health information searching [36]. *Attitudes* had the highest citation frequency, indicating that it had a great influence on the subsequent research. The cluster network shows that *attitudes* is associated with a large number of different keywords, but the relationship is not strong; among these different keywords, the closely related nodes are *community pharmacies*, *information and communication techniques*, and *electronic medical records*. The cluster network reveals that the nodes that are close to *physicians* are *interviews*, *patient centered care*, *patient–provider relationship*, and *teleconsultation*. The literature corresponding to *attitudes* and *physicians* reveals that the consumers' attitude toward, access to, and use of online information related to diseases [37], medicines, diet, and other health issues are the research hotspots in this period. For example, Amicizia et al [38] believes that information and communications technology provide an opportunity for health care workers to use mobile internet to disseminate vaccine-related knowledge interactively and entertainingly and to monitor adolescent attitudes toward vaccination through social media.

Table 4. Performance measures for the themes of the subperiod of 2009-2014.

Theme	Centrality	Density	Number of documents	H-index	Number of citations
Websites	77.12	8.45	20	9	252
Information seeking	69.64	37.42	10	7	129
Physicians	67.77	19.27	9	6	175
Attitudes	64.83	28.02	11	10	428
Technology	61.65	15.53	14	10	416
Risk	60.43	12.61	13	10	279
Food labeling	6.56	15.62	3	3	126
Patient	59.71	6.02	15	10	235
Strategies	5.5	11.11	2	2	22
Patient education	40.51	31.32	13	9	276
eHealth	30.79	37.44	5	3	42

Figure 8. Strategic diagram of themes for the subperiod of 2009-2014 (A: based on the publications, B: based on the citation frequency).



The themes of Q2 are *patient education*, *eHealth*, and *food labeling*, which have centralities of 40.51, 30.79, and 6.56, respectively, and density values of 31.32, 37.44, and 15.62, respectively. The cluster network shows that *patient education* is closely related to *smartphone*, *pamphlet*, *resources*, and *Alzheimer's disease*. Scholars mainly conducted research from 3 aspects: patients' health education needs [39], the effect of health education materials, and health information technology. For example, Kraschnewski et al [40] studied how pregnant women use the internet and mobile phone technology to acquire health care knowledge. They found that women thought that the educational materials they received during prenatal care were not helpful, and therefore, they turned to the internet and smartphone apps to fill the knowledge gap [40]. There are 9 nodes and 12 connections in the *eHealth* theme. Compared with the previous study period, in this study period, there are fewer nodes and connections. The research content of the theme is more concentrated. The keywords closely related to *eHealth* are *tailored intervention*, and *online health information*. Scholars are concerned about how to use information and communication technology to help patients improve their health status and strengthen their self-health management skills [41]. For example, as a health literacy intervention for Hispanic patients with AIDS, Jacobs et al [42] developed a Spanish-language computer application, which can provide information about HIV infection, treatment, and drugs. *Food labeling* is a new theme, for which the cluster network shows the related keywords are *consumer opinions*, *recommendations*, and *ratings*. King et al [43] investigated the consumers' understanding of common terms used to guide food consumption frequency and quantity and found that some terms are highly subjective and that more simple and clear terms need to be developed.

The themes of Q3 are *patients* and *strategies*, which have lower centralities and density values. The cluster network reveals that the keywords strongly related to *patients* were *health information service*, *information behavior*, and *conflict of interest*. *Strategies* is an emerging theme. The cluster network shows that the keywords associated with it are only *evidence-based practice* and *task* and that the weight of the connection is 0.17.

Having higher centralities and lower density values, the themes of Q4 are *websites*, *technology*, and *risk*. The cluster network shows that the keywords closely related to *websites* are *world wide web*, *evaluation studies*, *tools*, *email*. By studying the literature corresponding to *websites*, it is found that the research on the content evaluation of websites and health information intervention through websites are the research hotspots in this period [44]. Thakor et al [45] investigated the information quality of ecommerce websites selling *Hypericum perforatum* and found that most sites received poor reviews and lacked information on drug interactions, contraindications, and adverse reactions. The keywords closely related to *technology* are *hometelehealth* and *aging*. With the aggravation of the aging of the population, the research on home telehealth for the aged group has attracted much attention. Cimperman et al [46] studied the important factors affecting the elderly's adoption of home telehealth services and found that the following factors play an important role in the perception of home telehealth: perceived usefulness, expectation, social influence, perceived security, computer anxiety, convenience, and the doctor's opinion [46,47]. *Risk* is split from *medication*. The cluster network shows that the relationship between internal keywords is weak and that the relatively strong ones are *costs*, *patient decision aid*, and *cardiovascular disease*. For veterans with and without multiple sclerosis, Cameron et al [48] studied the relative risk of falling and requiring medical care and found that the adjusted odds ratio of falling was 3 times higher for female veterans with multiple sclerosis than for female veterans without multiple sclerosis.

Themes (n=14) in 2015-2019: Electronic Medical Records, Health Information Seeking, Attitudes, Health Communication, Breast Cancer, Health Literacy, Technology, Natural Language Processing, User-Centered Design, Pharmacy, Academic Libraries, Costs, Internet Utilization, and Online Health Information

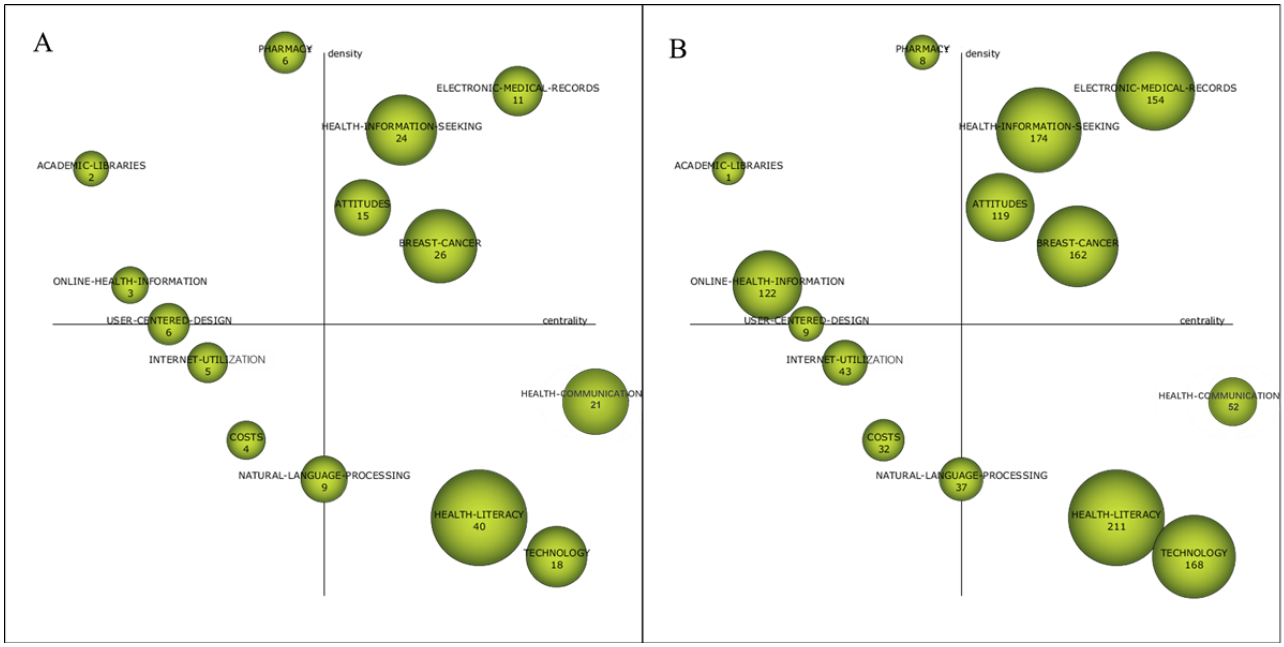
The themes of Q1 are *electronic medical records*, *health information seeking*, *attitudes*, and *breast cancer*, with centralities of 53.52, 47.78, 45.18, and 48.08, respectively, and density values of 19.71, 18.14, 8.89, and 7.11, respectively (Table 5 and Figure 9). The cluster network (Multimedia Appendix 2) shows that *electronic medical records* has a strong relationship with internal keywords, which include *personal health information management* and *care partner*. Scholars mainly focus on the patients' health information management, including the sharing of access to the patients' health records, personal health data visualization, and other contents. For example, Wolff et al [49] sent the doctor's medical records to patients and authorized nurses through OpenNotes. The scholars verified the acceptability and effect of this approach, and the results showed that it was acceptable for patients and their nursing staff to view the doctors' medical records; the results also reflected that this method improved the communication with patients and enhanced the patients' confidence in cooperating with nurses. Keywords closely related to *healthinformation seeking* are *users*, *behavior*, *health information needs*, and *consumer health information behavior*. An examination of the literature corresponding to important nodes revealed that the users of library, online health consultation platform, social media, and other media were often considered research objects and that a discussion of these users' health information needs and behaviors was a research hotspot in this period [50,51]. For example, taking the online health consultation case on "Taiwan eDoctor," an online health consultation platform, as the research object, Chiu et al [52] studied the length, time, communication mode, purpose, and identity disclosure of online consultation questions and described the communication mode of patients in the process of health information searching. *Attitudes* was closely related to internal keywords, such as *HPV vaccines*, *risk*, *evaluation studies*, and *physicians*. Systematically studying the rankings, quality, and contents of the web pages related to HPV vaccines, Fu et al [53] classified the included web pages into critical and noncritical ones and found that the quality of the critical web pages was poor but that the critical web pages often obtained higher rankings. Keywords closely related to *breast cancer* are *online*

healthcommunities, *support groups*, and *preferences*. The scholars mainly focused on the content analysis of the online health community [54] and its supporting role for patients with cancer. For example, through interviews and questionnaires, Huh et al [55] developed online health community roles to reflect the users' needs and requirements when using an online health community. These roles can help provide users with customized social support and patient support. The study found that the roles of online health communities can be divided into 4 categories: managers, opportunists, scientists, and adventurers. These roles reveal user interaction behavior and attitude patterns when using online health communities.

Table 5. Performance measures for the themes of the subperiod 2015-2019.

Theme	Centrality	Density	Number of documents	H-index	Number of citations
Electronic medical records	53.52	19.71	11	7	154
Health information seeking	47.78	18.14	24	5	174
Attitudes	45.18	8.89	15	7	119
Health communication	57.98	4.7	21	4	52
Breast cancer	48.08	7.11	26	7	162
Health literacy	50.29	3.55	40	8	211
Technology	54.79	3.46	18	7	168
Natural language processing	24.1	4.18	9	4	37
User-centered design	11.9	6.3	6	2	9
Pharmacy	16.52	28.49	6	2	8
Academic libraries	4.52	9.03	2	1	1
Costs	14.08	4.34	4	4	32
Internet utilization	12.67	4.72	5	3	43
Online health information	7.82	6.49	3	2	122

Figure 9. Strategic diagram of themes for the subperiod of 2015-2019 (A: based on the publications, B: based on the citation frequency).



The themes of Q2 are *pharmacy*, *online health information*, and *academic libraries*, with centralities of 16.52, 4.52, and 7.82, respectively, and density values of 28.49, 9.03, and 6.49, respectively. A mature professional theme (*pharmacy*) has the highest density and is close to the density axis. The keyword most closely related to *pharmacy* is *patient medication information*. Monkman and Kushniruk [56] studied consumer medication information in pharmacies and found that the organization and presentation of online consumer medication information need to be improved. The content of consumer medication information needs to be improved in order to promote the safety and effective use of drugs [57]. The keywords in the *online health information* cluster network include *video*, *web* 2.0, and *world wide web*. Scholars mainly pay close attention to the use and quality evaluation of online health information [58]. As an emerging theme, *academic libraries*

has the lowest centrality, the lowest amount of relevant literature, and the lowest citation frequency.

User-centered design, with centrality and density values of 11.9 and 6.3, respectively, is located on the centrality axis. One of the keywords clustering with *user-centered design* is *smartphone apps*. An examination of the key literature corresponding to cluster networks reveals that scholars focus on designing user-centered, smartphone-based mobile health apps, such as nutrition education apps that, for example, provide technical support for adolescents with overweight and obesity [59].

Having lower centralities and density values, the themes of Q3 are *internet utilization* and *costs*. The internal relationship between *internet utilization* and *costs* is weak; both are in an immature development stage and have little impact. The cluster network shows that the keywords in *costs* include *search engine* and *health records*; the keywords in *internet utilization* include *health behavior*, *systematic review*, and *patient education as topic*. The literature corresponding to *internet utilization* reveals that the content of relevant literature involved patients using the internet to search for health information for self-care. For example, Jamal et al [60] studied the online health information searching behavior of patients with type 2 diabetes in the Middle East and its influence on their self-care behavior. The results showed that most internet health information searchers had positive changes in their behaviors after searching and had a stronger awareness of diabetes self-care [60].

Natural language processing is located on the density axis, with centrality and density values of 24.1 and 4.18, respectively. The cluster network shows that *natural language processing* is connected with *consumer health vocabulary*, *state*, *online support groups*, and *text*. An examination of the core literature corresponding to the cluster shows that most scholars mined and dealt with the health-related contents of consumers through social media, health websites, health forums, and other ways to obtain the users' health needs, explore the users' health behaviors, and investigate the application of health information technology [61,62].

The themes of Q4 are *health communication*, *health literacy*, and *technology*, with centralities of 57.98, 50.29, and 54.79, respectively, and density values of 4.7, 3.55, and 3.46, respectively. The *technology* cluster network formed 15 nodes and 21 wires, but the internal relationship was very weak. *Health communication* has the highest centrality and has many connections with external themes. It is a key theme with development potential. The cluster network shows that *health communication* is strongly related to keywords such as *medicine*, *telehealth*, and *perspectives*. Benis et al [63] studied the use of communication channels between patients and health care organizations. By matching the communication channels with the patients' personal information, the communication between medical institutions and patients can be transformed into a more active mode, and the patient's participation can be improved. *Health literacy* is an emerging theme; it has the largest number of related literature and the largest citation frequency sum. The keywords closely related to *health literacy* include *skills*, *patient education*, and *public libraries*. Health literacy is the ability of individuals to access and understand health information and to use it to maintain and promote their health. An examination of the literature of important nodes reveals that scholars pay more attention to the assessment and intervention of health literacy [64] and the assessment of internet health information quality [65,66].

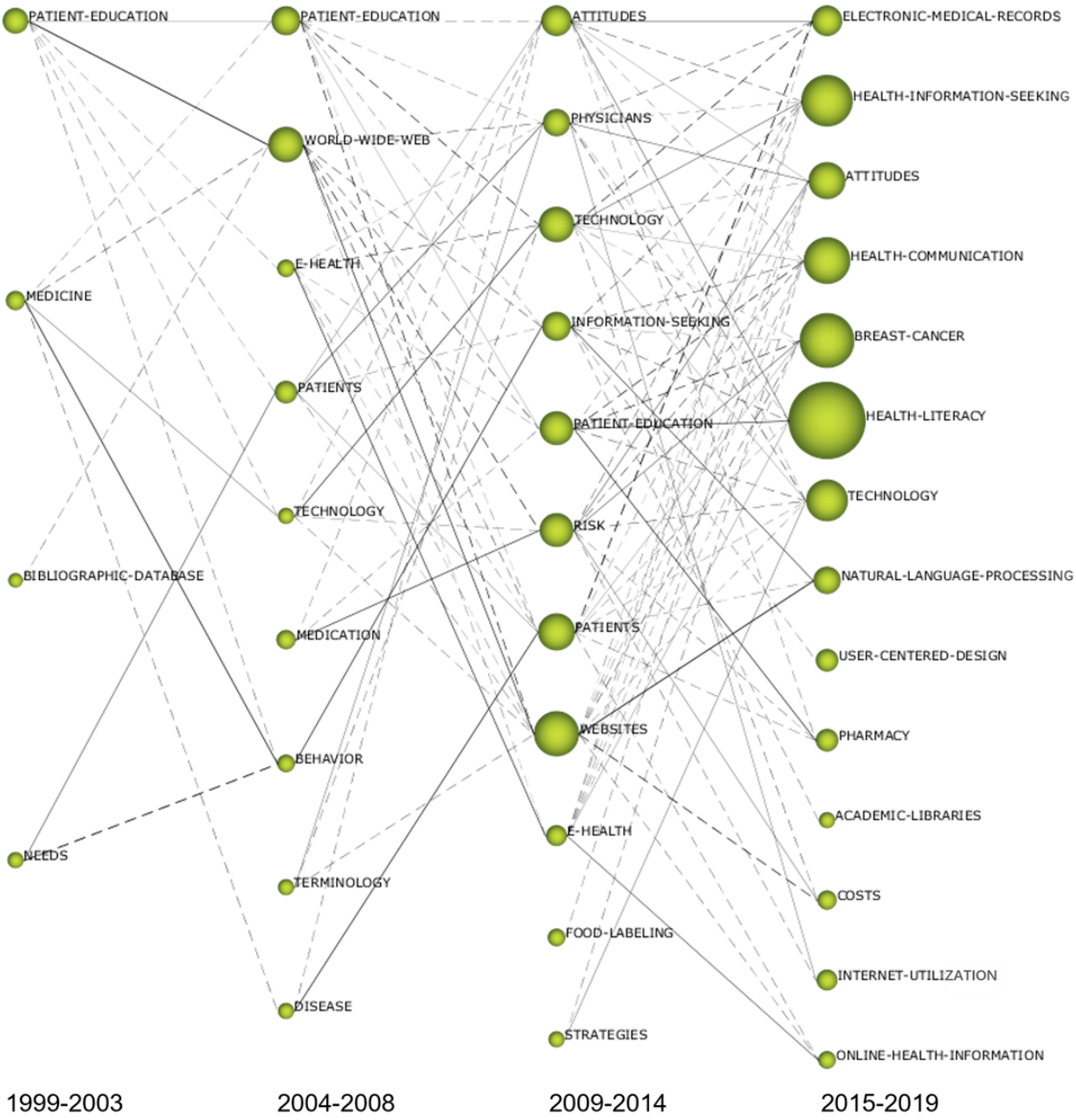
Evolutionary Path Analysis

Design and Overview

The theme evolution map shows, through a data flow, the evolution of the 4 study period themes of CHI research. It can analyze and track the dynamic evolution of the themes in the field of CHI research over a period. Figure 10 shows the CHI research field's theme evolution map. Each column in the figure represents a study period. In the figure, the nodes represent the research themes, and the size of the nodes is proportional to the number of related articles of the research themes. The connection between the nodes represents the theme data flow, the solid line represents the main keywords shared by 2 themes, and the dotted line represents the shared keywords are not the main keywords. The width of the connection is proportional to the Inclusion index [11].

From Figure 10, it can be seen intuitively that *patient education* and *technology* appear in 3 study periods, and that *eHealth*, *attitudes*, and *patients* appear in 2 study periods. With time, the circle in the theme evolution map becomes larger, the number of the research themes increases, and the data flow between the research themes becomes increasingly complex. This reflects the emergence of new research themes and contents from 1999 to 2019. According to the theme evolution map, the size of the research theme circle, and the data flow between themes, 10 theme evolution paths of 3 research directions in the field of CHI were determined.

Figure 10. Thematic evaluation map of the consumer health informatics research field (1999-2019).



Supply-Side Research of Consumer Health Informatics: Patient Education and Intervention Research

Consumer health education is the process of assisting one to acquire the correct information and understanding so that one will be able to make wise decisions about a certain health item [67]. *Patient education → patient education → patient education → health literacy*, *pharmacy* (subpath 1). *Patient education → world wide web → websites → health communication* (subpath 2). *Patient education → eHealth → eHealth → health literacy*, *breast cancer*, *health communication* (subpath 3).

The 3 subpaths above evolved from the main and branch directions of the *patient education* theme in the first study period. Over time, the nodes on the evolutionary path become larger, the number of research articles increases, and the data flow from node splitting is greater. The data flow of patient education and intervention research evolution is relatively clear and represents the main path of the evolution theme of CHI. The theme on the path mainly moves between the first quadrant and the fourth quadrant, and the split and fusion of themes are obvious. In the second study period, subpaths 1 and 3 became the focus of research. In the third study period, the theme density increased and the centrality decreased, demonstrating a professional and mature research direction in the field. In the fourth study period, the themes split into several themes with great influence and development potential (*health communication* and *health literacy*). In the third study period, the subpath 2 integrated multiple themes, and the density of themes decreased. In the fourth study period, the theme split

into several low-density but high centrality themes (*natural language processing*).

Consumer-Side Research of Consumer Health Informatics: Consumer Demand, Attitude, and Behavior

The evolutionary path of consumer demand, attitude, and behavior research consists of 3 subparts. *Needs → patients → attitudes*, *physicians → attitudes*, *health literacy*, *healthinformation seeking*, *internet utilization* (subpath 1); *medicine*, *needs → behavior → information seeking → healthinformation seeking*, *health literacy* (subpath 2); and *medicine → diseases → patients → electronic medical records* (subpath 3).

These 3 subpaths evolved from the main and branch directions of *medicine* and *needs* in the first study period. With time, the number of nodes on the evolutionary path increases, the number of research articles increases year by year, and the data flow of node splitting also increases. The themes in the evolutionary path of consumer demand, attitude, and behavior research mainly move between the first, third, and fourth quadrants, and the themes are divided and integrated. The theme of subpath 1 has gone through a stage in which centrality and density rise, and in this field, it finally forms many core themes, such as *healthinformation seeking*, *health literacy*, and *attitudes*. In the second study period, subpath 2 splits to form a new theme (*behavior*). Later, this theme attracted the attention of many scholars in the field; its density and centrality increased rapidly, and it became the core theme in the field. In the second and third study periods, subpath 3 is an edge theme that is not of high concern. In the fourth study period, some themes (*attitudes*, *physicians*, *technology*, *patients*, *websites*) split and merge into a high centrality and density key theme (*electronic medical records*).

The Technology Research of Consumer Health Informatics: The Application Research of Internet and Information Technology

Currently, big data, internet plus, artificial intelligence, and other emerging information technologies have been tightly integrated with traditional medical and health industries. The application of internet and information technology in the field of CHI has attracted more attention from computer science and communication science scholars. The evolution path of internet and information technology application research consists of 4 subpaths. *Medicine → technology → technology → technology*, *health communication* (subpath 1); *patient education → eHealth → eHealth → online health information* (subpath 2); *needs*, *bibliographic database → patients*, *world wide web → information seeking → natural language processing*, *user-centered design*, *academic libraries* (subpath 3); and *medicine → world wide web*, *terminology → websites → online health information*, *natural language processing*, *breast cancer* (subpath 4).

These 4 subpaths evolved from the main directions of *bibliographic database* in the first study period as well as the branch directions of *patient education* and *medicine*. The number of research articles is gradually increasing, and the phenomenon of node splitting and fusion is also obvious. The theme of the evolution path of the internet and information technology application research mainly moves between the first, second, and fourth quadrants. The evolution path is relatively complex, which may be related to the application of information technology in different research directions. Subpath 1 emerged from the theme of the first study period, and the density decreased in the second study period. In the third and fourth study periods, the centrality increased. The density decreased after the split of multiple themes, and some important themes with strong influence and development potential developed. Subpath 2 divides from the theme of *patient education* in the first study period. In the second study period, a key theme with high centrality and density was formed. In the third and fourth study periods, the centrality decreased, and the density increased; it became a professional theme in the field. The data flow of subpaths 3 and 4 is relatively complex and is formed by the convergence of 2 routes in the first and second study period.

Discussion

Overall Development Status of Consumer Health Informatics

In the past 21 years, CHI research has been on the rise and has gone through 3 stages: slow growth, stable growth, and rapid growth. In the early stage, the development was slow, the themes were few, and the intensity was low. In the later stages, the development became diversified. After 2015, the number of research papers on CHI increased substantially, and it became a hotspot of academic research. In 2019, it reached a peak of 112 papers, a research volume accounting for 12.03% of the total literature.

The period 1999-2003 comprises the basic research stage of consumer health information. Six of the top 10 highly cited literature in this field come from this study period. This literature constitutes the cornerstone of CHI research. Consumer health information evaluation tools, such as DISCERN, that were developed in this study period have become recognized mature tools in the field. The period 2004-2014 was one of steady growth in the research on CHI. During this period, fewer keywords disappeared, and more new ones were absorbed. In 2015-2019, the number of keywords was the largest, and the research content involved was increasingly more abundant. A similarity index for the fourth study period reached a higher level, reflecting that the research field of CHI had become more mature, the inheritance of research had become stronger, and researchers were continuing to focus on more research topics.

Evolutionary Characteristics Analysis

By analyzing the rise, decline, and the change in the density and centrality of the themes in different study periods, the evolution characteristics and status of the research themes can be reflected.

The motor themes include the following: *medicine* and *patient education* in 1999-2003; *world wide web*, *patient education*, *eHealth*, and *medication* in 2004-2008; *information seeking*, *physicians*, and *attitudes* in 2009-2014; and *electronic medical records*, *healthinformation seeking*, *attitudes*, and *breast cancer*

in 2015-2019. These themes have high centralities and density values. Besides, the research studies are popular and influential, and the internal relations within these themes are closely relevant. These themes are the core of the research field and their development is mature.

The highly developed and isolated themes include the following: *bibliographic database* in 1999-2003; *technology* in 2004-2008; *eHealth*, *patient education*, and *food labeling* in 2009-2014; and *pharmacy*, *academic libraries*, *online health information*, and *user-centered design* in 2015-2019. Having high-density values and low centralities, these themes do not represent the research center. They are mature and peripheral themes in this field.

The emerging or declining themes include the following: *behavior*, *terminology*, and *disease* in 2004-2008; *patient* and *strategies* in 2009-2014; and *costs* and *internet utilization* in 2015-2019. Because of their low centralities and density values, these themes have been weakly developed and are marginalized themes in this field.

The basic and transversal themes include the following: *needs* in 1999-2003; *patients* in 2004-2008; *websites*, *technology*, and *risk* in 2009-2014; and *health communication*, *breast cancer*, *health literacy*, *technology*, and *natural language processing* in 2015-2019. These themes have high centralities and low-density values, indicating that these have a high influence on the field of CHI but are not well developed.

By analyzing the evolutionary status of the themes in the 2015-2019 study period, the paper expects to predict the future development trend of the field. The centrality and density value of *electronic medical records* are extremely high. Although there are not many related documents, each document has been cited 14 times on average, indicating that it has received extremely high attention and may appear in the next research period. *Healthinformation seeking* is relatively stable because of its high centrality and density value. After an internal keyword splitting and recombination, the *attitudes* theme is still high in centrality and density value, and the research enthusiasm has not decreased. It will continue to become a hot topic in the field of CHI research. *Breast cancer* split from *risk* in the previous study period, with increased centrality and density value (transferred from Q4 to Q1). With the expansion of research influence, it became more mature and eventually became the research core in the field. Although the number of studies related to *online health information* is not large, the average citation frequency of each study is the highest, indicating that it has a great influence on future research. Online health information is one of the foundations of the CHI research field, has great development potential, and may continue to appear in the next study period. *Natural language processing* has been a hot topic in recent years, and it is likely to continue to appear in the next study period. *Health literacy* is an emerging topic, with the largest number of relevant studies and total cited frequency. This indicates that it has a high level of attention and has a strong evolution and development ability. It will continue to be a hot topic in this field. *Technology* has been transferred from Q2 to Q4. Its research influence and development potential have improved, reflecting that it will continue to be the basic theme in this field.

Evolutionary Path Analysis

By analyzing the evolution characteristics of patient education and intervention research themes, it can be found that the research in the slow growth period mainly focuses on the application and exploration of the internet and information systems and explores the health education for patients through computers and websites [13]. The research in the stable growth period mainly focuses on the use of network health information [41], the design of personalized or customized patient education [29], and the application of intelligent and mobile devices [30]. The research in the rapid growth period mainly focuses on the role and mechanism of the online health community in patient support [54], doctor–patient communication [63], and the quality evaluation of online health information [66].

By analyzing the evolution characteristics of consumer demand, attitude, and behavior research themes, it can be found that the research in the slow growth period mainly focuses on the users' demand for online health information search, as well as demand-driven behavioral research, such as an analysis of the characteristics of health information–retrieval terms [28]. Studies in the stable growth period mainly focus on the patients' attitude and cognition toward health information [23], health information searching behavior, and its influencing factors [68]. The research in the rapid growth period mainly focuses on personal health information management [69], health literacy assessment, and health literacy intervention [70]. The internet health information search behavior of patients with different diseases is still the research hotspot in this period.

By analyzing the evolution characteristics of the research themes regarding the internet and information technology application, it can be found that the research in the slow growth period mainly focuses on the technology applied to patient education [71] and the development of a bibliographic database network interface [26]. The studies in the stable growth period mainly focus on the use of the internet and information technology to intervene in people's health activities [72], help improve the patients' health status, and strengthen the patients' self-care skills [73]. They also examined the use of information technology in the patients' health information search behavior [74], identifying text difficulty of health information and classifying health care webpages using machine learning and natural language processing methods [75,76]. The research in the rapid growth period mainly focuses on the application of natural language processing [77] and the development and evaluation of user-centered mobile medical apps [78]. Wongchaisuwat et al [79] developed an algorithm to automatically answer health-related question by implementing a semisupervised learning algorithm. Park et al [80] analyzed the online discussion content of 3 online health communities employing text mining and k-means machine learning algorithm to compare the discussion topics.

Limitations

In this paper, we used the SciMAT tool to analyze the keywords of the literature data and constructed the overlapping map,

evolution map, strategic diagram, and a cluster network. Owing to the limitation of the scope of our school's database, we only obtained the literature from 1999 to 2019, failed to trace the earliest research literature on CHI, and failed to build a complete evolution path of the CHI research theme.

Conclusions

In this research, a bibliometric analysis was carried out to explore the dynamic evolution path and evolution laws of CHI research themes on a time dimension and from the perspective of strategic diagrams and data flows. CHI research focuses on themes, such as the patients' education, health information needs, information search behavior, health behavior intervention, health literacy, health information technology, and eHealth. The research content in 4 different study periods formed the 38 themes. These themes formed 10 evolution paths in 3 research directions: patient education and intervention, consumer demand attitude and behavior, and internet information technology application. Patient education and intervention research, consumer demand, attitude, and behavior research comprise the main theme evolution path. The path's evolution process has been relatively stable, and it will continue to represent the research hotspot in this field. Research on the internet and information technology application is a secondary theme evolution path, where obvious absorption, fragmentation, and extinction of themes have occurred. Its complex evolution process has also attracted the attention of many scholars. The research status and influence of this evolution path have gradually increased, making it a research direction with development potential.

According to the existing literature, themes that will continue to appear in the next study period include the following: *electronic medical records*, *online health information*, *health information seeking*, *attitudes*, *health literacy*, *technology*, and *natural language processing*. The first 2 themes are information resource elements *in* the consumer health information ecosystem. Because of the continuous advancement of medical informatization and the rapid development of social media, electronic health records, online health information, and online health question answering data have formed massive health information repositories. These health information big data are like gold mines waiting to be explored by researchers. The middle 3 themes are information subject elements, and the last 2 themes are the information technical elements in the environment. To maintain the sustainable development of the consumer health information ecosystem, it is necessary to start from the supply side and the demand side of health information and to solve the mismatch between service and demand by using computer and information technology.

Acknowledgments

We thank the State Key Program of National Social Science of China (grant no. 17AZD037) for assistance in this study. We also appreciate the support of all participants.

Authors' Contributions

AL and WX co-designed this research. WO, ZX, and HH conducted the data retrieval/construction, data collection, and analysis. YY, FL, and YC helped preprocess and analyze the data. WO interpreted the results and drafted the manuscript. TW, XP, and PD made major revisions. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Cluster network of the themes (2009-2014).
[PNG File , 5060 KB - jmir_v23i9e21974_app1.png]

Multimedia Appendix 2
Cluster network of the themes (2015-2019).
[PNG File , 4909 KB - jmir_v23i9e21974_app2.png]

References

1. Thornton K. Directions and Opportunities in Health Informatics in British Columbia. Methods Inf Med 1994 Jun;33(3):299-301. [doi: 10.1055/s-0038-1635019]
2. Ferguson T. Consumer health informatics. Healthc Forum J 1995;38(1):28-33. [Medline: 10154286]
3. Eysenbach G. Consumer health informatics. BMJ 2000 Jun 24;320(7251):1713-1716 [FREE Full text] [doi: 10.1136/bmj.320.7251.1713] [Medline: 10864552]
4. American Medical Informatics Association. Consumer Health Informatics. URL: https://www.amia.org/applications-informatics/consumer-health-informatics [accessed 2020-06-08]
5. Kokol P, Saranto K, Vosner HB. eHealth and health informatics competences: A systemic analysis of literature production based on bibliometrics. Kybernetes 2018 Mar 05;47(5):1018-1030. [doi: 10.1108/k-09-2017-0338]

6. Zhao YH, Zhang J. Consumer health information seeking in social media: a literature review. Health Information and Libraries Journal 2017 Oct 17;34(4):268-283. [doi: 10.1111/hir.12192]
7. Xie HL, Zhang YW, Duan KF. Evolutionary overview of urban expansion based on bibliometric analysis in Web of Science from 1990 to 2019. Habitat International 2020 Jan;95:102100. [doi: 10.1016/j.habitatint.2019.102100]
8. Guo YQ, Hao ZC, Zhao SC, Gong JQ, Yang F. Artificial Intelligence in Health Care: Bibliometric Analysis. Journal of Medical Internet Research 2020 Jul 29;22(7):e18228. [doi: 10.2196/18228]
9. Wang J, Deng H, Liu BT, Hu AB, Liang J, Fan LY, et al. Systematic Evaluation of Research Progress on Natural Language Processing in Medicine Over the Past 20 Years: Bibliometric Study on PubMed. Journal of Medical Internet Research 2020 Jan 23;22(1):e16816. [doi: 10.2196/16816]
10. Yonoff RE, Ochoa GV, Cardenas-Escorcia Y, Silva-Ortega JI, Merino-Stand L. Research trends in proton exchange membrane fuel cells during 2008-2018: A bibliometric analysis. Heliyon 2019 May;5(5):e01724. [doi: 10.1016/j.heliyon.2020.e04848]
11. Cobo MJ, López-Herrera AG, Herrera-Viedma E, Herrera F. SciMAT: A new science mapping analysis software tool. Journal of the American Society for Information Science and Technology 2012 Jul 02;63(8):1609-1630. [doi: 10.1002/asi.22688]
12. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the World Wide Web - A systematic review. JAMA-Journal of the American Medical Association 2002 May 22;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691]
13. Cline RJW, Haynes KM. Consumer health information seeking on the Internet: the state of the art. Health Education Research 2001 Dec;16(6):671-692. [doi: 10.1093/her/16.6.671]
14. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. Journal of Epidemiology and Community Health 1999 Feb 01;53(2):105-111. [doi: 10.1136/jech.53.2.105]
15. Cerminara C, Santarone ME, Casarelli L, Curatolo P, El Malhany N. Use of the DISCERN tool for evaluating web searches in childhood epilepsy. Epilepsy & Behavior 2014 Dec;41:119-121. [doi: 10.1016/j.yebeh.2014.09.053]
16. Alnaim L. Evaluation Breast Cancer Information on The Internet in Arabic. Journal of Cancer Education 2019 Aug;34(4):810-818. [doi: 10.1007/s13187-018-1378-9]
17. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. Journal of Medical Internet Research 2006 Jun 16;8(2):e9. [doi: 10.2196/jmir.8.2.e9]
18. Kim KA, Kim YJ, Choi M. Association of Electronic Health Literacy With Health-Promoting Behaviors in Patients With Type 2 Diabetes: A Cross-sectional Study. Cin-Computers Informatics Nursing 2018 Sep;36(9):438-447. [doi: 10.1097/cin.0000000000000438]
19. Cherid C, Baghdadli A, Wall M, Mayo NE, Berry G, Harvey EJ, et al. Current level of technology use, health and eHealth literacy in older Canadians with a recent fracture—a survey in orthopedic clinics. Osteoporosis International 2020 Feb 28;31(7):1333-1340. [doi: 10.1007/s00198-020-05359-3]
20. Gustafson DH, Hawkins R, Boberg E, Pingree S, Serlin RE, Graziano F, et al. Impact of a patient-centered, computer-based health information/support system. American Journal of Preventive Medicine 1999;16(1):1-9. [doi: 10.1016/s0749-3797(98)00108-1]
21. Or CKL, Karsh BT. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]
22. van den Berg MH, Schoones JW, Vlieland TPMV. Internet-Based Physical Activity Interventions: A Systematic Review of the Literature. Journal of Medical Internet Research 2007 Sep 30;9(3):e26. [doi: 10.2196/jmir.9.3.e26]
23. Dutta-Bergman MJ. Primary Sources of Health Information: Comparisons in the Domain of Health Attitudes, Health Cognitions, and Health Behaviors. Health Communication 2004 Jul;16(3):273-288. [doi: 10.1207/s15327027hc1603_1]
24. Hibbard JH, Peters E. Supporting informed consumer health care decisions: data presentation approaches that facilitate the use of information in choice. Annu Rev Public Health 2003;24:413-433. [doi: 10.1146/annurev.publhealth.24.100901.141005] [Medline: 12428034]
25. Wu JS, Li YC, Jian WS. Consumer health information websites in Taiwan. Journal of the Formosan Medical Association 2000 Aug;99(8):663-666. [Medline: 10969514]
26. Marton CF, Kalsbeek K, Thulin K. Evaluating the Web user interface design of a bibliographic database on women's health. 2003 Jul Presented at: 7th World Multiconference on Systemics, Cybernetics and Informatics; Jul 27-30, 2003; Orlando, fl p. 456-461.
27. Lewis D. Computers in Patient Education. Cin-Computers Informatics Nursing 2003 Mar;21(2):88-96. [doi: 10.1097/00024665-200303000-00010]
28. Zeng Q, Kogan S, Ash N, Greenes RA, Boxwala AA. Characteristics of Consumer Terminology for Health Information Retrieval. Methods of Information in Medicine 2002;41(4):289-298. [doi: 10.1055/s-0038-1634490]
29. Doupi P, van der Lei J. Design and implementation considerations for a personalized patient education system in burn care. International Journal of Medical Informatics 2005 Mar;74(2-4):151-157. [doi: 10.1016/j.ijmedinf.2004.04.021]

30. Tufano JT, Karras BT. Mobile eHealth Interventions for Obesity: A Timely Opportunity to Leverage Convergence Trends. Journal of Medical Internet Research 2005;7(5):97-104. [doi: 10.2196/jmir.7.5.e58]
31. Wallace LS, Keenum AJ, Roskos SE, Blake GH, Colwell ST, Weiss BD. Suitability and readability of consumer medical information accompanying prescription medication samples. Patient Education and Counseling 2008 Mar;70(3):420-425. [doi: 10.1016/j.pec.2007.11.017]
32. Webb J, Davis TC, Bernadella P, Clayman ML, Parker RM, Adler D, et al. Patient-centered approach for improving prescription drug warning labels. Patient Education and Counseling 2008 Sep;72(3):443-449. [doi: 10.1016/j.pec.2008.05.019]
33. Zeng QT, Divita G, Keselman A, Crowell J, Browne AC, Goryachev S, et al. Term Identification Methods for Consumer Health Vocabulary Development. Journal of Medical Internet Research 2007 Mar 14;9(1):e4. [doi: 10.2196/jmir.9.1.e4]
34. Hesse BK, Arora NK, Beckjord EB, Rutten LJF. Information support for cancer survivors. Cancer 2008 Jun 01;112(11):2529-2540. [doi: 10.1002/cncr.23445]
35. McInnes DK, Cleary PD, Stein KD, Ding L, Mehta CC, Ayanian JZ. Perceptions of cancer-related information among cancer survivors: A report from the American Cancer Society's Studies of Cancer Survivors. Cancer 2008 Sep 15;113(6):1471-1479. [doi: 10.1002/cncr.23713]
36. Cui LC, Carter R, Zhang GQ. Evaluation of a Novel Conjunctive Exploratory Navigation Interface for Consumer Health Information: A Crowdsourced Comparative Study. Journal of Medical Internet Research 2014 Feb 10;16(2):e45. [doi: 10.2196/jmir.3111]
37. Vodicka E, Mejilla R, Leveille SG, Ralston JD, Darer JD, Delbanco T, et al. Online Access to Doctors' Notes: Patient Concerns About Privacy. Journal of Medical Internet Research 2013 Sep 26;15(9):186-198. [doi: 10.2196/jmir.2670]
38. Amicizia D, Domnich A, Gasparini R, Bragazzi NL, Lai PL, Panatto D. An overview of current and potential use of information and communication technologies for immunization promotion among adolescents. Human Vaccines & Immunotherapeutics 2013 Dec 01;9(12):2634-2642. [doi: 10.4161/hv.26010]
39. Meesters JJL, Vlieland TPMV, Hill J, Ndosi ME. Measuring educational needs among patients with rheumatoid arthritis using the Dutch version of the Educational Needs Assessment Tool (DENAT). Clinical Rheumatology 2009 May 16;28(9):1073-1077. [doi: 10.1007/s10067-009-1190-3]
40. Kraschnewski JL, Chuang CH, Poole ES, Peyton T, Blubaugh I, Pauli J, et al. Paging "Dr. Google": Does Technology Fill the Gap Created by the Prenatal Care Visit Structure? Qualitative Focus Group Study With Pregnant Women. J Med Internet Res 2014 Jun 03;16(6):e147. [doi: 10.2196/jmir.3385]
41. Hewitt-Taylor J, Bond CS. What E-patients Want From the Doctor-Patient Relationship: Content Analysis of Posts on Discussion Boards. Journal of Medical Internet Research 2012 Nov 08;14(6):301-309. [doi: 10.2196/jmir.2068]
42. Jacobs RJ, Caballero J, Ownby RL, Kane MN. Development of a culturally appropriate computer-delivered tailored internet-based health literacy intervention for spanish-dominant hispanics living with HIV. BMC Medical Informatics and Decision Making 2014 Nov 30;14:103. [doi: 10.1186/s12911-014-0103-9]
43. King L, Watson WL, Chapman K, Kelly B, Louie JCY, Hughes C, et al. Do We Provide Meaningful Guidance for Healthful Eating? An Investigation into Consumers' Interpretation of Frequency Consumption Terms. Journal of Nutrition Education and Behavior 2012 Sep;44(5):459-463. [doi: 10.1016/j.jneb.2011.12.004]
44. Meesters JJL, de Boer IG, van den Berg MH, Fiocco M, Vlieland TPMV. Evaluation of a website providing information on regional health care services for patients with rheumatoid arthritis: an observational study. Clinical Rheumatology 2012 Apr;31(4):637-645. [doi: 10.1007/s10067-011-1897-9]
45. Thakor V, Leach MJ, Gillham D, Esterman A. The quality of information on websites selling St. John's wort. Complementary Therapies in Medicine 2011 Jun;19(3):155-160. [doi: 10.1016/j.ctim.2011.05.005]
46. Cimperman M, Brenčič MM, Trkman P, Stanonik MD. Older Adults' Perceptions of Home Telehealth Services. Telemedicine and E-Health 2013 Oct;19(10):786-790. [doi: 10.1089/tmj.2012.0272]
47. Reeder B, Le T, Thompson HJ, Demiris G. Comparing Information Needs of Health Care Providers and Older Adults: Findings from a Wellness Study. 2013 Presented at: 14th World Congress on Medical and Health Informatics (MEDINFO); Aug 20-23, 2013; Copenhagen, DENMARK p. 18-22.
48. Cameron MH, Poel AJ, Haselkorn JK, Linke A, Bourdette D. Falls requiring medical attention among veterans with multiple sclerosis: A cohort study. Journal of Rehabilitation Research and Development 2011;48(1):13-20. [doi: 10.1682/jrrd.2009.12.0192]
49. Wolff JL, Darer JD, Berger A, Clarke D, Green JA, Stametz RA, et al. Inviting patients and care partners to read doctors' notes: OpenNotes and shared access to electronic medical records. Journal of the American Medical Informatics Association 2017 Apr(E1):E166-E172. [doi: 10.1093/jamia/ocw108]
50. Ntlotlang T, Grand B. The role of libraries in the dissemination of health information in Botswana A study of Mochudi and Molepolole public libraries. Library Review 2016 Jul 04;65(4-5):320-349. [doi: 10.1108/lr-05-2015-0051]
51. Deng ZH, Liu S. Understanding consumer health information-seeking behavior from the perspective of the risk perception attitude framework and social support in mobile social media websites. International Journal of Medical Informatics 2017 Sep;105:98-109. [doi: 10.1016/j.ijmedinf.2017.05.014]
52. Chiu MHP. An investigation of the questions posted on medical consultation websites. Health Information and Libraries Journal 2016 Jul 06;33(4):283-294. [doi: 10.1111/hir.12153]

53. Fu LY, Zook K, Spoehr-Labutta Z, Hu P, Joseph JG. Search Engine Ranking, Quality, and Content of Web Pages That Are Critical Versus Noncritical of Human Papillomavirus Vaccine. Journal of Adolescent Health 2016 Jan;58(1):33-39. [doi: 10.1016/j.jadohealth.2015.09.016]
54. Lee YJ, Kamen C, Margolies L, Boehmer U. Online health community experiences of sexual minority women with cancer. Journal of the American Medical Informatics Association 2019 Aug;26(8-9):759-766. [doi: 10.1093/jamia/ocz103]
55. Huh J, Kwon BC, Kim SH, Lee S, Choo J, Kim J, et al. Personas in online health communities. Journal of Biomedical Informatics 2016 Oct;63:212-225. [doi: 10.1016/j.jbi.2016.08.019]
56. Monkman H, Kushniruk AW. Consumer Medication Information: Similarities and Differences Between Three Canadian Pharmacies. 2017 Presented at: Conference on Information Technology and Communications in Health (ITCH); Feb 16-19, 2017; Victoria, CANADA p. 238-242.
57. Monkman H, Kushniruk AW. All Consumer Medication Information Is Not Created Equal: Implications for Medication Safety. 2017 Presented at: Conference on Information Technology and Communications in Health (ITCH); Feb 16-19, 2017; Victoria, CANADA p. 233-237.
58. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: A systematic review. Health Informatics Journal 2015 Sep;21(3):173-194. [doi: 10.1177/1460458213512220]
59. LeRouge C, Durneva P, Sangameswaran S, Gloster AM. Design Guidelines for a Technology-Enabled Nutrition Education Program to Support Overweight and Obese Adolescents: Qualitative User-Centered Design Study. Journal of Medical Internet Research 2019 Jul 29;21(7):e14430. [doi: 10.2196/14430]
60. Jamal A, Khan SA, AlHumud A, Al-Duhyyim A, Alrashed M, Bin Shabr F, et al. Association of Online Health Information-Seeking Behavior and Self-Care Activities Among Type 2 Diabetic Patients in Saudi Arabia. Journal of Medical Internet Research 2015 Aug 12;17(8):e196. [doi: 10.2196/jmir.4312]
61. Lee J, Kim J, Hong YJ, Piao M, Byun A, Song H, et al. Health Information Technology Trends in Social Media: Using Twitter Data. Healthcare Informatics Research 2019;25(2):99-105. [doi: 10.4258/hir.2019.25.2.99]
62. Gu G, Zhang XT, Zhu XG, Jian Z, Chen K, Wen D, et al. Development of a Consumer Health Vocabulary by Mining Health Forum Texts Based on Word Embedding: Semiautomatic Approach. JMIR Medical Informatics 2019 May 23;7(2):216-224. [doi: 10.2196/12704]
63. Benis A, Harel N, Barkan RB, Srulovici E, Key C. Patterns of Patients' Interactions With a Health Care Organization and Their Impacts on Health Quality Measurements: Protocol for a Retrospective Cohort Study. JMIR Research Protocols 2018 Nov 07;7(11):e10734. [doi: 10.2196/10734]
64. Karnoe A, Furstrand D, Christensen KB, Norgaard O, Kayser L. Assessing Competencies Needed to Engage With Digital Health Services: Development of the eHealth Literacy Assessment Toolkit. Journal of Medical Internet Research 2018 May 10;20(5):e178. [doi: 10.2196/jmir.8347]
65. Koo K, Yap RL. How Readable Is BPH Treatment Information on the Internet? Assessing Barriers to Literacy in Prostate Health. American Journal of Mens Health 2017 Mar;11(2):300-307. [doi: 10.1177/1557988316680935]
66. Haymes AT. The Quality of Rhinoplasty Health Information on the Internet. Annals of Plastic Surgery 2016 Feb;76(2):143-149. [doi: 10.1097/sap.0000000000000660]
67. Abundo BJ. Consumer Health Education (MAPEH). 2015 Sep 15. URL: https://www.slideshare.net/senohj/consumer-health-education-mapeh [accessed 2020-06-08]
68. Boot CRL, Meijman FJ. The public and the Internet: Multifaceted drives for seeking health information. Health Informatics Journal 2010 Jun 23;16(2):145-156. [doi: 10.1177/1460458210364786]
69. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. The Invisible Work of Personal Health Information Management Among People With Multiple Chronic Conditions: Qualitative Interview Study Among Patients and Providers. Journal of Medical Internet Research 2015 Jun 04;17(6):e137. [doi: 10.2196/jmir.4381]
70. Huhta AM, Hirvonen N, Huotari ML. Health Literacy in Web-Based Health Information Environments: Systematic Review of Concepts, Definitions, and Operationalization for Measurement. Journal of Medical Internet Research 2018 Dec 19;20(12):e10273. [doi: 10.2196/10273]
71. Gustafson DH, Hawkins RP, Boberg EW, McTavish F, Owens B, Wise M, et al. CHESS: 10 years of research and development in consumer health informatics for broad populations, including the underserved. International Journal of Medical Informatics 2002 Nov;65(3):169-177. [doi: 10.1016/s1386-5056(02)00048-5]
72. Chang BL. Internet Intervention for Community Elders. Western Journal of Nursing Research 2016 Jun;26(4):461-466. [doi: 10.1177/0193945904263428]
73. Pemu PE, Quarshie AQ, Josiah-Willock R, Ojutalayo FO, Alema-Mensah E, Ofili EO. Socio-demographic Psychosocial and Clinical Characteristics of Participants in e-HealthyStrides (c): An Interactive ehealth Program to Improve Diabetes Self-Management Skill. Journal of Health Care for the Poor and Underserved 2011;22(4):146-164. [doi: 10.1353/hpu.2011.0162]
74. Puspitasari I, Moriyama K, Fukui K, Numao M. Effects of Individual Health Topic Familiarity on Activity Patterns During Health Information Searches. JMIR Medical Informatics 2015 Mar 17;3(1):e16. [doi: 10.2196/medinform.3803]

75. Wang YL. Automatic recognition of text difficulty from consumers health information. 2006 Presented at: 19th IEEE Symposium on Computer-Based Medical Systems; Jun 22-23, 2006; Salt Lake City, UT p. 131-136. [doi: 10.1109/cbms.2006.58]
76. Chen GC, Warren J, Riddle P. Semantic Space models for classification of consumer webpages on metadata attributes. Journal of Biomedical Informatics 2010 Oct;43(5):725-735. [doi: 10.1016/j.jbi.2010.06.005]
77. Rastegar-Mojarad M, Ye Z, Wall D, Murali N, Lin S. Collecting and Analyzing Patient Experiences of Health Care From Social Media. JMIR Research Protocols 2015 Jul 02;4(3):e78. [doi: 10.2196/resprot.3433]
78. Stec MA, Arbour MW, Hines HF. Client-Centered Mobile Health Care Applications: Using the Mobile Application Rating Scale Instrument for Evidence-Based Evaluation. Journal of Midwifery & Women's Health 2019 Mar 18;64(3):324-329. [doi: 10.1111/jmwh.12941]
79. Wongchaisuwat P, Klabjan D, Jonnalagadda SR. A Semi-Supervised Learning Approach to Enhance Health Care Community-Based Question Answering: A Case Study in Alcoholism. JMIR Medical Informatics 2016 Jul 02;4(3):e24. [doi: 10.2196/medinform.5490]
80. Park A, Conway M, Chen AT. Examining thematic similarity, difference, and membership in three online mental health communities from reddit: A text mining and visualization approach. Computers in Human Behavior 2018 Jan;78:98-112. [doi: 10.1016/j.chb.2017.09.001]

Abbreviations

CHI: consumer health informatics

Edited by R Kukafka; submitted 30.06.20; peer-reviewed by S Wen, L Alpay; comments to author 04.11.20; revised version received 23.11.20; accepted 13.07.21; published 09.09.21.

Please cite as:
Ouyang W, Xie W, Xin Z, He H, Wen T, Peng X, Dai P, Yuan Y, Liu F, Chen Y, Luo A
Evolutionary Overview of Consumer Health Informatics: Bibliometric Study on the Web of Science from 1999 to 2019
J Med Internet Res 2021;23(9):e21974
URL: https://www.jmir.org/2021/9/e21974
doi:10.2196/21974
PMID:34499042

Review

Smartphone-Based Interventions to Reduce Sedentary Behavior and Promote Physical Activity Using Integrated Dynamic Models: Systematic Review

Reza Daryabeygi-Khotbehsara[1], MSc; Sheikh Mohammed Shariful Islam[1], MBBS, MPH, FESC, PhD; David Dunstan[2,3], PhD; Jenna McVicar[1], MSc; Mohamed Abdelrazek[4], PhD; Ralph Maddison[1], PhD

[1]Institute for Physical Activity and Nutrition, Deakin University, Geelong, Australia

[2]Physical Activity Laboratory, Baker Heart and Diabetes Institute, Melbourne, Australia

[3]Behaviour, Environment and Cognition Research Program, Mary MacKillop Institute for Health Research, Australian Catholic University, Melbourne, Australia

[4]School of Information Technology, Deakin University, Geelong, Australia

Corresponding Author:
Reza Daryabeygi-Khotbehsara, MSc
Institute for Physical Activity and Nutrition
Deakin University
75 Pigdons Road, Waurn Ponds Victoria
Geelong, 3216
Australia
Phone: 61 3 924 45936
Email: reza.d@deakin.edu.au

Abstract

Background: Traditional psychological theories are inadequate to fully leverage the potential of smartphones and improve the effectiveness of physical activity (PA) and sedentary behavior (SB) change interventions. Future interventions need to consider dynamic models taken from other disciplines, such as engineering (eg, control systems). The extent to which such dynamic models have been incorporated in the development of interventions for PA and SB remains unclear.

Objective: This review aims to quantify the number of studies that have used dynamic models to develop smartphone-based interventions to promote PA and reduce SB, describe their features, and evaluate their effectiveness where possible.

Methods: Databases including PubMed, PsycINFO, IEEE Xplore, Cochrane, and Scopus were searched from inception to May 15, 2019, using terms related to mobile health, dynamic models, SB, and PA. The included studies involved the following: PA or SB interventions involving human adults; either developed or evaluated integrated psychological theory with dynamic theories; used smartphones for the intervention delivery; the interventions were adaptive or just-in-time adaptive; included randomized controlled trials (RCTs), pilot RCTs, quasi-experimental, and pre-post study designs; and were published from 2000 onward. Outcomes included general characteristics, dynamic models, theory or construct integration, and measured SB and PA behaviors. Data were synthesized narratively. There was limited scope for meta-analysis because of the variability in the study results.

Results: A total of 1087 publications were screened, with 11 publications describing 8 studies included in the review. All studies targeted PA; 4 also included SB. Social cognitive theory was the major psychological theory upon which the studies were based. Behavioral intervention technology, control systems, computational agent model, exploit-explore strategy, behavioral analytic algorithm, and dynamic decision network were the dynamic models used in the included studies. The effectiveness of quasi-experimental studies involved reduced SB (1 study; P=.08), increased light PA (1 study; P=.002), walking steps (2 studies; P=.06 and P<.001), walking time (1 study; P=.02), moderate-to-vigorous PA (2 studies; P=.08 and P=.81), and nonwalking exercise time (1 study; P=.31). RCT studies showed increased walking steps (1 study; P=.003) and walking time (1 study; P=.06). To measure activity, 5 studies used built-in smartphone sensors (ie, accelerometers), 3 of which used the phone's GPS, and 3 studies used wearable activity trackers.

Conclusions: To our knowledge, this is the first systematic review to report on smartphone-based studies to reduce SB and promote PA with a focus on integrated dynamic models. These findings highlight the scarcity of dynamic model–based smartphone studies to reduce SB or promote PA. The limited number of studies that incorporate these models shows promising findings. Future research is required to assess the effectiveness of dynamic models in promoting PA and reducing SB.

Trial Registration: International Prospective Register of Systematic Reviews (PROSPERO) CRD42020139350; https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=139350.

(J Med Internet Res 2021;23(9):e26315) doi:10.2196/26315

KEYWORDS

smartphone; mobile phone; physical activity; sedentary behavior; computational models; control systems; systematic review

Introduction

In the past decade, there has been a widespread proliferation in the use of digital technologies to deliver behavior change interventions for health [1]. Given their ubiquity, smartphones, in particular, have been used to improve a wide range of health-related behaviors, including physical activity (PA) and sedentary behavior (SB) [2,3]. Smartphones offer a host of relevant functions, including computational capabilities, built-in sensors (eg, accelerometers and GPS), and internet connectivity, enabling users to run software apps and connect with third-party sensors. Collectively, these features offer the potential for delivering real-time, context-aware, and interactive health care interventions [4].

Theory-based lifestyle interventions have been shown to be more effective than nontheoretical approaches [5]. Thus, to better leverage the potential of mobile technologies for health behavior interventions (mobile health [mHealth]), appropriate behavior change theories and models are needed. Such theories and models need to guide the design and development of complex smartphone interventions that can adapt rapidly in response to various inputs [4]. To date, many smartphone-based interventions to promote PA and reduce SB have relied predominantly on psychological theory, including social cognitive theory (SCT) and self-efficacy theory [2,6]. In a seminal paper, Riley et al [4] argued that current behavioral theories do not meet the need for a more dynamic and interactive nature of digital behavior change interventions, such as just-in-time adaptive interventions. These just-in-time adaptive interventions are complex interventions that adapt throughout time to an individual's time-varying context (where) and status (when) to meet an individual's changing needs for support [7-9]. Riley et al [4] argued that existing psychological theories are relatively static and linear and lack sufficient within-subject dynamic regulatory processes. Furthermore, current psychological theories have been used to tailor interventions based on preintervention data rather than deliver adaptive interventions.

To transform current theories into dynamic frameworks and fully maximize the potential of smartphone technologies, Riley et al [4] highlighted the need to incorporate theories from other disciplines (eg, computer science and engineering) for the future development of adaptive and dynamic digital behavior change interventions. One such theory is the control systems theory—derived from the *control theory* or *cybernetics*—which is a general concept for the understanding of regulatory processes [10] and has various applications in engineering, mathematics, medicine, and economics, among others. Control systems engineering explores how to influence and regulate a dynamic system (eg, time-varying adaptive PA intervention) [11,12]. Applying these dynamic models to health behaviors offers the potential to better predict behavior and provide greater insight into real-time changes, which, in turn, enable the optimization and maintenance of behaviors [9].

Since the study by Riley et al [4] was published, it is unclear how many smartphone-based interventions targeting PA and SB have integrated nonpsycholgical theories to create more dynamic models for digital behavior change interventions, what adaptive factors have been considered, and whether these dynamic interventions improve behaviors. Therefore, this review aims to (1) quantify the number of studies that have used integrated dynamic models to develop smartphone-based interventions to promote PA and reduce SB, (2) describe their features, and (3) evaluate their effectiveness, where possible.

Methods

Design

The systematic review was performed according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [13] and was registered with PROSPERO (International Prospective Register of Systematic Reviews; CRD42020139350) [14].

Study Criteria

This review included studies that developed or evaluated digital behavior change interventions targeting PA, SB, or both and integrated psychological theories with dynamic theories and computational models (eg, control systems engineering); were either adaptive or just-in-time adaptive interventions that included smartphones for delivery; involved human adult participants; included randomized controlled trials (RCTs), pilot RCTs, quasi-experimental, pre-post study designs; and were published from 2000 onward.

Exclusion Criteria

Studies that used conventional theories of behavior change alone without integration with dynamic theories or computational models, case studies, protocols, conference abstracts, dissertations, and reviews were excluded.

Definition

For this review, *dynamic theories* refer to dynamic models taken from other disciplines, including engineering (eg, control systems engineering) and computer science (eg, agent-based modeling). The defining features of dynamic approaches are that they are not static, nonlinear in nature, and capable of capturing complex and rapid changes in behaviors (ie, time-variant) and their influential factors (ie, multivariate). Furthermore, they are quantifiable, empirical, and testable models.

Search Strategy

Databases (IEEE, PubMed, PsycINFO, Cochrane, and Scopus) were searched from January 2000 to May 15, 2019, without language restriction. Keywords (including Medical Subject Headings terms) and phrases comprised 3 components (mHealth, dynamic models, and activity), where "OR" and "AND" Boolean operators were used for within and between component searching (Multimedia Appendix 1). The wild-card term "*" was used where necessary to potentiate sensitivity. Snowball searching was performed using the included studies to identify additional relevant research. The search results were exported to a reference manager software (EndNote X9; Clarivate Analytics) for review and extraction.

Screening Process and Data Extraction

Two researchers (JMV and RDK) independently screened and reviewed the titles and abstracts to identify eligible studies. The full text of the included papers was assessed based on the study criteria. The following information was collected: author and year, country, study design, duration of the study, recruitment and setting, the population of the study, sample size, inclusion criteria, participant characteristics, dynamic model, theory or constructs integrated, and outcomes measured (SB and PA behaviors).

Quality Assessment

Two researchers (JMV and RDK) assessed the risk of bias. The Cochrane Handbook for Systematic Reviews of Interventions [15] was used to evaluate randomized studies for selection bias, detection bias, attrition bias, performance bias, and reporting bias as the main sources of bias. Other sources of bias were also considered. In addition, the Joanna Briggs Institute Critical Appraisal Checklist for Quasi-Experimental Studies [16] was used to assess nonrandomized studies. Where available, protocols and trial registry data were found for risk of bias assessment. Where multiple reports existed for the same study, data were extracted from all reports and expressed together. The authors were contacted for further information, as needed.

Data Analysis

The data were synthesized narratively to address the aims of this review. Given the heterogeneity of the included studies in terms of methodology, outcome measures, and statistical approaches, a meta-analysis of effectiveness data was not conducted. Instead, a synthesis without a meta-analysis method—vote counting based on the direction of effects—was used to synthesize data [17]. The effect direction is a standardized binary metric based on the observed benefit (positive) or harm (negative). Vote counting is based on effect direction and compares the number of positive effects with the number of negative effects on an outcome. An effect direction plot is used for the visual representation of data and linking narrative synthesis to the overall conclusion [18,19]. In this review, the updated method of the effect direction plot is used as outlined elsewhere [20]. Changes within the intervention arm of controlled studies and changes from baseline in uncontrolled studies were considered for judgment. PA outcomes including light activity, walking (time and steps), moderate-to-vigorous PA (MVPA), nonwalking exercise, and total PA time from 6 studies were grouped as PA health domains. For studies with multiple PA outcomes, the effect direction was where 70% or more of the outcomes reported a similar direction (positive or negative). If less than 70% of outcomes showed a similar direction, they were reported as conflicting findings or no clear effect. A sign test was applied to test any evidence of an effect across studies. A 2-tailed *P* value was then calculated to show the probability of observing positive and negative findings for the PA health domain.

Results

Overview

A total of 1087 study reports were identified after removing duplicates. In addition, 9 studies were identified through a manual search. A total of 76 research articles underwent a full review, and 11 reports describing 8 studies were eligible and included in this systematic review. The characteristics of the included studies are summarized in Table 1. The inclusion process and reasons for exclusion are shown in the PRISMA flow diagram (Figure 1). The reasons for excluding 65 studies (in full-text review) are provided in Multimedia Appendix 2 [21-85].

Table 1. General characteristics of the included studies.

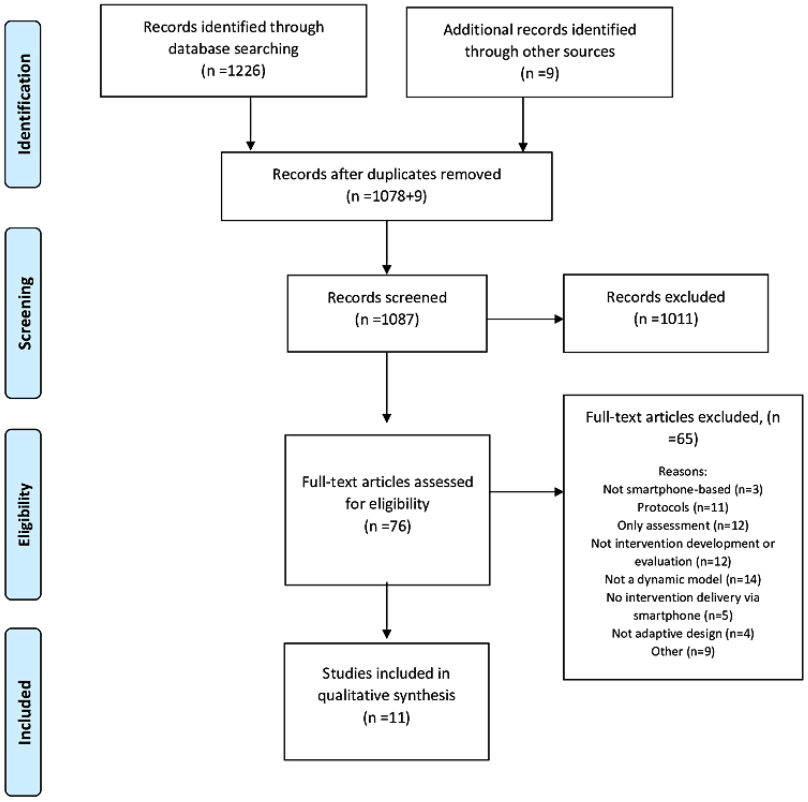
Author (year)	Country	Study design and duration	Recruitment setting	Sample, n	Inclusion criteria	Participants' characteristics
Baretta et al (2019) [86]	Italy	Intervention development; 8 weeks	Indoor activity settings (eg, gyms)	60	Not described	• People who did not meet PAa guidelines • Age (35-60 years) • Female (35/60, 58%)
Direito et al (2019) [87] (other related reference: Direito et al [88])	New Zealand	Pre-post single-arm intervention; 8 weeks	Community	69	• 17-69 years • Owning an Android phone	• Insufficiently active healthy adults (either those who did not meet PA recommendations or who intended to decrease sedentary behavior) • Mean age 34.5 (SD 11.8) years • Female (54/69, 78%) • Mean BMI 25.6 (SD 4.95) kg/m2 • Ethnicity: New Zealand European (38/69, 55%)
Conroy et al (2018) [12]	United States	Single-group microintervention; 16 weeks	Community (via advertisement)	10	Adults not meeting federally recommended levels of aerobic PA but had no contraindications to PA	• Mean age 34.4 (SD 9.0) years • Female (9/10, 90%) • Employed full time (8/10, 80%) • Parents (6/10, 60%) • Single (5/10, 50%), married (4/10, 40%), or divorced (1/10, 10%) • Education (6/10, 60% with at least a bachelor's degree) • White (9/10, 90%), Asian American (1/10, 10%), and none were Hispanic or Latino
Middelweerd et al (2020) [89] (Other related references: Klein et al [90] and Middelweerd et al [91])	The Netherlands	3-arm quasi-experimental; 12 weeks	Community (flyers, posters, social media, personal contacts, and snowball strategies)	104	Adults aged 18-30 years at the time of registration, in possession of a suitable smartphone running on Android or iOS, apparently healthy, Dutch-speaking, and signed the informed consent form	• Healthy young adults • Mean age 23.4 (SD 3.0) years • Female (83/104, 79.8%) • Students (72/104, 69.2%) • Mean BMI 22.8 (SD 3.4) kg/m2 • Previous experience with PA apps (33/104, 31.7%)
Korinek et al (2018) [92] (other related references: Freigoun et al [93] and Martin et al [22])	United States	Pre-post single-arm intervention; 14 weeks	Nationally via community advertising methods (eg, email to student listservs, word-of-mouth, and social media ads)	20	Generally healthy, insufficiently active, 40 to 65 years, BMI 25 to 45 kg/m^2, owned and regularly used an Android phone capable of connecting to a Fitbit Zip via Bluetooth 4.0	• Overweight and sedentary adults • Age (47 years) • Mean BMI 33.8 (SD 6.82) kg/m2 • Female (18/20, 90%) • Walked on average 4863 steps per day

Author (year)	Country	Study design and duration	Recruitment setting	Sample, n	Inclusion criteria	Participants' characteristics
Rabbi et al (2015) [94]	United States	Pilot RCT[b]; 3 weeks	Advertisement placed throughout the university campus	17 (intervention=9; control=8)	Owned an Android mobile phone, interested in fitness	• Adult students and staff • Mean age 28.3 (SD 6.96) years • Student (13/17, 76%) • Female (8/17, 47%) • All participants (low-to-moderate PA)
Rabbi et al (2018) [95]	United States	Pilot Pre-post single-arm intervention; 5 weeks	Via the Wellness Center and retiree mailing lists from Cornell University	10	People with a history of chronic back pain (≥6 months in duration); willing to use MyBehavior-CBP; having some reasonable level of outdoor movement (eg, traveling to and from work); not being significantly housebound; with a basic level of mobile phone proficiency; aged between 18 years and 65 years; and fluent in English	• Adults with chronic low back pain • Mean age 41.1 (SD 11.3; range 31-60) years • Female (7/10, 70%)
Zhou et al (2018) [96]	United States	RCT; 10 weeks	Email announcement; university campus	64 (intervention=34; control=30)	Staff member, intended to be physically active in the next 10 weeks; own an iPhone 5s or newer; willing to keep the phone in the pocket during the day; willing to install and use the study App; able to read and speak English	• Adult staff employees • Small fraction had the following conditions: high blood pressure (5/64, 8%), type 2 diabetes (5/64, 8%), hypercholesterolemia (7/64, 11%) • Married or cohabitating (34/64, 56%) • White or non-Hispanic (29/64, 45%) • Full-time job (45/64, 70%) • Mean age 41.1 (SD 11.3) years • Female (53/64, 83%)

[a]PA: physical activity.

[b]RCT: randomized controlled trial.

Figure 1. Flow of studies. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses [13].



General Description of the Studies

A total of 5 studies were conducted in the United States [12,92,94-96], 1 in Italy [86], 1 in New Zealand [87,88], and 1 in the Netherlands [89-91]. Of these, 3 studies used pre-post intervention designs [87,92,95], 2 were RCTs [94,96], 1 was a 3-arm quasi-experimental study [89], 1 was a single-group microrandomized trial [12], and 1 was development study [86]. The duration of the studies ranged from 3 weeks to 6 months. A total of 5 studies recruited participants from community settings [12,86,87,89], 1 from the university campus and community [92], 2 from the university only [94,96], and 1 from a university wellness center and retiree mailing list [95]. Populations included insufficiently active and sedentary healthy adults, healthy and highly educated young adults, overweight and sedentary adults, adults with chronic low back pain, and students and staff from a university setting. Sample sizes ranged from 10 to 104 participants in the intervention evaluation studies and 60 in the development study. Participants were predominantly women in all studies except one [94]. The general characteristics of the included studies are summarized in Table 1.

Theoretical Premise

SCT was the predominant psychological theory used [12,89,92,94,95]. A study incorporated self-efficacy theory [86], with a dynamic decision network—a sequence of simple Bayesian networks used to describe probabilistic computational models [97]. A study used an integrated behavior change model incorporating 33 behavior change techniques (eg, self-monitoring, goal setting, and review of goals) combined with the behavioral intervention technology model [87]. Two studies incorporated control systems engineering models integrated with SCT [12,92]. In a study, SCT, self-regulation theory, and health action process approaches were integrated with a computational agent model—an intelligent reasoning system [89]. Learning theory, the Fogg behavior model, and SCT were combined with the exploit-explore strategy in 2 studies [94,95]. Rather than using a theoretical framework, a study integrated a single behavior change technique (goal setting) with a behavioral analytic algorithm [96].

Featured Description of Interventions

All studies promoted PA, whereas 4 studies also involved interventions for reducing SB [12,87,94,95]. Few studies explicitly stated the inclusion of behavior change techniques [87,89,92] as part of their intervention, 2 of which included a range of behavior change techniques [87,89]. Conroy et al [12] did not describe specific behavior change techniques but stated using intervention messages, which targeted key SCT constructs (eg, outcome expectancies, risk awareness and planning, efficacy-building affirmations, social support, and evoking anticipated reward or regret). The most common behavior change technique used across all studies was goal setting [86,87,89,92,96]. In terms of PA, 3 studies [87,92,96] included daily goal setting to achieve PA targets, whereas 1 study promoted weekly goal setting [89]. In a study, weekly step goals were initially established and then broken down into daily short-term goals [86]. Only 1 study set goals for SB [87]. To help participants set PA and SB goals, 5 studies used past activity performance [86,87,89,92,96], whereas 2 also took into account individuals' perceptions of self-efficacy [86,92]. Instead of setting goals, 2 studies focused on habit formation by providing suggestions from an individual's past frequent and infrequent activities after manual and automatic logging and

clustering of past activities [94,95]. Habit formation was also accounted for in another study [87]. For SB, Direito et al [87] encouraged participants to replace periods of extended sedentary time with light-intensity walking and standing, whereas Conroy et al [12] included *sit less* and *move more* messages. Two other studies by Rabbi et al [94,95] targeted both SB and standing by promoting short walks. None of the studies measured standing as an outcome.

Monitoring and feedback on behavior was another widely used behavior change technique [86,87,89,92,94,96]. All 6 studies used visual and numerical feedback on behavior, whereas 2 used biofeedback to help monitor behavior [86,94]. Four studies included a reward in the form of social rewards [19,20,94] and material incentives [17]. In terms of the type of intervention, 2 studies used push notification messages [87,89], 3 used push notifications to present step goals or minutes of activity goals (eg, walking) [86,92,96], 2 had in-app suggestions selected from frequent and infrequent past activities [94,95], and 1 used text messages [12].

In total, 7 studies used mobile apps, 6 of which ran on Android [86,87,89,92,94,95] and 1 on iPhone operating systems (iOS) [96]; 1 study did not mention the operating system used [12]. Four studies including TODAY, MyBehavior, MyBehaviorCBP, and CalFit [87,94-96] used built-in smartphone sensors (ie, accelerometers) to measure activity, and 3 studies used wearable activity trackers (Fitbit One, Fitbit Zip, and ActivPAL3) [12,89,92]. A heart rate sensor was used to measure activity in the study by Baretta et al [86]. Furthermore, 3 studies used the phone GPS to identify geo-locations [89,94,95]. Some studies used built-in phone GPS and apps to capture and account for environmental contexts such as location (eg, workplace) [87,89,94,95], weather [89,92], and weekend or weekday [12,92]. The JustWalk intervention incorporated psychological states (eg, stress) and measures of busyness and sleep quality. Further details have been provided in Table 2.

Table 2. Features of smartphone-based physical activity intervention development or evaluation.

Author (year)	Intervention	Control	Theoretical premise	Primary outcome	Other outcomes	Technology feature	Results
Baretta et al (2019) [86]	Weekly tailored PA[a] goals • Starting goal (first week): 120 METb • Long-term goal: 600 METs per week of PA • Weekly goals broken down into daily goals • Factors not considered in the intervention development but proposed for the next study: working hours, time of the day, day of the week, health and illness, weather, etc	N/A[c]	Self-efficacy theory and dynamic decision network	PA measured by HR[d] sensor, self-efficacy beliefs	N/A	• Android app: Muoviti (visualizing the heart-beat rate graph of the last training session, the curves of weight and waistline variations week by week, the burned calories graph, session by session, and the percentage of vigorous activity with respect to moderate activity) • Other: HR wristbands (MioAlpha and PulseON)	• N/A
Direito et al (2018 and 2019) [87,88]	Daily individualized and adaptive PA and SB[e] goals: • Daily activities (eg, transport to or from work, PA at work) • Light-intensity activity to replace SB (eg, walking to a colleague's desk rather than call or email, stand up while on the phone) • Leisure-time moderate-to-vigorous PA (eg, cycling) • Daily goals, visual and numerical feedback on past day and historical data, tips or suggestions, infographics, videos, and links, frequently asked questions, reminders, and push notifications • Context: workplace (location)	N/A	Intervention mapping taxonomy to identify behavior change techniques (eg, self-monitoring, goal setting, or review of goals) from literature. Integrated behavior change model constructs and behavioral intervention technology; 33 behavior change techniques were included	Test the acceptability and feasibility of just-in-time adaptive intervention on PA and SB	Pilot-testing the TODAY[f] app	• Android apps: Art of Living app and TODAY app. Other: built-in phone sensors for SB and activity (ie, accelerometer)	• TODAY app: low-effort and pleasant (54.3%), provides guidance on changing activity profile (52.6%), positively framed messages (64.4%), the app sustained interest over the 8 weeks (28.8%) • Most favorable behavior change techniques for the users (goal setting, discrepancy between current behavior and goal, feedback on behavior, instruction on how to perform the behavior, and behavior substitution) • Only significant improvement was occurred on light PA (see the results for statistics)

Author (year)	Intervention	Control	Theoretical premise	Primary outcome	Other outcomes	Technology feature	Results
Conroy et al (2018) [12]	Five daily text messages (between 8 AM to 8 PM). Three message types (move more, sit less, general facts or trivia [unrelated to PA or SB]). Message receipt was confirmed with a reply. Factors: context (weekday and weekend)	N/A	Social cognitive theory and control systems engineering	Stepping time	N/A	• No app or text message • ActivPAL3 (activity tracker)	• (Proof-of-concept study) 50% of the sample: more pronounced behavioral responses to text messages on weekends than weekdays; 50% had similar weekend or weekday responses; 50% of responders increased stepping time in response to "move more" messages, and 50% increased stepping time in response to "sit less" messages
Middelweerd et al (2020) [89], Klein et al (2017) [90], and Middelweerd et al (2018) [91]	Weekly moderate-to-vigorous PA goals: 30 minutes of moderate PA for at least 5 days a week or 20 minutes of vigorous PA for 3 days a week Contexts (location, weather, occupation) Connected friends (Facebook API[g]), if 2 participants of the intervention are connected Up to 3 messages a day	N/A	Social cognitive theory, self-regulation theory and health action process approach and computational agent model	To increase the total time spent in moderate-to-vigorous PA	N/A	• Android app: Active2Gether • Fitbit One (for self-monitoring only), ActiGraph wGT3XBT and GT3X+ (activity trackers)	• No significant intervention effects were found for the Active2Gether-full and Active2Gether-ight conditions on levels of PA compared with the Fitbit condition: larger effect size for Active2Gether-ight (β=3.1, 95% CI −6.66 to 12.78, for minutes of moderate-to-vigorous PA; β=5.2, 95% CI −1334 to 1345, for steps). Smaller effect size for Active2Gether-full (β=1.2, 95% CI −8.7 to 11.1, for minutes of moderate-to-vigorous PA; β=−389, 95% CI −1750 to 972, for steps)
Korinek et al (2018) [92] and Freigoun et al (2017) [93]. More information is available in Martin et al (2018) [22]		N/A	Social cognitive theory (particularly self-efficacy construct), goal setting and control systems engineering (system identification)	Feasibility, daily steps	N/A	• Android app: JustWalk • Fitbit Zip (activity tracker) • Other: web-based mobile questionnaire	

Author (year)	Intervention	Control	Theoretical premise	Primary outcome	Other outcomes	Technology feature	Results
	Daily step goal: • Pseudorandomly assigned daily step goal (doable [based on baseline median daily step] and ambitious [ie, up to 2.5×baseline median])+ rewards (points>Amazon Gift Cards) • Six 16-day cyclesh (cycle 0 [baseline], cycles 1 to 5 [step goals assigned]) • Step goals prompted every morning+there were daily, weekly and monthly surveys • Morning and evening EMAi assessed constructs including (eg, confidence in achieving the goal, predicted busyness for that day, previous night's sleep quality) • Factors considered: perceived stress, perceived busyness, weather information, sleep quality						• Linear mixed effect model: each individual walked below 5000 steps at baseline with significant variation; mean intercept value 4863.3 steps (SD 1838.42), t_{98}=10.49; P<.001. • Daily steps increased by 2650 steps per day on average from day 0 to day 16 (cycle 0 to cycle 1); t_{98}=6.54, P<.001. • Quadratic mixed effect model: each individual walked roughly 5000 steps at baseline with significant variations; mean intercept value 5301.5 steps (SD 1862.04); t_{98}=11.29, P<.001. • Daily steps increased by 1500 steps per day on average from cycle 0 to cycle 1 (1505 steps; t=5.52, P<.001); however, daily steps decreased by 247.3 steps per day on average from day 0 to day 16 (cycle 0 to cycle 1); t_{98}=-5.01, P<.001 • High adherence was observed (only 10 days of having missing step data; only 40 days of nonwear; <500 step counts). Common problem: sync lag with Fitbit
Rabbi et al (2015) [94]		Nonpersonalized generic recommendations	Learning theory, Fogg behavior model, social cognitive theory, and exploit-explore strategy[j]	Adherence, acceptability, behavior change	N/A	• Android app: MyBehavior; other: phone accelerometer and GPS	

Author (year)	Intervention	Control	Theoretical premise	Primary outcome	Other outcomes	Technology feature	Results
	Daily personalized context-sensitive suggestions (PA and stationery). Manual and automatic logging to track activity and user location. Start of each day: 10 in-app activity suggestions (90% users' most frequent activities [exploit]; 10% from users' infrequent activities [explore]). MyBehavior app included both PA and dietary interventions						• Intervention participants more intended to follow personalized suggestions than control (effect size=0.99, 95% CI 0 to 1.001; *P*<.001). Most intervention participants (78%) had a positive trend in walking behavior (also increased daily walking by 10 minutes during the intervention), whereas most control participants (75%) showed a negative trend. The users found MyBehavior app suggestion very actionable and wanted to follow them
Rabbi et al (2018) [95]	Context-sensitive suggestions (PA and stationery). Manual and automatic logging to track activity and user location. In-app suggestions (80% users' most frequent activities [exploit]; 20% from users' infrequent activities [explore]); total time for each selected activity must not exceed 60 minutes. End of day reward score	Static suggestions	Learning theory, Fogg behavior model, social cognitive theory (self-efficacy) and exploit-explore strategy[j]	Use, acceptability, early efficacy	Qualitative feedback	• Android app: MyBehaviorCBP; other: phone accelerometer and GPS	• Intervention condition increased daily walking by 4.9 minutes (β=4.9; *P*=.02) significantly. Exercise time was increased nonsignificantly by 9.5 minutes (β=9.5; *P*=.31). MyBehaviorCBP was opened 3.2 times a day (on average). MyBehaviorCBP suggestions were perceived as low-burden (β=.42; *P*<.001). Back pain was reduced in the intervention condition, but not significantly (β=−.19; *P*=.24). Participants suggested consideration of weather, weekend or weekday, and level of pain for future interventions
Zhou et al (2018) [96]		Steady step goals (10,000 per day)	Goal setting and behavioral analytics algorithm[k]	Change in daily step		• iOS app: CalFit; other: built-in health chip in the iPhone	

Author (year)	Intervention	Control	Theoretical premise	Primary outcome	Other outcomes	Technology feature	Results
	Daily step goals (real-time, automated adaptive). Push notifications via app. Daily notifications at 8 AM. If the goal was accomplished before 8 PM, a congratulation notification was sent.				Step goal attainment, weight, height, barriers to being active quiz, IPAQ[l]-short form		• Mean daily step count was decreased by 390 steps (SD 490) per day in the intervention versus 1350 steps (SD 420) per day in the control from baseline to 10 weeks (net difference: 960 steps, *P*=.03)

[a]PA: physical activity.

[b]MET: metabolic equivalents.

[c]N/A: not applicable.

[d]HR: heart rate.

[e]SB: sedentary behavior.

[f]TODAY: Tailored Daily Activity.

[g]API: application programming interface.

[h]Step goals did not increase between cycles.

[i]EMA: ecological momentary assessment.

[j]Grounded in artificial intelligence and a subcategory of a broader decision-making framework called multiarmed bandit, which stems from probability theory.

[k]Behavioral analytics algorithm uses machine learning to build a predictive model–based on historical and goal steps for a particular person and then uses this estimation to generate challenging yet realistic and adaptive step goals based on a predictive model that would maximize the physical activity in the future.

[l]IPAQ: International Physical Activity Questionnaire.

Effectiveness of Interventions

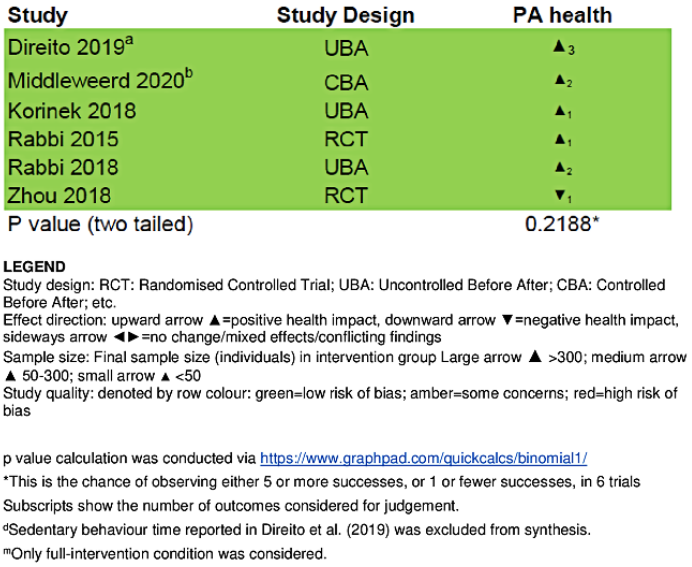
Narrative Synthesis of Individual Studies

A total of 6 studies reported on the effectiveness of the intervention [87,89,92,94-96]; the details are presented in Table 1. The intervention by Direito et al [87] increased the time spent in light and moderate-to-vigorous intensity PA and total PA time; however, only light-intensity PA achieved statistical significance from pre- to postintervention assessments (adjusted mean difference 2.2 minutes, SE of difference 1.0; 95% CI 0.78-3.56; *P*=.002). A small, but statistically nonsignificant, decrease in SB was observed (adjusted mean difference −9.5 minutes, SE of difference 7.5; 95% CI 19.98-1.05; *P*=.08). The Active2Gether intervention involved 3 arms of Active2Gether-full (tailored coaching messages, self-monitoring, and social comparison), Active2Gether-light (self-monitoring and social comparison), and Fitbit app control condition (self-monitoring). The Active2Gether did not show an effect on PA levels (average daily minutes of MVPA and step counts) compared with the Fitbit app (β=1.2, 95% CI −8.7 to 11.1, *P*=.81, for minutes of MVPA; β=−389, 95% CI −1750 to 972, *P*=.57, for steps and β=3.1, 95% CI −6.66 to 12.78, *P*=.53, for minutes of MVPA; β=5.2, 95% CI −1334 to 1345, *P*=.99, for steps, for the full and light app, respectively). The JustWalk intervention increased the average daily steps by 2650 steps in 16 days (t_{98}=6.54; *P*<.001). This effect decreased from day 16 to day 96 (average daily change −109.1 steps; t_{98}=−1.42; *P*=.15), suggesting acceptable maintenance. Users of the MyBehavior app walked an average of 10 minutes per day more from the first to the third week. There was no change in the control group (between-group differences were statistically significant (t_{15}=2.1; *P*=.06; 95% CI −0.23 to 19.05; *d*=0.9). In the second study by Rabbi et al [95], MyBehaviorCBP was associated with an increased daily walking time of 4.9 minutes (β=4.9; *P*=.02; 95% CI 0.8-0.89; *d*=0.31) among adults with chronic back pain. Nonwalking exercise time also increased by 9.5 minutes, but it was not statistically significant (β=9.5; *P*=.31; 95% CI −6.3 to 21.8; *d*=0.03). The Cal Fitness trial showed that the mean daily step count decreased in the 10-week intervention for both the intervention (mean −390, SD 490) and control group (mean −1350, SD 420; net mean difference 960; 95% CI 90-1830; *P*=.03). The Conroy et al [12] study was conducted to determine proof-of-concept and did not report effectiveness data (for descriptive results, see Table 2).

Effect Direction Plot

This study included 6 interventional studies. Figure 2 shows the effect direction plot for the PA health outcome domain; 5 of 6 interventions reported a positive effect direction, with 1 study showing a negative effect on PA health. The *P* value for the sign test for PA health was *P*=.21.

Figure 2. Effect direction plot summarizing the direction of impact from smartphone-based physical activity interventions.



Risk of Bias Assessment of the Included Interventions

Judgments on the risk of bias for the 2 RCTs and 4 quasi-experimental studies are presented in Multimedia Appendix 3 [12,87,89,92,94-96]. Overall, the included studies were of relatively high quality. The 2 RCTs [94,96] were judged to be low risk in all domains except one (ie, blinding of participants and personnel). All included quasi-experimental studies lacked a control group because of a pre-post [87,92,95] or single-group intervention [12] design. These interventions did not introduce additional risks to the remaining eight domains.

Discussion

Principal Findings

This review aims to quantify the number of studies that have integrated traditional psychological theories with dynamic computational models in the development or evaluation of smartphone interventions to reduce SB and promote PA. Although we showed that a few studies—mainly pilot, feasibility, and proof-of-concept—have taken an integrated dynamic approach, there was no consensus on what dynamic model–based approach should be used and how. Overall, it was difficult to draw a conclusion on the effectiveness of the included smartphone interventions; however, preliminary findings on PA are promising, less so for SB. Moreover, an effect direction plot was used to illustrate the direction of the intervention effect on PA outcomes, regardless of their statistical significance.

This review was driven in part by a paper by Riley et al [4] who argued that to truly capture the benefits of smartphones to deliver real-time and adaptive interventions, they need to adopt principles from other disciplines, such as control systems engineering, and integrate them with traditional health behavior theories. In total, we found only 8 studies that had adopted this notion, most of which used SCT for integration, with considerable complexity in the approaches used, ranging from a basic use of behavioral analytic algorithms to a more sophisticated approach using control systems.

Advancements in smartphone technology have enabled the collection of intensive contextual and longitudinal (time-variant) data, which facilitate the delivery of automated, real-time, and adaptive behavior change interventions such as just-in-time adaptive interventions. These features permit the testing of specific intervention components (eg, behavioral messaging comprising behavior change techniques). Control systems appear to offer an excellent fit for the development of adaptive smartphone interventions. It explores ways to influence a dynamic system (eg, time-varying adaptive PA intervention) and how to regulate it [11,12]. In other words, control systems engineering provides a dynamic approach to designing tailored interventions that adapt over time and are based on real-time data (ie, intensive longitudinal data) [98]. Despite the variability in the application of dynamic models outlined in this review, existing evidence suggests that their integration with traditional behavior change and psychological theories offer exciting opportunities to better understand human behavior (eg, SB and PA), identify patterns of behavior, and optimize individually adapted behavior change interventions.

Few of the included studies evaluated the effectiveness of the interventions, and small effects were observed on PA and SB. Possible reasons for the small effect sizes may have included inappropriate design (nonrandom allocation) [89], lack of exposure to automated intervention because of technical problems [89], use of nonpersonalized behavioral interventions [94], lack of participant engagement with the intervention [87], and insufficient inclusion of behavior change techniques [96]. Moreover, a binary sign test conducted in this review attempts to provide additional information and contribute to transparency in interpreting the effect direction. However, this should be interpreted carefully, as the small number of studies may have underpowered the test.

Most of the studies included in the review focused on PA, whereas only a few targeted SB; none assessed standing as a distinct outcome. Moreover, most smartphone-based SB and

PA interventions used built-in smartphone accelerometers and sensors as a tool to capture individual behaviors to inform behavioral interventions (ie, step counts were used to help participants set goals and monitor progress or provide activity suggestions) [87,94-96].

The benefits of smartphone interventions include the ability to collect and measure contextual factors (eg, location, weather, and emotional or psychological states), which could be used to personalize behavior interventions [99]. Existing research evidence has shown that contextually aware reminders increase the effectiveness of mHealth PA interventions [21,100]. Furthermore, leveraging contextual information in PA interventions enables the triggering of more frequent reminders without annoying the individual receiving the reminder, and these types of interventions are considered more acceptable [100]. Despite these proposed benefits, most of the included studies lacked an assessment of contextual factors. A likely reason for the lack of contextual factors in the reviewed studies is the technical challenges, such as system requirements. For example, high battery consumption and low localization speed by a built-in smartphone GPS compromise mobile app performance [101]. Another important reason might be the privacy implications for smartphone users [102]. Privacy breaches are most probable when context-sensitive information such as location is monitored [103]. Moreover, people generally refuse to be monitored for where they go or what they do [104]. A limitation of using native smartphone sensors is that they do not provide research-grade precision for measuring PA and SB. Commonly used accelerometers (eg, built-in smartphone accelerometers and Actigraph GT3X) measure SB by focusing on periods where the device records activity counts below a certain cutoff point, such as less than 100 counts per minute [105]. This leads to the miscategorization of SB [106]. Although postural devices (inclinometers) such as activPAL have excellent accuracy in measuring SB [107], they require proprietary software (activPAL Professional Research Edition, PAL Technologies) to process and collect the data and thus have low utility for real-time interventions. Finally, as highlighted above, none of the included studies assessed standing as an outcome, despite 3 studies promoting standing in their intervention messages [87,94,95]. This might be explained by the inability to measure standing in real time for a dynamic intervention purpose and limited evidence advocating standing as a distinct activity that brings health benefits. However, short-term and small-scale studies that support standing are emerging. In a lab-based study, breaking up every 30 minutes of sitting by 5 minutes of standing was shown to reduce postprandial blood glucose (34% reduction) compared with prolonged sitting in postmenopausal women [108]. Moreover, an office-based study has shown that an afternoon of standing reduced postprandial glucose (43% reduction) compared with sitting while performing computer work [109].

The included interventions comprised pre-post, RCT, and 3-arm quasi-experimental designs. These commonly used experimental designs are unable to assess rich context and time-intensive data. For example, RCTs do not provide information on the particular time when the intervention had an effect and the moderators that affected the behavior change [110]. In fact, RCTs typically consider the overall impact of an intervention package on behavior or health outcomes, not specific components of that intervention. Other study designs, such as factorial designs, are capable of investigating the effects of each intervention component and the interactions between components and the dosing of the intervention. However, they are not sufficient to delineate when the intervention was most effective and what moderators influenced the intervention [110]. A microrandomized trial may address these design limitations. The microrandomized trial is a novel experimental design to determine the optimal delivery of just-in-time adaptive interventions [110]. A key advantage is that microrandomized trials not only assess the effect of specific intervention components but also changes in effects over time and moderators, including contextual and psychological factors [110]. Microrandomization can help elucidate potential causal relationships between each randomized intervention feature and proximal effects (what happens in a limited time window, for example within 1 hour, following a randomized intervention) and allow assessment of time-varying contextual and psychological factors moderating those proximal effects [110].

Most of the included studies lacked comprehensive incorporation and testing of behavior change techniques, although they were theory-based. The precise specification of behavior change techniques—which are active ingredients of behavior change interventions and specification of intervention features of PA (eg, mode of delivery and frequency)—help provide accumulative evidence for effective and replicable interventions [111]. Smartphone-based interventions undertaking dynamic approaches with a proper experimental design (ie, microrandomized), while testing various behavior change techniques, are expected to provide more robust evidence than traditional theory approaches.

Limitations and Strengths

A limitation of this review is the heterogeneity in the reported effectiveness data that prevented a pooled meta-analysis. Other limitations include the small sample size and short duration of the included interventions and nonrandomized study designs. Moreover, women exceeded men in most studies, and all studies involved adult populations, which might limit the generalizability of the findings. A key strength of this review is that it focuses on the integration of dynamic models in smartphone-based PA and SB studies, as such dynamic models fit best with mobile technologies. Another strength is the use of the effect direction plot to present the direction of the effectiveness results. This methodology is superior to narrative synthesis, as it helps with the overall interpretation of the findings. Future studies, in the context of SB and PA behaviors, are suggested to incorporate and assess the effect of relevant environmental and internal contextual moderators, use computational models, and investigate SB, in particular, as there is a significant evidence gap.

Conclusions

In conclusion, despite the recommendation for integrating dynamic models such as control systems to better harness the potential of mobile technologies, this review showed that few studies have actually adopted this approach to promote PA and

reduce SB. To some extent, this research gap may be because of the complex and multifaceted nature of dynamic models, such as control systems, in integrating adaptive contexts and real-time measurement of outcomes.

Acknowledgments

The authors would like to thank Rachel West (Liaison Librarian at Deakin University) and Dr Gavin Abbott (Senior Research Fellow and Biostatistician at Deakin University) for their guidance. RDK received a Deakin University Postgraduate Research Scholarship for his PhD.

Authors' Contributions

RDK, RM, and SMSI designed the study. RDK conducted the search and removed duplicates; RDK and JM screened the titles, abstracts, and full text of the studies and extracted data; and RDK drafted the systematic review manuscript. RM, SMSI, DWD, and MA contributed to verifying screening; all coauthors—SMSI, DWD, JM, MA, and RM—contributed to the critical revision of the manuscript and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 13 KB - jmir_v23i9e26315_app1.docx]

Multimedia Appendix 2
Reason for exclusion.
[DOCX File , 16 KB - jmir_v23i9e26315_app2.docx]

Multimedia Appendix 3
Quality assessment.
[DOCX File , 15 KB - jmir_v23i9e26315_app3.docx]

References

1. Taj F, Klein MC, van Halteren A. Digital health behavior change technology: bibliometric and scoping review of two decades of research. JMIR Mhealth Uhealth 2019 Dec 13;7(12):e13311 [FREE Full text] [doi: 10.2196/13311] [Medline: 31833836]
2. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
3. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr 18;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
4. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar 24;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
5. Michie S, Abraham C. Interventions to change health behaviours: evidence-based or evidence-inspired? Psychol Health 2004 Feb;19(1):29-49. [doi: 10.1080/0887044031000141199]
6. Salwen-Deremer JK, Khan AS, Martin SS, Holloway BM, Coughlin JW. Incorporating health behavior theory into mHealth: an examination of weight loss, dietary, and physical activity interventions. J Technol Behav Sci 2019 Nov 04;5(1):51-60. [doi: 10.1007/s41347-019-00118-6]
7. Intille S, Kukla C, Farzanfar R, Bakr W. Just-in-time technology to encourage incremental, dietary behavior change. AMIA Annu Symp Proc 2003:874 [FREE Full text] [Medline: 14728379]
8. Spruijt-Metz D, Wen CK, O'Reilly G, Li M, Lee S, Emken BA, et al. Innovations in the use of interactive technology to support weight management. Curr Obes Rep 2015 Dec 12;4(4):510-519 [FREE Full text] [doi: 10.1007/s13679-015-0183-6] [Medline: 26364308]
9. Spruijt-Metz D, Nilsen W. Dynamic models of behavior for just-in-time adaptive interventions. IEEE Pervasive Comput 2014 Jul;13(3):13-17. [doi: 10.1109/mprv.2014.46]
10. Carver CS, Scheier MF. Control theory: a useful conceptual framework for personality–social, clinical, and health psychology. Psychol Bullet 1982;92(1):111-135. [doi: 10.1037/0033-2909.92.1.111]

11. Rivera DE, Pew MD, Collins LM. Using engineering control principles to inform the design of adaptive interventions: a conceptual introduction. Drug Alcohol Depend 2007 May;88 Suppl 2:31-40 [FREE Full text] [doi: 10.1016/j.drugalcdep.2006.10.020] [Medline: 17169503]
12. Conroy DE, Hojjatinia S, Lagoa CM, Yang C, Lanza ST, Smyth JM. Personalized models of physical activity responses to text message micro-interventions: a proof-of-concept application of control systems engineering methods. Psychol Sport Exerc 2019 Mar;41:172-180 [FREE Full text] [doi: 10.1016/j.psychsport.2018.06.011] [Medline: 30853855]
13. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
14. Daryabeygi-Khotbehsara R. Systematic review of smartphone-based interventions to reduce sedentary behaviour and promote physical activity using dynamic models. National Institute for Health Research. 2013. URL: https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=139350 [accessed 2020-01-22]
15. Higgins J, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Br Med J 2011 Oct 18;343:d5928. [doi: 10.1136/bmj.d5928] [Medline: 22008217]
16. Tufanaru C, Munn Z, Aromataris E, Campbell J, Hopp L. Chapter 3: Systematic reviews of effectiveness. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. University of Adelaide, South Australia: The Joanna Briggs Institute; 2020.
17. McKenzie JE, Brennan SE. Synthesizing and presenting findings using other methods. In: Cochrane Handbook for Systematic Reviews of Interventions. London: John Wiley & Sons Ltd; Sep 20, 2019:321-347.
18. Thomson HJ, Thomas S. The effect direction plot: visual display of non-standardised effects across multiple outcome domains. Res Synth Methods 2013 Mar 12;4(1):95-101 [FREE Full text] [doi: 10.1002/jrsm.1060] [Medline: 23795209]
19. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. Br Med J 2020 Jan 16;368:l6890 [FREE Full text] [doi: 10.1136/bmj.l6890] [Medline: 31948937]
20. Boon MH, Thomson H. The effect direction plot revisited: Application of the 2019 Cochrane Handbook guidance on alternative synthesis methods. Res Synth Methods 2021 Jan 05;12(1):29-33 [FREE Full text] [doi: 10.1002/jrsm.1458] [Medline: 32979023]
21. Klasnja P, Smith S, Seewald NJ, Lee A, Hall K, Luers B, et al. Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of HeartSteps. Ann Behav Med 2019 May 03;53(6):573-582 [FREE Full text] [doi: 10.1093/abm/kay067] [Medline: 30192907]
22. Martín CA, Rivera DE, Hekler EB, Riley WT, Buman MP, Adams MA, et al. Development of a control-oriented model of social cognitive theory for optimized mhealth behavioral interventions. IEEE Trans Control Syst Technol 2020 Mar;28(2):331-346 [FREE Full text] [doi: 10.1109/tcst.2018.2873538] [Medline: 33746479]
23. Adams MA, Sallis JF, Norman GJ, Hovell MF, Hekler EB, Perata E. An adaptive physical activity intervention for overweight adults: a randomized controlled trial. PLoS One 2013;8(12):e82901 [FREE Full text] [doi: 10.1371/journal.pone.0082901] [Medline: 24349392]
24. Bickmore TW, Schulman D, Sidner CL. A reusable framework for health counseling dialogue systems based on a behavioral medicine ontology. J Biomed Inform 2011 Apr;44(2):183-197 [FREE Full text] [doi: 10.1016/j.jbi.2010.12.006] [Medline: 21220044]
25. Hurling R, Catt M, Boni MD, Fairley BW, Hurst T, Murray P, et al. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. J Med Internet Res 2007 Apr 27;9(2):e7 [FREE Full text] [doi: 10.2196/jmir.9.2.e7] [Medline: 17478409]
26. Fukuoka Y, Kamitani E, Bonnet K, Lindgren T. Real-time social support through a mobile virtual community to improve healthy behavior in overweight and sedentary adults: a focus group analysis. J Med Internet Res 2011 Jul 14;13(3):e49 [FREE Full text] [doi: 10.2196/jmir.1770] [Medline: 21752785]
27. Hughes DC, Andrew A, Denning T, Hurvitz P, Lester J, Beresford S, et al. BALANCE (Bioengineering Approaches for Lifestyle Activity and Nutrition Continuous Engagement): developing new technology for monitoring energy balance in real time. J Diabetes Sci Technol 2010 Mar 01;4(2):429-434 [FREE Full text] [doi: 10.1177/193229681000400224] [Medline: 20307404]
28. Liao P, Dempsey W, Sarker H, Hossain SM, Al'absi M, Klasnja P, et al. Just-in-time but not too much: determining treatment timing in mobile health. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Dec;2(4):179 [FREE Full text] [doi: 10.1145/3287057] [Medline: 30801052]
29. Pirolli P, Youngblood GM, Du H, Konrad A, Nelson L, Springer A. Scaffolding the mastery of healthy behaviors with fittle+ systems: evidence-based interventions and theory. Hum Comp Interact 2018 Oct 15;36(2):73-106. [doi: 10.1080/07370024.2018.1512414]
30. Yang P, Stankevicius D, Marozas V, Deng Z, Liu E, Lukosevicius A, et al. Lifelogging data validation model for internet of things enabled personalized healthcare. IEEE Trans Syst Man Cybern Syst 2018 Jan;48(1):50-64. [doi: 10.1109/tsmc.2016.2586075]

31. Stroulia E, Fairbairn S, Bazelli B, Gibbs D, Lederer R, Faulkner R, et al. Smart-phone application design for lasting behavioral changes. In: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems. 2013 Presented at: 26th IEEE International Symposium on Computer-Based Medical Systems; June 20-22, 2013; Porto, Portugal. [doi: 10.1109/CBMS.2013.6627804]
32. Naved MM, Uddin MY. Adaptive notifications generation for smartphone users based on their physical activities. In: Proceedings of the 5th International Conference on Networking, Systems and Security (NSysS). 2018 Presented at: 5th International Conference on Networking, Systems and Security (NSysS); Dec. 18-20, 2018; Dhaka, Bangladesh. [doi: 10.1109/NSysS.2018.8631366]
33. Zheng H, Nugent C, McCullagh P, Huang Y, Zhang S, Burns W, et al. Smart self management: assistive technology to support people with chronic disease. J Telemed Telecare 2010;16(4):224-227. [doi: 10.1258/jtt.2010.004017] [Medline: 20511581]
34. Pirolli P. A computational cognitive model of self-efficacy and daily adherence in mHealth. Transl Behav Med 2016 Dec;6(4):496-508 [FREE Full text] [doi: 10.1007/s13142-016-0391-y] [Medline: 27848213]
35. Zhao ZW, Ma Q, Li CC, Liu L, Li WD. A machine-to-machine based framework for diabetes lifestyle management. In: Proceedings of the 10th IEEEE International Conference on Networking, Sensing and Control (ICNSC). 2013 Presented at: 10th IEEEE International Conference on Networking, Sensing and Control (ICNSC); April 10-12, 2013; Evry, France. [doi: 10.1109/ICNSC.2013.6548800]
36. Huang CY, Yang MC, Huang CY, Chen YJ, Wu ML, Chen KW. A chatbot-supported smart wireless interactive healthcare system for weight control and health promotion. In: Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). 2018 Presented at: IEEE International Conference on Industrial Engineering and Engineering Management (IEEM); Dec. 16-19, 2018; Bangkok, Thailand. [doi: 10.1109/IEEM.2018.8607399]
37. Müller AM, Blandford A, Yardley L. The conceptualization of a Just-In-Time Adaptive Intervention (JITAI) for the reduction of sedentary behavior in older adults. Mhealth 2017;3:37 [FREE Full text] [doi: 10.21037/mhealth.2017.08.05] [Medline: 29184889]
38. Mohadis HM, Mohamad Ali N. Designing persuasive application to encourage physical activity at workplace among older workers. In: Proceedings of the Sixth International Conference on Digital Information and Communication Technology and its Applications (DICTAP).: July 21-23, 2016; 2016 Presented at: Konya, Turkey; Konya, Turkey; Sixth International Conference on Digital Information and Communication Technology and its Applications (DICTAP) p. 1102-1114. [doi: 10.1109/DICTAP.2016.7544013]
39. Boulton E, Hawley-Hague H, French DP, Mellone S, Zacchi A, Clemson L, et al. Implementing behaviour change theory and techniques to increase physical activity and prevent functional decline among adults aged 61-70: The PreventIT project. Prog Cardiovasc Dis 2019;62(2):147-156 [FREE Full text] [doi: 10.1016/j.pcad.2019.01.003] [Medline: 30685469]
40. Martin SS, Feldman DI, Blumenthal RS, Jones SR, Post WS, McKibben RA, et al. mActive: A randomized clinical trial of an automated mhealth intervention for physical activity promotion. J Am Heart Assoc 2015 Nov 09;4(11):e002239 [FREE Full text] [doi: 10.1161/JAHA.115.002239] [Medline: 26553211]
41. Buman MP, Epstein DR, Gutierrez M, Herb C, Hollingshead K, Huberty JL, et al. BeWell24: development and process evaluation of a smartphone "app" to improve sleep, sedentary, and active behaviors in US Veterans with increased metabolic risk. Transl Behav Med 2016 Sep;6(3):438-448 [FREE Full text] [doi: 10.1007/s13142-015-0359-3] [Medline: 27528532]
42. Lim S, Kang SM, Kim KM, Moon JH, Choi SH, Hwang H, et al. Multifactorial intervention in diabetes care using real-time monitoring and tailored feedback in type 2 diabetes. Acta Diabetol 2016 Apr;53(2):189-198. [doi: 10.1007/s00592-015-0754-8] [Medline: 25936739]
43. Rawstorn J, Gant N, Meads A, Warren I, Maddison R. Remotely delivered exercise-based cardiac rehabilitation: design and content development of a novel mHealth platform. JMIR Mhealth Uhealth 2016 Jun 24;4(2):e57 [FREE Full text] [doi: 10.2196/mhealth.5501] [Medline: 27342791]
44. Vankipuram M, McMahon S, Fleury J. ReadySteady: app for accelerometer-based activity monitoring and wellness-motivation feedback system for older adults. AMIA Annu Symp Proc 2012;2012:931-939 [FREE Full text] [Medline: 23304368]
45. Martin CK, Gilmore LA, Apolzan JW, Myers CA, Thomas DM, Redman LM. Smartloss: A personalized mobile health intervention for weight management and health promotion. JMIR Mhealth Uhealth 2016 Mar 16;4(1):e18 [FREE Full text] [doi: 10.2196/mhealth.5027] [Medline: 26983937]
46. Ryan P, Papanek P, Csuka ME, Brown ME, Hopkins S, Lynch S, Striving to be Strong Team. Background and method of the Striving to be Strong study a RCT testing the efficacy of a m-health self-management intervention. Contemp Clin Trials 2018 Aug;71:80-87. [doi: 10.1016/j.cct.2018.06.006] [Medline: 29894865]
47. Clark DO, Srinivas P, Bodke K, Keith N, Hood S, Tu W. Addressing people and place microenvironments in weight loss disparities (APP-Me): Design of a randomized controlled trial testing timely messages for weight loss behavior in low income Black and White Women. Contemp Clin Trials 2018 Apr;67:74-80 [FREE Full text] [doi: 10.1016/j.cct.2018.01.006] [Medline: 29357313]

48. Thomas JG, Bond DS. Behavioral response to a just-in-time adaptive intervention (JITAI) to reduce sedentary behavior in obese adults: implications for JITAI optimization. Health Psychol 2015 Dec;34S:1261-1267 [FREE Full text] [doi: 10.1037/hea0000304] [Medline: 26651467]
49. Marquet O, Alberico C, Hipp A. Pokémon GO and physical activity among college students. A study using ecological momentary assessment. Comp Hum Behav 2018 Apr;81:215-222. [doi: 10.1016/j.chb.2017.12.028]
50. Srinivas P, Bodke K, Ofner S, Keith NR, Tu W, Clark DO. Context-sensitive ecological momentary assessment: application of user-centered design for improving user satisfaction and engagement during self-report. JMIR Mhealth Uhealth 2019 Apr 03;7(4):e10894 [FREE Full text] [doi: 10.2196/10894] [Medline: 30942698]
51. Cruciani F, Nugent C, Cleland I, McCullagh P. Rich context information for just-in-time adaptive intervention promoting physical activity. In: 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2017 Presented at: 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 11-15, 2017; Jeju, Korea (South). [doi: 10.1109/EMBC.2017.8036957]
52. Dunton GF, Dzubur E, Intille S. Feasibility and performance test of a real-time sensor-informed context-sensitive ecological momentary assessment to capture physical activity. J Med Internet Res 2016 Jun 01;18(6):e106 [FREE Full text] [doi: 10.2196/jmir.5398] [Medline: 27251313]
53. Dunton GF, Dzubur E, Kawabata K, Yanez B, Bo B, Intille S. Development of a smartphone application to measure physical activity using sensor-assisted self-report. Front Public Health 2014;2:12 [FREE Full text] [doi: 10.3389/fpubh.2014.00012] [Medline: 24616888]
54. Dunton GF, Liao Y, Kawabata K, Intille S. Momentary assessment of adults' physical activity and sedentary behavior: feasibility and validity. Front Psychol 2012;3:260 [FREE Full text] [doi: 10.3389/fpsyg.2012.00260] [Medline: 22866046]
55. Rofey DL, Hull EE, Phillips J, Vogt K, Silk JS, Dahl RE. Utilizing Ecological Momentary Assessment in pediatric obesity to quantify behavior, emotion, and sleep. Obesity (Silver Spring) 2010 Jun;18(6):1270-1272 [FREE Full text] [doi: 10.1038/oby.2009.483] [Medline: 20019675]
56. Smith A, de Salas K, Lewis I, Schüz B. Developing smartphone apps for behavioural studies: The AlcoRisk app case study. J Biomed Inform 2017 Aug;72:108-119 [FREE Full text] [doi: 10.1016/j.jbi.2017.07.007] [Medline: 28709856]
57. Jones KK, Zenk SN, McDonald A, Corte C. Experiences of African-American women with smartphone-based ecological momentary assessment. Public Health Nurs 2016 Jul;33(4):371-380. [doi: 10.1111/phn.12239] [Medline: 26530781]
58. Li Z, Das S, Codella JW, Hao T, Lin K, Maduri C, et al. An adaptive, data-driven personalized advisor for increasing physical activity. IEEE J Biomed Health Inform 2019 May;23(3):999-1010. [doi: 10.1109/JBHI.2018.2879805] [Medline: 30418890]
59. Fahim M, Baker T, Khattak AM, Shah B, Aleem S, Chow F. Context mining of sedentary behaviour for promoting self-awareness using a smartphone. Sensors (Basel) 2018 Mar 15;18(3):874 [FREE Full text] [doi: 10.3390/s18030874] [Medline: 29543763]
60. Brannon EE, Cushing CC, Crick CJ, Mitchell TB. The promise of wearable sensors and ecological momentary assessment measures for dynamical systems modeling in adolescents: a feasibility and acceptability study. Transl Behav Med 2016 Dec;6(4):558-565 [FREE Full text] [doi: 10.1007/s13142-016-0442-4] [Medline: 27678501]
61. Valentiner LS, Thorsen IK, Kongstad MB, Brinkløv CF, Larsen RT, Karstoft K, et al. Effect of ecological momentary assessment, goal-setting and personalized phone-calls on adherence to interval walking training using the InterWalk application among patients with type 2 diabetes-A pilot randomized controlled trial. PLoS One 2019;14(1):e0208181 [FREE Full text] [doi: 10.1371/journal.pone.0208181] [Medline: 30629601]
62. Alian S, Li J, Pandey V. A personalized recommendation system to support diabetes self-management for American Indians. IEEE Access 2018;6:73041-73051. [doi: 10.1109/ACCESS.2018.2882138]
63. Hales S, Turner-McGrievy GM, Wilcox S, Fahim A, Davis R, Huhns M, et al. Social networks for improving healthy weight loss behaviors for overweight and obese adults: a randomized clinical trial of the social pounds off digitally (Social POD) mobile app. Int J Med Inform 2016 Oct;94:81-90. [doi: 10.1016/j.ijmedinf.2016.07.003] [Medline: 27573315]
64. Salvi D, Ottaviano M, Muuraiskangas S, Martínez-Romero A, Vera-Muñoz C, Triantafyllidis A, et al. An m-Health system for education and motivation in cardiac rehabilitation: the experience of HeartCycle guided exercise. J Telemed Telecare 2018 May;24(4):303-316. [doi: 10.1177/1357633X17697501] [Medline: 28350282]
65. Tabak M, op den Akker H, Hermens H. Motivational cues as real-time feedback for changing daily activity behavior of patients with COPD. Patient Educ Couns 2014 Mar;94(3):372-378. [doi: 10.1016/j.pec.2013.10.014] [Medline: 24332934]
66. Maddison R, Rawstorn JC, Stewart RA, Benatar J, Whittaker R, Rolleston A, et al. Effects and costs of real-time cardiac telerehabilitation: randomised controlled non-inferiority trial. Heart 2019 Jan;105(2):122-129 [FREE Full text] [doi: 10.1136/heartjnl-2018-313189] [Medline: 30150328]
67. Kramer JN, Künzler F, Mishra V, Presset B, Kotz D, Smith S, et al. Investigating intervention components and exploring states of receptivity for a smartphone app to promote physical activity: protocol of a microrandomized trial. JMIR Res Protoc 2019 Jan 31;8(1):e11540 [FREE Full text] [doi: 10.2196/11540] [Medline: 30702430]
68. Militello L, Melnyk BM, Hekler EB, Small L, Jacobson D. Automated behavioral text messaging and face-to-face intervention for parents of overweight or obese preschool children: results from a pilot study. JMIR Mhealth Uhealth 2016 Mar 14;4(1):e21 [FREE Full text] [doi: 10.2196/mhealth.4398] [Medline: 26976387]

69. Brakenridge CL, Healy GN, Winkler EA, Fjeldsoe BS. Usage, acceptability, and effectiveness of an activity tracker in a randomized trial of a workplace sitting intervention: mixed-methods evaluation. Interact J Med Res 2018 Mar 02;7(1):e5 [FREE Full text] [doi: 10.2196/ijmr.9001] [Medline: 29500158]
70. Yom-Tov E, Feraru G, Kozdoba M, Mannor S, Tennenholtz M, Hochberg I. Encouraging physical activity in patients with diabetes: intervention using a reinforcement learning system. J Med Internet Res 2017 Oct 10;19(10):e338 [FREE Full text] [doi: 10.2196/jmir.7994] [Medline: 29017988]
71. Lin PH, Intille S, Bennett G, Bosworth HB, Corsino L, Voils C, et al. Adaptive intervention design in mobile health: Intervention design and development in the Cell Phone Intervention for You trial. Clin Trials 2015 Dec;12(6):634-645 [FREE Full text] [doi: 10.1177/1740774515597222] [Medline: 26229119]
72. Alexander B, Karakas K, Kohout C, Sakarya H, Singh N, Stachtiaris J, et al. A behavioral sensing system that promotes positive lifestyle changes and improves metabolic control among adults with type 2 diabetes. In: Proceedings of the Systems and Information Engineering Design Symposium (SIEDS). 2017 Presented at: Systems and Information Engineering Design Symposium (SIEDS); April 28-28, 2017; Charlottesville, VA, USA. [doi: 10.1109/SIEDS.2017.7937732]
73. Al-Ozairi E, Ridge K, Taghadom E, de Zoysa N, Tucker C, Stewart K, et al. Diabetes and TelecommunicationS (DATES) study to support self-management for people with type 2 diabetes: a randomized controlled trial. BMC Public Health 2018 Nov 12;18(1):1249 [FREE Full text] [doi: 10.1186/s12889-018-6136-8] [Medline: 30419893]
74. Baert A, Clays E, Bolliger L, De Smedt D, Lustrek M, Vodopija A, et al. A Personal Decision Support System for Heart Failure Management (HeartMan): study protocol of the HeartMan randomized controlled trial. BMC Cardiovasc Disord 2018 Sep 27;18(1):186. [doi: 10.1186/s12872-018-0921-2]
75. Bardus M, Hamadeh G, Hayek B, Al Kherfan R. A self-directed mobile intervention (WaznApp) to promote weight control among employees at a Lebanese University: protocol for a feasibility pilot randomized controlled trial. JMIR Res Protoc 2018 May 16;7(5):e133 [FREE Full text] [doi: 10.2196/resprot.9793] [Medline: 29769174]
76. Evans EH, Araújo-Soares V, Adamson A, Batterham AM, Brown H, Campbell M, et al. The NULevel trial of a scalable, technology-assisted weight loss maintenance intervention for obese adults after clinically significant weight loss: study protocol for a randomised controlled trial. Trials 2015 Sep 22;16:421 [FREE Full text] [doi: 10.1186/s13063-015-0931-7] [Medline: 26395774]
77. Goldstein SP, Thomas JG, Vithiananthan S, Blackburn GA, Jones DB, Webster J, et al. Multi-sensor ecological momentary assessment of behavioral and psychosocial predictors of weight loss following bariatric surgery: study protocol for a multicenter prospective longitudinal evaluation. BMC Obes 2018;5:27 [FREE Full text] [doi: 10.1186/s40608-018-0204-6] [Medline: 30410772]
78. Hassoon A, Schrack J, Naiman D, Lansey D, Baig Y, Stearns V, et al. Increasing physical activity amongst overweight and obese cancer survivors using an alexa-based intelligent agent for patient coaching: protocol for the Physical Activity by Technology Help (PATH) trial. JMIR Res Protoc 2018 Feb 12;7(2):e27 [FREE Full text] [doi: 10.2196/resprot.9096] [Medline: 29434016]
79. Hurley JC, Hollingshead KE, Todd M, Jarrett CL, Tucker WJ, Angadi SS, et al. The Walking Interventions Through Texting (WalkIT) Trial: Rationale, design, and protocol for a factorial randomized controlled trial of adaptive interventions for overweight and obese, inactive adults. JMIR Res Protoc 2015 Sep 11;4(3):e108 [FREE Full text] [doi: 10.2196/resprot.4856] [Medline: 26362511]
80. Orme MW, Weedon AE, Esliger DW, Saukko PM, Morgan MD, Steiner MC, et al. Study protocol for Chronic Obstructive Pulmonary Disease-Sitting and ExacerbAtions Trial (COPD-SEAT): a randomised controlled feasibility trial of a home-based self-monitoring sedentary behaviour intervention. BMJ Open 2016 Oct 03;6(10):e013014 [FREE Full text] [doi: 10.1136/bmjopen-2016-013014] [Medline: 27697880]
81. Pellegrini CA, Hoffman SA, Collins LM, Spring B. Optimization of remotely delivered intensive lifestyle treatment for obesity using the Multiphase Optimization Strategy: Opt-IN study protocol. Contemp Clin Trials 2014 Jul;38(2):251-259 [FREE Full text] [doi: 10.1016/j.cct.2014.05.007] [Medline: 24846621]
82. Phillips SM, Collins LM, Penedo FJ, Courneya KS, Welch W, Cottrell A, et al. Optimization of a technology-supported physical activity intervention for breast cancer survivors: Fit2Thrive study protocol. Contemp Clin Trials 2018 Mar;66:9-19 [FREE Full text] [doi: 10.1016/j.cct.2018.01.001] [Medline: 29330081]
83. Downs DS, Savage JS, Rivera DE, Smyth JM, Rolls BJ, Hohman EE, et al. Individually tailored, adaptive intervention to manage gestational weight gain: protocol for a randomized controlled trial in women with overweight and obesity. JMIR Res Protoc 2018 Jun 08;7(6):e150 [FREE Full text] [doi: 10.2196/resprot.9220] [Medline: 29884603]
84. Vidmar AP, Salvy SJ, Pretlow R, Mittelman SD, Wee C, Fink C, et al. An addiction-based mobile health weight loss intervention: protocol of a randomized controlled trial. Contemp Clin Trials 2019 Mar;78:11-19 [FREE Full text] [doi: 10.1016/j.cct.2019.01.008] [Medline: 30654026]
85. Block G, Azar KM, Romanelli RJ, Block TJ, Hopkins D, Carpenter HA, et al. Diabetes prevention and weight loss with a fully automated behavioral intervention by email, web, and mobile phone: a randomized controlled trial among persons with prediabetes. J Med Internet Res 2015 Oct 23;17(10):e240 [FREE Full text] [doi: 10.2196/jmir.4897] [Medline: 26499966]

86. Baretta D, Sartori F, Greco A, D'Addario M, Melen R, Steca P. Improving physical activity mhealth interventions: development of a computational model of self-efficacy theory to define adaptive goals for exercise promotion. Adv Hum Comp Interact 2019 Mar 04;2019:1-11. [doi: 10.1155/2019/3068748]
87. Direito A, Tooley M, Hinbarji M, Albatal R, Jiang Y, Whittaker R, et al. Tailored daily activity: an adaptive physical activity smartphone intervention. Telemed J E Health 2020 Apr 01;26(4):426-437. [doi: 10.1089/tmj.2019.0034] [Medline: 31063038]
88. Direito A, Walsh D, Hinbarji M, Albatal R, Tooley M, Whittaker R, et al. Using the intervention mapping and behavioral intervention technology frameworks: development of an mhealth intervention for physical activity and sedentary behavior change. Health Educ Behav 2018 Jun 07;45(3):331-348. [doi: 10.1177/1090198117742438] [Medline: 29216765]
89. Middelweerd A, Mollee J, Klein MM, Manzoor A, Brug J, Te Velde SJ. The use and effects of an app-based physical activity intervention "Active2Gether" in young adults: quasi-experimental trial. JMIR Form Res 2020 Jan 21;4(1):e12538 [FREE Full text] [doi: 10.2196/12538] [Medline: 31961330]
90. Klein M, Manzoor A, Mollee J. Active2Gether: A personalized m-health intervention to encourage physical activity. Sensors (Basel) 2017 Jun 19;17(6):1436 [FREE Full text] [doi: 10.3390/s17061436] [Medline: 28629178]
91. Middelweerd A, Te Velde SJ, Mollee JS, Klein MC, Brug J. App-based intervention combining evidence-based behavior change techniques with a model-based reasoning system to promote physical activity among young adults (Active2Gether): descriptive study of the development and content. JMIR Res Protoc 2018 Dec 21;7(12):e185 [FREE Full text] [doi: 10.2196/resprot.7169] [Medline: 30578198]
92. Korinek EV, Phatak SS, Martin CA, Freigoun MT, Rivera DE, Adams MA, et al. Adaptive step goals and rewards: a longitudinal growth model of daily steps for a smartphone-based walking intervention. J Behav Med 2018 Feb 16;41(1):74-86. [doi: 10.1007/s10865-017-9878-3] [Medline: 28918547]
93. Freigoun M, Martin C, Magann A, Rivera DE, Phatak S, Korinek E, et al. System identification of Just Walk: A behavioral mHealth intervention for promoting physical activity. In: Proceedings of the American Control Conference (ACC). 2017 May 24 Presented at: American Control Conference (ACC); May 24-26, 2017; Seattle, WA, USA. [doi: 10.23919/acc.2017.7962940]
94. Rabbi M, Pfammatter A, Zhang M, Spring B, Choudhury T. Automated personalized feedback for physical activity and dietary behavior change with mobile phones: a randomized controlled trial on adults. JMIR Mhealth Uhealth 2015 May 14;3(2):e42 [FREE Full text] [doi: 10.2196/mhealth.4160] [Medline: 25977197]
95. Rabbi M, Aung MS, Gay G, Reid MC, Choudhury T. Feasibility and acceptability of mobile phone-based auto-personalized physical activity recommendations for chronic pain self-management: pilot study on adults. J Med Internet Res 2018 Oct 26;20(10):e10147 [FREE Full text] [doi: 10.2196/10147] [Medline: 30368433]
96. Zhou M, Fukuoka Y, Mintz Y, Goldberg K, Kaminsky P, Flowers E, et al. Evaluating machine learning-based automated personalized daily step goals delivered through a mobile phone app: randomized controlled trial. JMIR Mhealth Uhealth 2018 Jan 25;6(1):e28 [FREE Full text] [doi: 10.2196/mhealth.9117] [Medline: 29371177]
97. Ghahramani Z. Learning dynamic Bayesian networks. In: Adaptive Processing of Sequences and Data Structures. Berlin, Heidelberg: Springer; 1998.
98. Riley W, Cesar A, Martin CA, Rivera DE. The importance of behavior theory in control system modeling of physical activity sensor data. In: Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2014 Presented at: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Aug. 26-30, 2014; Chicago, IL, USA p. 6880-6883 URL: http://europepmc.org/abstract/MED/25571577 [doi: 10.1109/EMBC.2014.6945209]
99. Nahum-Shani I. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2017 Dec 12;52(6):446-462. [doi: 10.1007/s12160-016-9830-8]
100. Ding X, Xu J, Wang H, Chen G, Herpreet T, Zhang Y. WalkMore: Promoting walking with just-in-time context-aware prompts. In: Proceedings of the IEEE Wireless Health (WH). 2016 Presented at: IEEE Wireless Health (WH); October 25-27, 2016; Bethesda, MD, USA. [doi: 10.1109/WH.2016.7764558]
101. Lin Y. Motivate: A Context-Aware Mobile Application for Physical Activity Promotion. Eindhoven: Technology University Eindhoven; Feb 01, 2013.
102. Minch RP. Privacy issues in location-aware mobile devices. In: Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 2004. 2004 Jan 05 Presented at: 37th Annual Hawaii International Conference on System Sciences, 2004; January 5-8, 2004; Big Island, HI, USA. [doi: 10.1109/hicss.2004.1265320]
103. Satyanarayanan M. Pervasive computing: vision and challenges. IEEE Pers Commun 2020;8(4):10-17. [doi: 10.1109/98.943998]
104. Lee H, Kobsa A. Understanding user privacy in Internet of Things environments. In: Proceedings of the IEEE 3rd World Forum on Internet of Things (WF-IoT). 2016 Dec 12 Presented at: IEEE 3rd World Forum on Internet of Things (WF-IoT); Dec. 12-14, 2016; Reston, VA, USA. [doi: 10.1109/wf-iot.2016.7845392]
105. Clemes S, David BM, Zhao Y, Han X, Brown W. Validity of two self-report measures of sitting time. J Phys Act Health 2012 May;9(4):533-539. [doi: 10.1123/jpah.9.4.533] [Medline: 21946087]

106. Gibbs B, Hergenroeder AL, Katzmarzyk PT, Lee IM, Jakicic JM. Definition, measurement, and health risks associated with sedentary behavior. Med Sci Sports Exerc 2015 Jun;47(6):1295-1300. [doi: 10.1249/MSS.0000000000000517] [Medline: 25222816]
107. Edwardson CL, Winkler EA, Bodicoat DH, Yates T, Davies MJ, Dunstan DW, et al. Considerations when using the activPAL monitor in field-based research with adult populations. J Sport Health Sci 2017 Jun;6(2):162-178 [FREE Full text] [doi: 10.1016/j.jshs.2016.02.002] [Medline: 30356601]
108. Henson J, Davies MJ, Bodicoat DH, Edwardson CL, Gill JM, Stensel DJ, et al. Breaking up prolonged sitting with standing or walking attenuates the postprandial metabolic response in postmenopausal women: a randomized acute study. Diabetes Care 2016 Jan;39(1):130-138. [doi: 10.2337/dc15-1240] [Medline: 26628415]
109. Buckley JP, Mellor DD, Morris M, Joseph F. Standing-based office work shows encouraging signs of attenuating post-prandial glycaemic excursion. Occup Environ Med 2014 Feb 02;71(2):109-111. [doi: 10.1136/oemed-2013-101823] [Medline: 24297826]
110. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. Health Psychol 2015 Dec;34S(Suppl):1220-1228 [FREE Full text] [doi: 10.1037/hea0000305] [Medline: 26651463]
111. Cradock KA, ÓLaighin G, Finucane FM, Gainforth HL, Quinlan LR, Ginis KA. Behaviour change techniques targeting both diet and physical activity in type 2 diabetes: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2017 Feb 08;14(1):18 [FREE Full text] [doi: 10.1186/s12966-016-0436-0] [Medline: 28178985]

Abbreviations

mHealth: mobile health
MVPA: moderate-to-vigorous physical activity
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
SB: sedentary behavior
SCT: social cognitive theory

Edited by R Kukafka; submitted 07.12.20; peer-reviewed by R Krukowski, F Tabei; comments to author 16.12.20; revised version received 29.12.20; accepted 30.04.21; published 13.09.21.

Please cite as:
Daryabeygi-Khotbehsara R, Shariful Islam SM, Dunstan D, McVicar J, Abdelrazek M, Maddison R
Smartphone-Based Interventions to Reduce Sedentary Behavior and Promote Physical Activity Using Integrated Dynamic Models: Systematic Review
J Med Internet Res 2021;23(9):e26315
URL: https://www.jmir.org/2021/9/e26315
doi:10.2196/26315
PMID:34515637

Review

Promoting Physical Activity Through Conversational Agents: Mixed Methods Systematic Review

Tiffany Christina Luo[1], MSW; Adrian Aguilera[1,2,3], PhD; Courtney Rees Lyles[3,4,5,6], PhD; Caroline Astrid Figueroa[1], MD, PhD

[1]School of Social Welfare, University of California, Berkeley, Berkeley, CA, United States

[2]Department of Psychiatry, Zuckerberg San Francisco General Hospital, University of California, San Francisco, San Francisco, CA, United States

[3]Center for Vulnerable Populations, Zuckerberg San Francisco General Hospital, University of California, San Francisco, San Francisco, CA, United States

[4]Department of Epidemiology & Biostatistics, University of California, San Francisco, San Francisco, CA, United States

[5]Department of Medicine, University of California, San Francisco, San Francisco, CA, United States

[6]School of Public Health, University of California, Berkeley, Berkeley, CA, United States

Corresponding Author:
Tiffany Christina Luo, MSW
School of Social Welfare
University of California, Berkeley
Haviland Hall
Berkeley, CA, 94720-7400
United States
Phone: 1 650 228 3514
Email: tiffany.luo@berkeley.edu

Abstract

Background: Regular physical activity (PA) is crucial for well-being; however, healthy habits are difficult to create and maintain. Interventions delivered via conversational agents (eg, chatbots or virtual agents) are a novel and potentially accessible way to promote PA. Thus, it is important to understand the evolving landscape of research that uses conversational agents.

Objective: This mixed methods systematic review aims to summarize the usability and effectiveness of conversational agents in promoting PA, describe common theories and intervention components used, and identify areas for further development.

Methods: We conducted a mixed methods systematic review. We searched seven electronic databases (PsycINFO, PubMed, Embase, CINAHL, ACM Digital Library, Scopus, and Web of Science) for quantitative, qualitative, and mixed methods studies that conveyed primary research on automated conversational agents designed to increase PA. The studies were independently screened, and their methodological quality was assessed using the Mixed Methods Appraisal Tool by 2 reviewers. Data on intervention impact and effectiveness, treatment characteristics, and challenges were extracted and analyzed using parallel-results convergent synthesis and narrative summary.

Results: In total, 255 studies were identified, 7.8% (20) of which met our inclusion criteria. The methodological quality of the studies was varied. Overall, conversational agents had moderate usability and feasibility. Those that were evaluated through randomized controlled trials were found to be effective in promoting PA. Common challenges facing interventions were repetitive program content, high attrition, technical issues, and safety and privacy concerns.

Conclusions: Conversational agents hold promise for PA interventions. However, there is a lack of rigorous research on long-term intervention effectiveness and patient safety. Future interventions should be based on evidence-informed theories and treatment approaches and should address users' desires for program variety, natural language processing, delivery via mobile devices, and safety and privacy concerns.

(J Med Internet Res 2021;23(9):e25486) doi:10.2196/25486

KEYWORDS

physical activity; health behavior; behavior change; conversational agent; virtual agent; chatbot; digital health; eHealth; mHealth; mobile health; mobile phone

Introduction

Background

Physical activity (PA) is crucial to health and well-being, and regular exercise can reduce the risk of disease, improve mental health, and boost quality of life [1]. In 2016, 28% of adults globally did not meet the World Health Organization's PA guidelines for 150 minutes of aerobic activity per week [2]. Global PA levels have not improved since 2001, and the prevalence of inactivity has steadily risen in high-income countries [2]. Therefore, innovative interventions are required to increase PA.

Recently, there has been an increase in digital health interventions that promote healthy lifestyle changes through technologies such as smartphone apps, web-based programs, and text messages [3]. Some of these interventions are as effective as in-person interventions at modifying behavior [4]. Programs may include virtual health coaching, workout or diet plans, progress monitoring, and positive reinforcement for healthy eating and PA. Tailored feedback based on individual goals, habits, and circumstances can create a more personalized experience for users. Furthermore, some digital platforms offer users the option of pairing activity trackers such as pedometers, accelerometers, and heart rate monitors to improve the accuracy of data tracking and performance feedback.

In addition to their customizability, digital interventions allow health programs to have a wide reach. In 2018, mobile phone ownership rates ranged from 83% in emerging economies to >90% in advanced economies worldwide [5]. Smartphone ownership and internet use are nearly universal in most advanced economies and continue to grow rapidly in emerging economies [5]. With the advent of technology, demographic groups that previously did not have access to health coaching because of prohibitive costs can now access that support. Low-income Hispanic adults and Black adults in the United States, in particular, may benefit, as they have a significantly higher prevalence of physical inactivity than non-Hispanic White adults [6]. Smartphone ownership and use are more common in Hispanic and Black households than in non-Hispanic White households [7], making mobile platforms suitable for disseminating health-related interventions to underserved communities.

Digital interventions can take the form of a conversational agent, also known as a chatbot or virtual agent. Conversational agents are software programs that mimic written or spoken human conversations. They come in many forms, from chatbots engaging in written conversations to avatars simulating face-to-face discussions through synthetic speech [8]. Depending on their form, conversational agents may be deployed through standalone computer software, messaging apps, web-based platforms, mobile apps, and SMS text messaging or multimedia messaging services (MMSs). Interacting with conversational agents typically does not require much digital literacy beyond chatting or typing.

Simple conversational agents operate according to expert systems or rule-based systems, meaning they generate conversations based on questions and responses written by program developers [9]. In such cases, users are often restricted to selecting predefined answers. Conversational agents with more advanced capabilities are programmed to conduct natural language processing and integrate machine learning. Users are free to enter any command, and conversational agents formulate appropriate responses based on artificial intelligence algorithms.

Conversational agents have been increasingly used in the health care sector to help patients achieve their health goals, owing to their ability to provide interactive and personalized content [8]. Many of these conversational agents provide daily feedback, encouragement, and adaptive goals based on objective data received from fitness trackers. In contrast to in-person health coaching, conversational agents can be accessed around the clock for the duration of the intervention.

An example of a conversational agent that supports individuals in reaching their health goals is Ally, a smartphone-based chatbot that incorporates self-monitoring prompts, exercise planning, and financial incentives (cash and donations to a charity organization) to motivate users to walk more [10]. Another example, FitChat, uses goal setting, discussions of barriers, and motivational messages to encourage older adults to engage in aerobic activity and muscle-strengthening exercises [11]. A third example, Laura, falls into the subset of conversational agents termed *relational agents* [12-14]. Relational agents are computational artifacts, often with humanlike appearance and speech, designed to establish social-emotional relationships with users [12]. Relational agents such as Laura use social dialog, empathy, humor, and self-disclosure to keep users engaged over time and motivate them to create and maintain exercise habits [12].

Rationale

Systematic and scoping reviews have been conducted on the use of digital interventions to increase PA [15-18] and the use of conversational agents in health care [8,19-21]. Previous reviews have found that many digital interventions are not theoretically based or evidence informed [4]. These interventions may be limited in their impact, as they do not include established constructs for behavior change. Although there is emerging evidence that most behavior change interventions are suitable for adaptation to a digital platform [22], few studies have addressed how digital content is linked to empirically tested frameworks and how program content and dialog flows are translated from face-to-face to virtual delivery.

It is unknown whether previous findings extend to PA conversational agents. To our knowledge, no systematic reviews have focused exclusively on PA conversational agents and analyzed their use of theories, treatment approaches, and intervention components. Research in this domain may help elucidate the successes and shortcomings of current interventions, thus guiding the development of program content and dialog flows that will have maximum impact on users.

Objectives

Our objective is to conduct a systematic review to (1) summarize the usability and effectiveness of PA conversational agents; (2) describe common theoretical frameworks, treatment approaches,

and intervention techniques; and (3) identify areas for further development.

Methods

Overview

We conducted a mixed methods systematic review following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines [23] (Multimedia Appendix 1 [23]). The protocol for this systematic review was registered on the Open Science Framework registries [24].

We chose a mixed methods systematic review as conversational agents are still relatively new. As such, there is a shortage of randomized controlled trials (RCTs) investigating their efficacy and effectiveness in the health care sector [8]. Many studies of conversational agents include both quantitative data (eg, step counts and participant ratings on Likert scales) and qualitative data (eg, quotes from individual interviews or focus group sessions); a mixed methods design produces a more comprehensive overview of conversational agents than synthesizing quantitative or qualitative data only.

Eligibility Criteria

The formulation of the eligibility criteria was based on the PICOS (patient problem, intervention, comparison, outcomes, and studies) framework (Textbox 1) [25].

Textbox 1. Inclusion and exclusion criteria using the PICOS (patient problem, intervention, comparison, outcomes, and studies) framework.

Inclusion criteria

- Patient problem: studies that targeted physical activity in users
- Intervention: interventions that involved an automated conversational agent
- Comparison: another intervention type or delivery method (eg, face-to-face and app), treatment as usual, no treatment, or one group pre-post comparison
- Outcomes: reporting of intervention impact on participants or participants' experiences with the conversational agent; some description of theoretical basis, dialog flow development, or intervention components of the program
- Study type: quantitative, qualitative, and mixed methods studies

Exclusion criteria

- Patient problem: studies that did not target physical activity in users
- Intervention: interventions that did not involve an automated conversational agent
- Comparison: studies without a comparison condition were not excluded, provided they still included sufficient outcome data
- Outcomes: no mention of intervention impact or participant experiences; no description of the applied intervention
- Study type: literature reviews, conference abstracts, dissertations, protocol papers, and tutorials

The inclusion criteria for this review included primary literature that involved an automated conversational agent. We focused on studies describing existing conversational agents, as opposed to studies exploring hypothetical uses of conversational agents, in an attempt to present concrete findings with external validity. We did not place any limitations on the conversational agent type, delivery platform, dialog technique, or input and output modalities. PA had to be one of the targets of the intervention. No restrictions were imposed on the target population or setting.

Studies were excluded if there was no primary research conducted or if the intervention did not use an automated conversational agent to target PA. Studies were not excluded for the lack of a comparison condition, provided they still offered outcome data on intervention impact or participant experiences and described the intervention in sufficient detail. Protocol papers and tutorials on building conversational interfaces were excluded as they did not provide any outcome data.

Information Sources

We searched seven relevant electronic databases (PsycINFO, PubMed, Embase, CINAHL, ACM Digital Library, Scopus, and Web of Science) from their inception through July 22, 2020. We also reviewed the reference lists of relevant papers.

Search Strategy

We based our search strategy on a preliminary scan of the literature on digital health interventions. We also consulted a librarian at the University of California, Berkeley, to generate search strings for selected databases, using Boolean operators and thesaurus terms where applicable. We combined search terms for two major topic areas: conversational agents and PA (complete search strategy available in Multimedia Appendix 2).

Study Selection

One author conducted the initial search in each database and imported all references into Covidence (Veritas Health Innovation), a web-based software program that facilitates collaboration among reviewers. Duplicate records were identified and removed.

The titles and abstracts of all the citations were independently screened by 2 authors for eligibility. Potentially relevant articles were retrieved in full for review. Full-text studies that did not meet the predefined eligibility criteria were excluded. Any

discrepancies regarding the inclusion of an article were resolved through discussion between the 2 reviewers. Cohen κ was calculated to measure intercoder agreement.

Data Management and Collection

Data from the selected studies were charted in a spreadsheet developed by the authors for this review (Multimedia Appendix 3). Data extraction was performed by one reviewer, with a second reviewer cross-checking the data extraction table for accuracy.

Data Items

Descriptive Data

The following descriptive data were extracted from each study: authors, publication year, title, study design, targeted behaviors (in addition to PA), population (eg, clinical vs nonclinical samples), geographic focus, initial and final sample size, conversational agent name, conversational agent type, delivery method, delivery platform, conversational agent output modality, user input modality, comparison conditions, control type, and outcome measures. Data were also analyzed for the variables given in the following sections.

Intervention Effectiveness and Impact

Evaluation measures for assessing changes in users' activity levels or motivation to exercise as a result of the intervention included data derived from subjective measures (eg, questionnaires and self-reports) and objective measures (eg, pedometers).

Theory

Theories attempt to explain how and why a behavior occurs. Theoretical frameworks may guide the design and selection of the program content. In addition, the integration of theoretical content may boost the effectiveness of behavior change interventions [4]. Examples of established theories of PA promotion that have guided some of the interventions discussed in this review include behavior change theory, the habit formation model, and the health action process approach.

Dialog Flow Development

Dialog flows for conversational agents are often adapted from counseling techniques for a specific treatment approach, such as motivational interviewing or cognitive behavioral therapy. These approaches can help enhance motivation for behavior change and identify barriers to PA.

Intervention Components

Conversational agents implement specific program elements to help users overcome exercise barriers and increase their activity levels. Examples include health education, self-monitoring, goal setting, and exercise reminders.

Challenges and Areas for Improvement

Study limitations, ethical considerations, barriers to program development or implementation, and key areas for improving the conversational agent were noted.

Outcomes and Prioritization

The primary outcomes for which we collected data were (1) usability and effectiveness of PA conversational agents; (2) theories, intervention components, and cognitive and behavioral constructs used to motivate individuals to engage in PA; and (3) challenges and areas for improvement. Quantitative and qualitative data were collected to assess the outcomes.

Appraisal of Studies

The methodological quality of the included studies was assessed using the Mixed Methods Appraisal Tool (MMAT) [26]. The MMAT is a valid, reliable, and efficient tool that allows the simultaneous appraisal of qualitative, quantitative, and mixed methods studies [27]. The methods section of each included study was read by 2 reviewers independently, and each study was categorized as qualitative research, RCT, nonrandomized study, quantitative descriptive study, or mixed methods study. Then, studies were rated based on their fulfillment of the MMAT criteria in each of their respective categories. Examples of methodological quality indicators include the appropriateness of study design, choice of sampling strategy, adherence to data collection methods, intervention integrity, and integration of results. Any disagreements on ratings were resolved through discussion between the 2 reviewers.

Assigning studies an overall numerical score based on the ratings of each criterion is discouraged because a single number cannot provide insight into which aspects of the study methodology are problematic [26]. Instead, we classified studies as having lower methodological quality when they met ≤60% of the MMAT criteria and higher quality when they met >60% of the criteria. In addition, we included a detailed overview of our ratings of each criterion. All eligible studies were discussed in this review regardless of their MMAT ratings, as it is discouraged to exclude studies on the basis of low methodological quality [28].

Data Synthesis

A meta-analysis was not conducted because of the heterogeneity of study types and outcome data. Instead, data were analyzed using parallel-results convergent synthesis, which allows qualitative and quantitative evidence to be synthesized concurrently, without data transformation [29]. Parallel-results convergent synthesis is suitable for systematic reviews that pose two or more complementary review questions [29]. Following evidence synthesis, we presented a narrative summary of our findings and made recommendations for future work.

Results

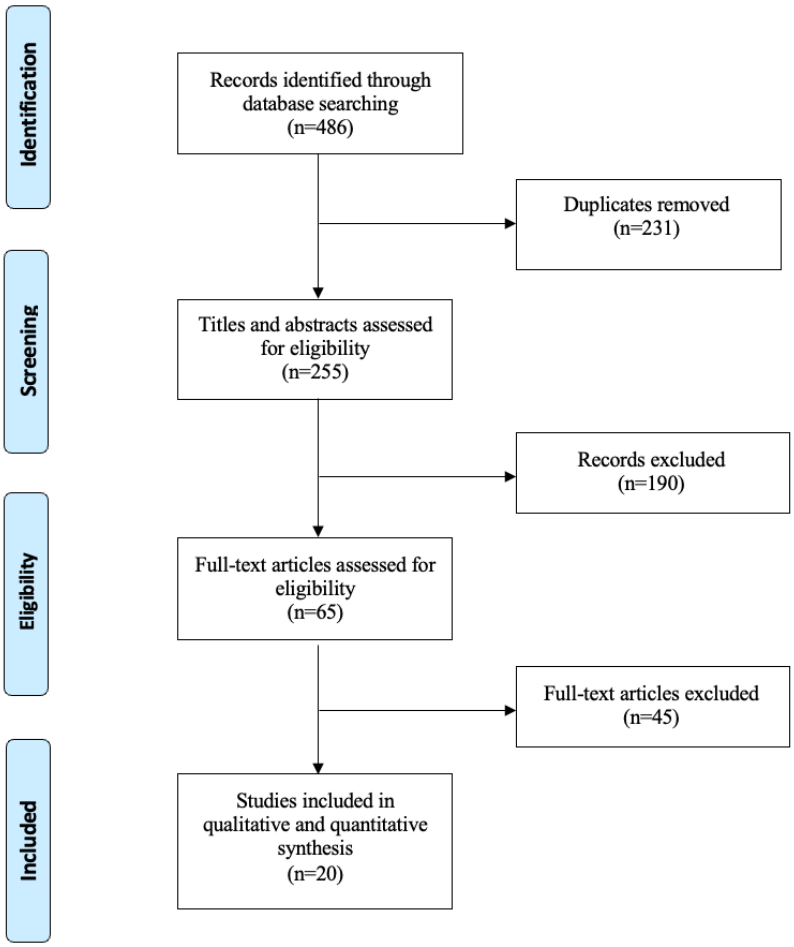
Search Results

Our literature search retrieved 486 citations. After the removal of duplicates, 255 studies remained. An additional 74.5% (190/255) of studies were excluded after the title and abstract screening. Of the 65 remaining studies, 20 (31%) were selected for inclusion after full-text screening. Our review of the reference lists of relevant papers did not yield any additional records. The study selection process is illustrated in Figure 1.

Excluded studies with reasons for exclusion are listed in Multimedia Appendix 4.

Interrater reliability was assessed at both screening stages. The κ coefficients were 0.71 (moderate agreement) for the title and abstract screening and 0.65 (moderate agreement) for the full-text screening.

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



Overview of Included Studies

We included 20 studies evaluating 17 unique conversational agents in this review (Table 1) [12-14,30-46]. Out of the 20 studies, 10 (50%) were RCTs, 8 (40%) were quasi-experimental studies, and 2 (10%) were qualitative studies. PA was the sole target of intervention in half of the studies [12,13,32,35,37,38,41,43,45,46]. In the other half of the studies, PA was a primary target, but there were additional targets such as diet [33,34,36,39,44], fruit and vegetable consumption [30,31,40], medication adherence [14], mental well-being [33,36], stress management [33,34,36,44], and sun protection [42]. A total of 60% (12/20) studies used subjective measures to gauge intervention effectiveness and user satisfaction, and the other 40% (8/20) studies relied on objective data from pedometers or accelerometers.

The studies were conducted in 8 different countries. Studies were primarily conducted in nonclinical populations (eg, healthy adults and college students), with only 5 studies recruiting from clinical settings (eg, clinics and hospitals) [12,14,32,36,45]. The sample size ranged from 4-958 participants (median 55; mean 117, SD 206.3). Half of the studies were published in the last 3 years (2017-2020 [33,34,36-42,46]), and the other half were published between 2005 and 2014 [12-14,30-32,35,43-45].

Table 1. Study characteristics.

Characteristics and study	Targeted behaviors	Population	Location	Initial sample size[a], n	Final sample size[b], n (%)
RCT[c]					
Bickmore et al [12]	PA[d]	Geriatric ambulatory clinic patients	United States	21	16 (76.2)
Bickmore et al [13]	PA	Healthy adults	United States	101	91 (90.1)
Bickmore et al [31]	PA and fruit or vegetable consumption	Healthy adults	United States	122	113 (92.6)
Bickmore et al [32]	PA	Geriatric ambulatory clinic patients	United States	263	250 (95.1)
Friederichs et al [35]	PA	Healthy adults	Netherlands	958	500 (52.2)
Gardiner et al [36]	PA, diet, mental well-being, and stress	Primary care clinic patients	United States	61	57 (93.4)
Kramer et al [38]	PA	Insurees of an insurance company	Switzerland	274	274 (100)
Piao et al [41]	PA	Office employees	South Korea	121	106 (87.6)
Vainio et al [44]	PA, diet, and stress	Healthy adults	Finland	66	38 (57.6)
Watson et al [45]	PA	Hospital patients	United States	70	62 (88.6)
Quasi-experimental					
Bickmore et al [14]	PA and medication	Patients with schizophrenia	United States	20	16 (80)
Bickmore et al [30]	PA and fruit or vegetable consumption	Healthy adults	United States	8	8 (100)
Fadhil and AbuRa'ed [33]	PA, diet, mental well-being, and stress	Healthy adults	Iraq	43	43 (100)
Fadhil et al [34]	PA, diet, and stress	University students	Italy	22	19 (86.4)
Kocielnik et al [37]	PA	Healthy adults	United States	33	33 (100)
Maher et al [39]	PA and diet	Healthy adults	Australia	31	28 (90.3)
Olafsson et al [40]	PA and fruit or vegetable consumption	College students	United States	39	39 (100)
Zhou et al [46]	PA	Chinese adults living in the United States	United States	49	49 (100)
Qualitative					
Sillice et al [42]	PA and sun protection	Healthy adults	United States	34	34 (100)
Simila et al [43]	PA	Older adults in exercise groups or home care	Finland	4	4 (100)

[a]Number of participants who began the study.

[b]Number of participants who completed the intervention.

[c]RCT: randomized controlled trial.

[d]PA: physical activity.

Results of Appraisal

Of the 20 included studies, 10 (50%) were categorized as quantitative research (RCT or nonrandomized study), 8 (40%) as mixed methods studies, and 2 (10%) as qualitative research. Overall, the methodological quality of the 20 studies varied: 55% (11/20) of the studies met ≤60% of the criteria outlined by the MMAT (lower methodological quality), and 45% (9/20) of the studies met >60% of the criteria (higher methodological quality). Reviewers' ratings for each methodological quality criterion are presented in Multimedia Appendix 5 [12-14,30-46].

Overview of Conversational Agents

The 20 included studies evaluated 17 unique conversational agents (Table 2). A conversational agent, Laura, was used in 15% (3/20) of the studies [12-14], and another agent, Karen, was used in 10% (2/20) of the studies [30,31]. Conversational agents Steps to Health [32], Gabby [36], Emily [40], and Elsie/Meimei [46] were designed with similar architectural systems; however, they used distinct dialog flows tailored to different populations (eg, older adults, racially diverse city-dwelling women, and Chinese adults living in the United

States), so they were categorized as unique agents. For example, the conversational agent developed for racially diverse city-dwelling women delivered culturally aware patient strategies and health information and mentioned prayers and spiritual traditions [36]. Similarly, the conversational agent developed for Chinese adults emphasized values common to the Chinese culture, including collectivism [46].

Table 2. Conversational agent characteristics.

Conversational agent or program name	Delivery method (computer or phone)	Delivery platform	Conversational agent output (speech or text)	User input (constrained or unconstrained)
ECA[a]				
Laura or FitTrack [12-14]	Computer	Software	Speech	Constrained
Karen [30,31]	Computer	Software	Speech	Constrained
Steps to Health [32]	Computer	Software	Speech	Constrained
Gabby [36]	Computer	Web-based	Speech	Constrained
Emily [40]	Computer	Software	Speech	Constrained
Project RAISE [42]	Computer	Software	Speech	Not specified
Virtual Coach [45]	Computer	Software	Speech	Constrained
Elsie or Meimei [46]	Computer	Software	Speech	Constrained
Chatbot				
Ollobot [33]	Both	Messaging app	Text	Unconstrained
CoachAI [34]	Both	Messaging app	Text	Unconstrained
Reflection Companion [37]	Phone	SMS or MMS[b]	Text	Unconstrained
Ally [38]	Phone	Mobile app	Text	Constrained
Paola or MedLiPal [39]	Both	Messaging app	Text	Unconstrained
Healthy Lifestyle Coaching Chatbot [41]	Both	Messaging app	Text	Unconstrained
Both				
I Move [35]	Computer	Web-based	Text	Both
AmIE Project [43]	Computer	Software	Both	Constrained
Mindless Change [44]	Phone	Mobile app	Both	Constrained

[a]ECA: embodied conversational agent.

[b]MMS: multimedia messaging service.

Of the 17 conversational agents, 10 (59%) were computer-based [12-14,30-32,35,36,40,42,43,45,46], 4 (24%) could be used on computers or phones [33,34,39,41], and 3 (18%) were designed for mobile devices only [37,38,44]. Conversational agents were implemented using standalone computer software [12-14,30-32,40,42,43,45,46], messaging apps [33,34,39,41], web-based platforms [35,36], mobile apps [38,44], and SMS text messaging or MMS [37].

In total, of the 17 agents, 8 (47%) were embodied conversational agents (ECAs) with synthesized speech [12-14,30-32,36,40,42,45,46], 6 (35%) were text-only chatbots [33,34,37-39,41], and 3 (18%) had both an ECA and chatbot option [35,43,44]. With all 17 conversational agents, participants gave input by typing on a keyboard or selecting answer options with a mouse, touchpad, or touchscreen; 59% (10/17) of the conversational agents limited users to constrained input, whereby users selected answers from a multiple-choice list of options, and conversational agents responded according to predefined templates [12-14,30-32,36,38,40,43-46]. Only 29% (5/17) of the conversational agents accepted free-text responses and used machine learning and natural language processing to understand users' input and generate replies [33,34,37,39,41], and 6% (1/17) of the conversational agents accepted free-text responses and multiple-choice answers [35]. The remaining 6% (1/17) conversational agents did not specify what user inputs were accepted [42].

Intervention Effectiveness and Impact

RCTs

Of the 10 RCTs, 6 (60%) found that participants in the conversational agent group outperformed participants in the control group on various PA measures. Intervention groups increased daily walking more quickly [31], achieved >30 minutes of exercise or 10,000 steps per day more times per week [13], significantly increased step count during the study period [12,32], significantly increased self-reported PA at 1 month [35], and maintained step counts throughout time [45]. Only

10% (1/10) RCTs did not find significant differences in activity levels between the intervention and control groups [36].

The remaining 30% (3/10) of RCTs used conversational agents in both experimental and control groups but varied the conversational agent conditions (eg, cash incentives vs charity incentives vs no incentives [38], rewards vs no rewards [41], and ECA vs text-only chatbot [44]). In 67% (2/3) of these studies, interacting with a conversational agent significantly increased step counts and self-reported activity across all conditions; however, including financial incentives and rewards further boosted activity levels [38,41]. The last RCT determined that conversational agents were useful but limited by low adherence [44].

Quasi-Experimental Studies

Of the 8 quasi-experimental studies, 6 (75%) used within-subjects pre-post designs [14,30,33,34,37,39] and 2 (25%) included comparator groups [40,46] (Multimedia Appendix 6 [12-14,30-46]). Of the 8 quasi-experimental studies, 3 (38%) measured changes in activity level as a result of interacting with a conversational agent [14,34,39]; 2 found positive impacts in the form of increased enjoyment during walking [14], higher frequency of step-goal achievement [14], and increased weekly exercise time [39], and 1 did not find any differences in activity levels [34].

An additional 38% (3/8) of the quasi-experimental studies measured participants' attitudes toward exercise before and after the intervention [37,40,46]. Conversational agents successfully triggered reflection on new exercise routines [37], increased participants' self-efficacy and motivation to exercise for at least 30 minutes every day [40] and persuaded participants to start regular exercise [46].

The remaining 25% (2/8) of the quasi-experimental studies discussed users' preliminary experiences with conversational agents [30,33]. Overall, these conversational agents had moderately high usability and feasibility. Participants perceived them to be satisfactory [30,33], trustworthy [30], empathetic [30], useful [33], and easy to use [33].

Qualitative Studies

Of the 20 included studies, only 2 (10%) were qualitative studies [42,43]. In one study, most participants had positive, satisfying interactions with the relational agent and found the agent humanlike, caring, and supportive [42]. About half of the participants viewed the relational agent as informative and felt motivated to maintain regular exercise. Another qualitative study compared two different PA conversational agents: a text-based chatbot and an ECA [43]. Participants had positive experiences with both systems and felt that conversational agents could provide motivation and serve as information channels.

ECAs Versus Chatbots

ECAs and text-only chatbots performed similarly, with 88% (7/8) of the ECAs and 83% (5/6) of the chatbots positively affecting participants' PA levels, motivation to exercise, or perceptions of conversational agents. Of all 20 studies, 3 (15%) directly compared ECAs with chatbots; one study found that both were equally effective at building social relationships and increasing PA [35], one study suggested that ECAs could provide a slightly more engaging user experience than chatbots [42], and the remaining study described the benefits and drawbacks of each conversational agent [43].

Intervention Characteristics

Theory

Of the 20 studies, 11 (55%) cited a theory that guided their intervention development (Table 3). Of these 11 studies, 6 (55%) designed the intervention and selected program elements according to the referenced theories [37,38,40,41,44,46], and 5 (45%) mentioned a theory as their overarching framework but did not explicitly link intervention components with corresponding theoretical constructs [13,30,31,34,42].

The used theories could be broadly categorized into learning theories, which describe how people receive and process knowledge, and behavior change theories, which explain how behaviors develop and shift throughout time. Four interventions were based on a combination of theories [30,31,40,44], and 1 intervention used the Hofstede cultural dimensions theory to develop culturally appropriate dialog for an American and a Chinese conversational agent [46].

Table 3. Distribution of theories.

Theoretical model or framework	Study
Learning theories	
Learning theory (broad) [47]	Kocielnick et al [37]
Social learning theory [48]	Bickmore et al [13]
Social cognitive theory [49]	Bickmore et al [30,31]
Constructivist learning theory [47]	Vainio et al [44]
Cognitive dissonance theory [50]	Olafsson et al [40]
Behavior change theories	
Behavior change theory (broad) [51]	Bickmore et al [31], Kramer et al [38]
Habit formation model [52]	Piao et al [41], Vainio et al [44]
Health action process approach [53]	Fadhil et al [34]
Transtheoretical model [54]	Bickmore et al [30,31], Olafsson et al [40], Sillice et al [42]
Other	
Hofstede's cultural dimensions theory [55]	Zhou et al [46]

Dialog Flow Development

Of the 20 studies, 9 (45%) discussed the use of one or more treatment approaches to guide the development of dialog flows for conversational agents. The most commonly used approach was motivational interviewing [30,31,35-37,40], followed by cognitive behavioral therapy [13,33,34,45] and behavioral therapy [13,45].

Of the 9 studies, 4 (44%) described how dialog flows were adapted from face-to-face counseling and prepared for virtual delivery. Techniques included using transcripts from videotaped counseling sessions as a basis for the conversational structure [30,40], using a dialog interpreter to convert statements from counseling sessions into interactive virtual conversations [31], and developing scripts through literature reviews and consultations with physicians, computer scientists, and exercise trainers [45]. The remaining 56% (5/9) studies did not explain how dialog flows for conversational agents were written.

Intervention Components

The most common program components were health education, motivational messages, problem-solving barriers to exercise, goal setting, self-monitoring, and exercise tips (Table 4). Additional components included reminders, homework, workout planning, incentives, and reflection.

Participants found health education helpful [36,42], as it allowed them to learn new ways of increasing PA [40]. They also enjoyed receiving tips for new exercise routines [40] and periodic exercise reminders [31,42]. Positive feedback motivated participants [37], built rapport [42], and increased agent likeability [31]. Participants appreciated progress tracking features [34] and visual step charts [31,32]. Conversational agents helped participants formulate concrete goals, action plans, and overcome obstacles [37]. However, participants mentioned that they would have liked to talk more about how their health problems affected their ability to exercise [12]. Change talk and reflection helped participants increase their commitment to positive health behaviors [37,40]. Finally, rewards were implemented with moderate success, with one study finding that daily cash incentives increased step-goal achievement by 8.1% [38] and another study finding that intrinsic rewards improved habit formation and enhanced intervention sustainability [41].

Table 4. Distribution of intervention components.

Study	Goal setting	Positive reinforcement	Self-monitoring	Problem-solving barriers	Education	Tips	Reminders	Homework	Workout planning	Rewards	Change talk or reflection (motivational interviewing)
Bickmore et al [12]	✓[a]	✓	✓	✓	✓	✓					
Bickmore et al [13]	✓	✓	✓	✓	✓						
Bickmore et al [14]	✓	✓	✓	✓	✓	✓					
Bickmore et al [30]	✓	✓		✓	✓	✓		✓			✓
Bickmore et al [31]	✓			✓	✓			✓			
Bickmore et al [32]	✓	✓	✓	✓		✓					
Fadhil and Abu-Ra'ed [33]		✓	✓		✓	✓			✓		✓
Fadhil et al [34]		✓	✓		✓	✓	✓				
Friederichs et al [35]		✓		✓	✓				✓		✓
Gardiner et al [36]	✓		✓	✓	✓	✓		✓			
Kocielnick et al [37]	✓		✓	✓							✓
Kramer et al [38]	✓		✓	✓		✓			✓	✓	
Maher et al [39]	✓		✓	✓	✓						
Olafsson et al [40]		✓			✓						✓
Piao et al [41]	✓	✓					✓			✓	
Sillice et al [42]		✓	✓	✓			✓	✓			
Simila et al [43]		✓	✓		✓	✓	✓				
Vainio et al [44]	✓	✓	✓		✓	✓			✓		✓
Watson et al [45]	✓	✓	✓	✓	✓	✓					
Zhou et al [46]				✓	✓				✓		

[a]Intervention component present.

Challenges and Areas for Improvement

Conversational Agent Constraints

The most common challenges were related to the capabilities of conversational agents. In 59% (10/17) of the conversational agents, users were required to respond via multiple-choice answers. This format limited user freedom [12,13] and lacked the personalization necessary to address more complex issues [14]. Although researchers acknowledged the need for more sophisticated dialog systems, they were concerned about the difficulty of implementing machine learning and the increased chance of misunderstanding users' intents [13].

Another area for improvement was communication modality. None of the conversational agents were built to accept spoken input. Participants were required to type out their answers or select answers using a mouse, touchpad, or touchscreen. In one study, participants universally stated that they would have preferred speaking to the conversational agent [12].

Studies have presented mixed findings on the value of ECAs with synthesized speech. According to qualitative data, talking ECAs seemed more versatile than text-only chatbots [43] and provided a closer approximation of face-to-face conversations with health care providers [32]. However, 45% (5/11) ECAs were criticized by participants for their robotic voices, slow pace, unnatural movements, and limited relational skills [35,36,40,42,46].

Program Delivery

Participants encountered more issues with computer-based than with phone-based conversational agents. Some participants had limited access to computers, limited time to sit in front of computers [36], or difficulties installing software and entering information [12]. Internet access was also an issue, with network breaks preventing participants from starting apps, synchronizing devices and databases, and connecting fitness trackers [43]. Many participants across studies felt that having the conversational agent on their phone would be more convenient and accessible, allowing them to complete the program "on the go" [32,36,42].

Mobile interventions were well-liked, particularly those that used familiar messaging apps, as they did not require participants to download and learn to use additional applications [41]. However, some participants had minimal smartphone skills

and did not know how to send text messages, thus limiting their engagement with the intervention [39]. In addition, one mobile app suffered from poor usability because of slow performance on older smartphones [44].

Program Content

Of the 20 studies, 7 (35%) studies mentioned the repetitiveness of program content as a key area for improvement [12,13,30,31,37,42,43]. This included dialog flows that were often repeated, leading to lower satisfaction [30] and increased boredom [37,43]. Participants desired more personalized responses and suggestions based on their health information, preferences, and PA history [37,40]. Owing to repetitiveness, participants felt that continued use would not lead to any additional impact [42].

User engagement waned throughout time [45], and high attrition rates limited the efficacy of the interventions. In one study, participants responded to 50% of the self-monitoring prompts and completed only a few exercise and coping plans, explaining that weekly planning was too difficult and time-consuming [38]. In another study, participants found the conversational agent engaging, but without external support, almost half of them discontinued the use of the service [44]. Participants who lapsed for a short period were more likely to quit the program [41].

Ethical Issues

Many relational agents relied on social dialog, humor, empathic statements, and personal stories to build rapport with users [14,31,32,42,46]. The use of these techniques may have increased the potential for misperceptions and false illusions, as virtual agents do not have emotions or personal histories. Humans tend to anthropomorphize advanced technology [13], and conversational agents may have deceived some users into thinking they were interacting with a human. One study pointed out that patients with schizophrenia who are experiencing a psychotic episode could be more likely to confuse relational agents with real people, develop parasocial relationships with relational agents, or become paranoid that relational agents or their programmers are monitoring their behavior [14]. Researchers attempted to address this matter by having the relational agent periodically remind users that it was “just a computer character with limited capabilities” [14].

Standards of Care

Of the 20 studies, only 1 (5%) compared the quality of care between a human and a conversational agent. This study found that a human agent was often more motivating, engaging, and supportive than a virtual agent [34].

Most studies did not address privacy features or data storage and access procedures despite participants expressing concerns that conversational agents could collect and share their personal information [12,14]. One study discussed security measures, such as requiring usernames and passwords and automatically logging users out after a period of inactivity [36]. Another study described weekly backup procedures to mitigate the possibility of data loss due to system crashes or computer theft [12].

Finally, 10% (2/20) of studies discussed user safety issues. One conversational agent provided videos demonstrating exercises that a participant with arthritis could not safely perform without the help of an elastic band [43]. Another study discussed the necessity of improving automated dialog flows because of conversational agents’ inadequate responses to safety concerns mentioned in users’ free-text answers [40].

Discussion

Principal Findings

This literature review charted data from 20 studies that evaluated 17 PA conversational agents. Overall, conversational agent interventions were feasible and promising for increasing PA. Of the 10 RCTs, 6 (60%) found that participants assigned to the conversational agent group outperformed participants in the control group on PA measures, such as step counts and exercise frequency and duration. Conversational agents had moderate usability and acceptability, as measured by subjective data in the form of questionnaires, interviews, activity logs, and diaries. The interventions were generally found to be useful, easy to use, and satisfactory to participants; however, they faced some implementation challenges, including high attrition, technical issues, limited options for user input, and privacy and security risks. Methodological quality varied across studies, and few studies adequately addressed issues of user engagement, safety, and ethics.

Comparison With Prior Work

To the best of our knowledge, this is the first systematic review to evaluate PA conversational agents. Previous reviews have reported on the effectiveness of digital interventions for increasing PA [15-18]. Our results are consistent with their findings that digital interventions have a modest effect on activity levels, particularly in the short term; however, user engagement tends to decline over time [16-18]. Our findings are also in line with other reviews’ evaluations of health care conversational agents, which show that natural language processing and machine learning are underused, high-quality evidence and attention to patient safety are lacking, and study methods and evaluation measures are often inconsistently reported [8,20,21].

Recommendations

On the basis of the findings of this review, we propose several recommendations for the future design and implementation of PA conversational agents.

Program Content

Participant feedback indicated that many intervention programs lost their novelty over time, resulting in decreased user engagement. More diverse program content is required to maintain long-term user satisfaction. A way to reduce repetitiveness is through just-in-time adaptive interventions (JITAIs), which provide dynamically tailored support when users need it while minimizing user burden [10]. JITAIs can inform participants when they have been sedentary for long periods or when they are behind on their step goals. In addition, JITAIs can offer exercise suggestions based on weather conditions, time of day, and users’ physical surroundings. JITAIs

for conversational agents are currently being explored and developed through microrandomized trials [10].

Another way to improve the sustainability of interventions is to base their programming on relevant behavior change theories and evidence-based treatment approaches. Behavior change theories may help identify intervention techniques that tap into users' motivations and result in increased engagement. Similarly, dialog flows based on treatment approaches, such as motivational interviewing and cognitive behavioral therapy, can help users explore and resolve barriers to PA. Owing to the heterogeneity of the studies, we were unable to determine if the inclusion of a theoretical framework or treatment approach increased intervention effectiveness in this review. Future work should assess this as the number of studies increases.

Programming conversational agents to send periodic tips and exercise reminders may help decrease the high attrition rates reported in a few studies [38,41,44,45]. In addition, as many PA interventions are self-guided, encouraging users to share goals and progress with their social circles may increase accountability.

Conversational Agent Delivery

Computer-based ECAs were the most common agents used; however, qualitative interviews revealed that participants desired mobile delivery platforms. Phone ownership rates are higher than computer ownership rates [7]; thus, conversational agents operating via SMS or MMS text messaging may increase scalability. They are also appropriate for those with low digital literacy. For computer-based agents, web-based platforms and familiar messaging apps that do not need to be installed or regularly updated may be more accessible than standalone software.

ECAs have the potential to improve human-computer interactions; however, they are commonly criticized as robotic and unnatural. ECAs can be improved by replacing synthesized speech with human voice, giving users control over pacing of messages, and designing higher-quality animation. Automatic speech recognition is highly desirable, particularly among populations with low vision or difficulty typing. In addition, although artificially intelligent conversational agents may take more time to develop, they afford users more freedom and personalized content to sustain engagement and maximize treatment efficacy.

Safety and Ethics

Most conversational agent programs were designed for healthy and able-bodied adults; however, programs should also be equipped with education and exercise tips for users of different age groups and users with physical limitations. Conversational agents should offer suggestions for exercise-related injuries or pain, such as performing pre- and postworkout stretches, modifying activities, and consulting with health care providers. Users may mention mental health conditions such as depression or anxiety that prevent them from exercising. Thus, researchers should consider incorporating dialog flows that refer users to mental health resources and crisis hotlines. If interventions are designed specifically for clinical populations, additional safety features may be necessary, such as periodic check-ins with a human advisor. Furthermore, for individuals with severe mental illnesses, such as psychosis, additional consideration may be warranted, including ensuring agents are not too anthropomorphic.

Users often share sensitive health information with conversational agents. However, only a few studies have discussed privacy and security issues. User privacy should be protected through measures such as requiring logins and passwords for apps and software, deidentifying user data, and archiving past conversations.

Finally, efforts must be made to uphold the quality of digital interventions. There are currently no regulations regarding the standards of care for conversational agents. Similar to health interventions provided by human coaches, conversational agent programs should be based on a relevant theory and treatment approach to ensure that they are grounded in evidence-based practice.

Limitations

The findings of this review must be considered in the context of a few limitations. First, we may have missed relevant studies in additional databases despite our search strategy being fairly broad. In particular, we lacked quantitative descriptive studies and qualitative studies without comparison conditions, which could suggest that our PICOS criteria were better suited for effectiveness studies that included comparison conditions. Although we aimed to include usability studies without comparison conditions, we had to exclude many such studies because of insufficient data on study participants' experiences, the intervention's impact on activity levels, or the intervention's theoretical mechanisms of change.

Second, because of the heterogeneity of study designs and outcome data, we could not conduct a meta-analysis or directly compare different interventions. We synthesized the main findings from the existing literature; however, without effect sizes, it was difficult to draw definitive conclusions about intervention effectiveness. This field of research would benefit from more longitudinal RCTs that evaluate the long-term sustainability of conversational agents.

Third, we appraised the methodological quality of the included studies following the appropriate method-related standards using the MMAT, one of the few tools designed specifically for mixed methods reviews. However, the MMAT is not designed to grade the level of evidence or the risk of bias in effectiveness studies. We chose not to apply an ad hoc tool to appraise the risk of bias of effectiveness studies because only half of the included studies were RCTs that reported on treatment effectiveness. To date, there is no single, unified approach for assessing confidence in findings generated from combined quantitative and qualitative evidence [56]. More research is needed on best practices for critically appraising included studies in mixed methods reviews.

Fourth, we refrained from analyzing the more technical aspects of conversational agents (eg, programming and interfaces), choosing instead to focus on intervention components and guiding frameworks. Additional questions regarding technical design should be studied in systematic reviews to maximize the user-friendliness of conversational agents.

Fifth, intervention techniques were difficult to identify, as some studies embedded them within figures rather than discussing them descriptively, and there was no uniform language across studies regarding techniques.

Finally, more than half of the included studies focused exclusively on healthy adults, thus limiting the generalizability of their results. As conversational agents are often designed for a broad audience, future studies should also consider sampling from youth and clinical populations (eg, individuals with mental illness or pre-existing health conditions).

Conclusions

On the basis of current evidence, conversational agents appear to be a feasible and effective modality for delivering PA interventions. However, more research comparing conversational agents with other forms of interventions, including human-delivered interventions, is required. Most conversational agents reviewed were computer-based and constrained users to written, predefined inputs. Future conversational agents should consider accessibility and inclusive design and consider supporting automatic speech recognition, natural language processing, and mobile phone platforms. In addition, program content should be further personalized and diversified by using relevant evidence-based frameworks and their accompanying behavior change methods. Researchers should provide a clear overview of how they select intervention components and how these components affect health behavior. This can lead to a deeper understanding of the mechanisms of change in interventions, and consequently, increase the effectiveness of these interventions. Personalization of program content may also lead to higher user satisfaction and engagement while supporting user choice and agency. Finally, in addition to user experiences, safety, privacy, and ethical concerns should be prioritized in the design of PA conversational agents.

Acknowledgments

The authors would like to thank Margaret Phillips from the University of California, Berkeley, for her help in building search strings for the review. The authors received no specific funding for this study.

Authors' Contributions

This study was designed by TCL and CAF, with input from AA and CRL. Screening and appraisal of articles were completed by TCL and CAF. Data extraction and analysis were performed by TCL. The first draft of the manuscript was written by TCL. Revisions and subsequent drafts were completed by TCL, AA, CRL, and CAF.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 104 KB - jmir_v23i9e25486_app1.pdf]

Multimedia Appendix 2
Search strategy.
[PDF File (Adobe PDF File), 58 KB - jmir_v23i9e25486_app2.pdf]

Multimedia Appendix 3
Data extraction form with descriptions of data items.
[PDF File (Adobe PDF File), 86 KB - jmir_v23i9e25486_app3.pdf]

Multimedia Appendix 4
Excluded studies with reasons for exclusion.
[PDF File (Adobe PDF File), 95 KB - jmir_v23i9e25486_app4.pdf]

Multimedia Appendix 5
Mixed Methods Appraisal Tool (MMAT) quality appraisal profile.
[PDF File (Adobe PDF File), 87 KB - jmir_v23i9e25486_app5.pdf]

Multimedia Appendix 6
Characteristics of comparators in each included study.
[PDF File (Adobe PDF File), 79 KB - jmir_v23i9e25486_app6.pdf]

References

1. World Health Organization. Global Action Plan on Physical Activity 2018–2030: More Active People for a Healthier World. Geneva Switzerland: World Health Organization; 2018.
2. Global Health Observatory data repository: Prevalence of insufficient physical activity among adults. World Health Organization. 2018 Nov 05. URL: https://apps.who.int/gho/data/view.main.2463 [accessed 2020-09-05]
3. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
4. Salwen-Deremer JK, Khan AS, Martin SS, Holloway BM, Coughlin JW. Incorporating health behavior theory into mHealth: an examination of weight loss, dietary, and physical activity interventions. J Technol Behav Sci 2019 Nov 04;5:51-60. [doi: 10.1007/s41347-019-00118-6]
5. Taylor K, Silver L. Smartphone ownership is growing rapidly around the world, but not always equally. Pew Research Center. 2019 Feb 05. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2020-09-04]
6. Adult physical inactivity prevalence maps by race/ethnicity. Centers for Disease Control and Prevention. 2020 Jan. URL: https://www.cdc.gov/physicalactivity/data/inactivity-prevalence-maps/index.html [accessed 2020-06-30]
7. Ryan C. Computer and internet use in the United States: 2016. American Community Survey Reports - U.S. Census Bureau. 2018 Aug. URL: https://www.census.gov/content/dam/Census/library/publications/2018/acs/ACS-39.pdf [accessed 2020-11-02]
8. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The Personalization of conversational agents in health care: systematic review. J Med Internet Res 2019 Nov 07;21(11):e15360 [FREE Full text] [doi: 10.2196/15360] [Medline: 31697237]
9. O'Shea J, Bandar Z, Crockett K. Systems engineering and conversational agents. In: Tolk A, Jain LC, editors. Intelligence-Based Systems Engineering. Berlin, Germany: Springer; 2011:201-232.
10. Kramer J, Künzler F, Mishra V, Presset B, Kotz D, Smith S, et al. Investigating intervention components and exploring states of receptivity for a smartphone app to promote physical activity: protocol of a microrandomized trial. JMIR Res Protoc 2019 Jan 31;8(1):e11540 [FREE Full text] [doi: 10.2196/11540] [Medline: 30702430]
11. Wiratunga N, Cooper K, Wijekoon A. FitChat: conversational artificial intelligence interventions for encouraging physical activity in older adults. arXivcs. 2020 Apr 29. URL: http://arxiv.org/abs/2004.14067 [accessed 2020-09-03]
12. Bickmore TW, Caruso L, Clough-Gorr K, Heeren T. 'It's just like you talk to a friend' relational agents for older adults. Interact Comput 2005 Dec;17(6):711-735. [doi: 10.1016/j.intcom.2005.09.002]
13. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005 Oct;59(1):21-30. [doi: 10.1016/j.pec.2004.09.008] [Medline: 16198215]
14. Bickmore TW, Puskar K, Schlenk EA, Pfeifer LM, Sereika SM. Maintaining reality: Relational agents for antipsychotic medication adherence. Interact Comput 2010 Jul;22(4):276-288. [doi: 10.1016/j.intcom.2010.02.001]
15. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
16. Muellmann S, Forberger S, Möllers T, Bröring E, Zeeb H, Pischke CR. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: a systematic review. Prev Med 2018 Mar;108:93-110. [doi: 10.1016/j.ypmed.2017.12.026] [Medline: 29289643]
17. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
18. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, et al. Can smartphone apps increase physical activity? Systematic review and meta-analysis. J Med Internet Res 2019 Mar 19;21(3):e12053 [FREE Full text] [doi: 10.2196/12053] [Medline: 30888321]
19. Kramer LL, Stal ST, Mulder BC, de Vet E, van Velsen L. Developing embodied conversational agents for coaching people in a healthy lifestyle: scoping review. J Med Internet Res 2020 Feb 06;22(2):e14058 [FREE Full text] [doi: 10.2196/14058] [Medline: 32022693]
20. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
21. Car LT, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020 Aug 07;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]
22. Bendig E, Erb B, Schulze-Thuesing L, Baumeister H. The Next Generation: Chatbots in clinical psychology and psychotherapy to foster mental health – a scoping review. Verhaltenstherapie 2019 Aug 20:1-13 [FREE Full text] [doi: 10.1159/000501812]

23. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
24. Luo TC, Figueroa C, Aguilera A, Lyles C. Promoting physical activity through conversational agents: a mixed-method systematic review. OSF Home. 2020. URL: https://osf.io/p4v6y/ [accessed 2021-08-24]
25. Tacconelli E. Systematic reviews: CRD's guidance for undertaking reviews in health care. Lancet Infect Dis 2010 Apr;10(4):226. [doi: 10.1016/s1473-3099(10)70065-7]
26. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Educ Inf 2018 Dec 18;34(4):285-291. [doi: 10.3233/EFI-180221]
27. Pace R, Pluye P, Bartlett G, Macaulay AC, Salsberg J, Jagosh J, et al. Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. Int J Nurs Stud 2012 Jan;49(1):47-53. [doi: 10.1016/j.ijnurstu.2011.07.002] [Medline: 21835406]
28. Hong Q, Pluye P, Fabregues S. Mixed Methods Appraisal Tool (MMAT), version 2018: user guide. National Collaborating Centre for Methods and Tools. 2018 Aug 01. URL: http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/fetch/127916259/MMAT_2018_criteria-manual_2018-08-01_ENG.pdf [accessed 2020-11-06]
29. Hong QN, Pluye P, Bujold M, Wassef M. Convergent and sequential synthesis designs: implications for conducting and reporting systematic reviews of qualitative and quantitative evidence. Syst Rev 2017 Mar 23;6(1):61 [FREE Full text] [doi: 10.1186/s13643-017-0454-2] [Medline: 28335799]
30. Bickmore TW, Schulman D, Sidner CL. A reusable framework for health counseling dialogue systems based on a behavioral medicine ontology. J Biomed Inform 2011 Apr;44(2):183-197 [FREE Full text] [doi: 10.1016/j.jbi.2010.12.006] [Medline: 21220044]
31. Bickmore TW, Schulman D, Sidner C. Automated interventions for multiple health behaviors using conversational agents. Patient Educ Couns 2013 Aug;92(2):142-148 [FREE Full text] [doi: 10.1016/j.pec.2013.05.011] [Medline: 23763983]
32. Bickmore TW, Silliman RA, Nelson K, Cheng DM, Winter M, Henault L, et al. A randomized controlled trial of an automated exercise coach for older adults. J Am Geriatr Soc 2013 Oct;61(10):1676-1683. [doi: 10.1111/jgs.12449] [Medline: 24001030]
33. Fadhil A, AbuRa'ed A. OlloBot - towards a text-based Arabic health conversational agent valuation and results. In: Proceedings of Recent Advances in Natural Language Processing. 2019 Presented at: Recent Advances in Natural Language Processing; Sep 2–4, 2019; Varna, Bulgaria p. 295-303. [doi: 10.26615/978-954-452-056-4_034]
34. Fadhil A, Wang Y, Reiterer H. Assistive conversational agent for health coaching: a validation study. Methods Inf Med 2019 Jun;58(1):9-23. [doi: 10.1055/s-0039-1688757] [Medline: 31117129]
35. Friederichs S, Bolman C, Oenema A, Guyaux J, Lechner L. Motivational interviewing in a web-based physical activity intervention with an avatar: randomized controlled trial. J Med Internet Res 2014 Feb 13;16(2):e48 [FREE Full text] [doi: 10.2196/jmir.2974] [Medline: 24550153]
36. Gardiner PM, McCue KD, Negash LM, Cheng T, White LF, Yinusa-Nyahkoon L, et al. Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: a feasibility randomized control trial. Patient Educ Couns 2017 Sep;100(9):1720-1729 [FREE Full text] [doi: 10.1016/j.pec.2017.04.015] [Medline: 28495391]
37. Kocielnik R, Xiao L, Avrahami D, Hsieh G. Reflection companion: a conversational system for engaging users in reflection on physical activity. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Jun;2(2):1-26. [doi: 10.1145/3214273]
38. Kramer J, Künzler F, Mishra V, Smith SN, Kotz D, Scholz U, et al. Which components of a smartphone walking app help users to reach personalized step goals? Results from an optimization trial. Ann Behav Med 2020 Jun 12;54(7):518-528 [FREE Full text] [doi: 10.1093/abm/kaaa002] [Medline: 32182353]
39. Maher CA, Davis CR, Curtis RG, Short CE, Murphy KJ. A physical activity and diet program delivered by artificially intelligent virtual health coach: proof-of-concept study. JMIR Mhealth Uhealth 2020 Jul 10;8(7):e17558 [FREE Full text] [doi: 10.2196/17558] [Medline: 32673246]
40. Olafsson S, O'Leary T, Bickmore T. Coerced change-talk with conversational agents promotes confidence in behavior change. In: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2019 May Presented at: PervasiveHealth'19: The 13th International Conference on Pervasive Computing Technologies for Healthcare; May 20-23, 2019; Trento, Italy p. 31-40. [doi: 10.1145/3329189.3329202]
41. Piao M, Ryu H, Lee H, Kim J. Use of the healthy lifestyle coaching chatbot app to promote stair-climbing habits among office workers: exploratory randomized controlled trial. JMIR Mhealth Uhealth 2020 May 19;8(5):e15085 [FREE Full text] [doi: 10.2196/15085] [Medline: 32427114]
42. Sillice MA, Morokoff PJ, Ferszt G, Bickmore T, Bock BC, Lantini R, et al. Using relational agents to promote exercise and sun protection: assessment of participants' experiences with two interventions. J Med Internet Res 2018 Feb 07;20(2):e48 [FREE Full text] [doi: 10.2196/jmir.7640] [Medline: 29415873]
43. Similä H, Merilahti J, Ylikauppila M, Muuraiskangas S, Perälä J, Kivikunnas S. Comparing two coaching systems for improving physical activity of older adults. In: Romero RL, editor. XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013. Cham: Springer; 2014:1197-1200.

44. Vainio J, Korhonen I, Kaipainen K, Kenttä O, Järvinen J. Learning healthy habits with a mobile self-intervention. In: Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare. 2014 Jul 23 Presented at: 8th International Conference on Pervasive Computing Technologies for Healthcare; May 20-23, 2014; Oldenburg, Germany. [doi: 10.4108/icst.pervasivehealth.2014.254951]
45. Watson A, Bickmore T, Cange A, Kulshreshtha A, Kvedar J. An internet-based virtual coach to promote physical activity adherence in overweight adults: randomized controlled trial. J Med Internet Res 2012 Jan 26;14(1):e1 [FREE Full text] [doi: 10.2196/jmir.1629] [Medline: 22281837]
46. Zhou S, Zhang Z, Bickmore T. Adapting a persuasive conversational agent for the Chinese culture. In: Proceedings of the International Conference on Culture and Computing (Culture and Computing). 2017 Presented at: International Conference on Culture and Computing (Culture and Computing); Sep 10-12, 2017; Kyoto, Japan. [doi: 10.1109/culture.and.computing.2017.42]
47. Harasim L. Learning Theory and Online Technology. Oxfordshire, England: Routledge; May 16, 2017.
48. Bandura A. Social Learning Theory. Englewood Cliffs, New Jersey: Prentice Hall; 1977.
49. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
50. Festinger L. A Theory of Cognitive Dissonance. Palo Alto, California: Stanford University Press; 1957.
51. Glanz K, Rimer B, Viswanath K. Health Behavior and Health Education: Theory, Research, and Practice. 4th Edition. San Francisco, California: Jossey-Bass; 2008:1-592.
52. Gardner B, Lally P, Wardle J. Making health habitual: the psychology of 'habit-formation' and general practice. Br J Gen Pract 2012 Dec;62(605):664-666 [FREE Full text] [doi: 10.3399/bjgp12X659466] [Medline: 23211256]
53. Schwarzer R, Luszczynska A. How to overcome health-compromising behaviors. Eur Psychol 2008 Jan;13(2):141-151. [doi: 10.1027/1016-9040.13.2.141]
54. Marcus BH, Simkin LR. The transtheoretical model: applications to exercise behavior. Med Sci Sports Exerc 1994 Nov;26(11):1400-1404. [Medline: 7837962]
55. Hofstede G. Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations. Thousand Oaks, CA: SAGE Publications; Apr 20, 2001.
56. Noyes J, Booth A, Moore G, Flemming K, Tunçalp Ö, Shakibazadeh E. Synthesising quantitative and qualitative evidence to inform guidelines on complex interventions: clarifying the purposes, designs and outlining some methods. BMJ Glob Health 2019 Jan 25;4(Suppl 1):e000893-e00089- [FREE Full text] [doi: 10.1136/bmjgh-2018-000893] [Medline: 30775016]

Abbreviations

ECA: embodied conversational agent
JITAI: just-in-time adaptive intervention
MMAT: Mixed Methods Appraisal Tool
MMS: multimedia messaging service
PA: physical activity
PICOS: patient problem, intervention, comparison, outcomes, and studies
RCT: randomized controlled trial

Edited by R Kukafka; submitted 04.11.20; peer-reviewed by A Tapuria, M Sillice, P Pluye; comments to author 06.02.21; revised version received 01.03.21; accepted 19.07.21; published 14.09.21.

Please cite as:
Luo TC, Aguilera A, Lyles CR, Figueroa CA
Promoting Physical Activity Through Conversational Agents: Mixed Methods Systematic Review
J Med Internet Res 2021;23(9):e25486
URL: https://www.jmir.org/2021/9/e25486
doi:10.2196/25486
PMID:34519653

Review

Recent Academic Research on Clinically Relevant Digital Measures: Systematic Review

Md Mobashir Hasan Shandhi[1], PhD; Jennifer C Goldsack[2], MChem, MA, MBA; Kyle Ryan[3]; Alexandra Bennion[3]; Aditya V Kotla[3]; Alina Feng[3]; Yihang Jiang[1], BSc; Will Ke Wang[1], BA; Tina Hurst[4], PhD, RNutr; John Patena[5], MPH, MA; Simona Carini[6], MA; Jeanne Chung[2], MBA; Jessilyn Dunn[1,7], PhD

[1]Department of Biomedical Engineering, Duke University, Durham, NC, United States

[2]Digital Medicine Society, Boston, MA, United States

[3]Big Ideas Lab, Department of Biomedical Engineering, Duke University, Durham, NC, United States

[4]Activinsights Ltd, Cambridgeshire, United Kingdom

[5]Brown-Lifespan Center for Digital Health, Brown University, Providence, RI, United States

[6]Division of General Internal Medicine, University of California, San Francisco, CA, United States

[7]Department of Biostatistics & Bioinformatics, Duke University, Durham, NC, United States

Corresponding Author:
Jessilyn Dunn, PhD
Department of Biomedical Engineering
Duke University
1427 FCIEMAS, Box 90281
Durham, NC, 27708
United States
Phone: 1 919 660 5131
Fax: 1 919 684 4488
Email: jessilyn.dunn@duke.edu

Abstract

Background: Digital clinical measures collected via various digital sensing technologies such as smartphones, smartwatches, wearables, ingestibles, and implantables are increasingly used by individuals and clinicians to capture health outcomes or behavioral and physiological characteristics of individuals. Although academia is taking an active role in evaluating digital sensing products, academic contributions to advancing the safe, effective, ethical, and equitable use of digital clinical measures are poorly characterized.

Objective: We performed a systematic review to characterize the nature of academic research on digital clinical measures and to compare and contrast the types of sensors used and the sources of funding support for specific subareas of this research.

Methods: We conducted a PubMed search using a range of search terms to retrieve peer-reviewed articles reporting US-led academic research on digital clinical measures between January 2019 and February 2021. We screened each publication against specific inclusion and exclusion criteria. We then identified and categorized research studies based on the types of academic research, sensors used, and funding sources. Finally, we compared and contrasted the funding support for these specific subareas of research and sensor types.

Results: The search retrieved 4240 articles of interest. Following the screening, 295 articles remained for data extraction and categorization. The top five research subareas included operations research (research analysis; n=225, 76%), analytical validation (n=173, 59%), usability and utility (data visualization; n=123, 42%), verification (n=93, 32%), and clinical validation (n=83, 28%). The three most underrepresented areas of research into digital clinical measures were ethics (n=0, 0%), security (n=1, 0.5%), and data rights and governance (n=1, 0.5%). Movement and activity trackers were the most commonly studied sensor type, and physiological (mechanical) sensors were the least frequently studied. We found that government agencies are providing the most funding for research on digital clinical measures (n=192, 65%), followed by independent foundations (n=109, 37%) and industries (n=56, 19%), with the remaining 12% (n=36) of these studies completely unfunded.

Conclusions: Specific subareas of academic research related to digital clinical measures are not keeping pace with the rapid expansion and adoption of digital sensing products. An integrated and coordinated effort is required across academia, academic partners, and academic funders to establish the field of digital clinical measures as an evidence-based field worthy of our trust.

(J Med Internet Res 2021;23(9):e29875) doi:10.2196/29875

KEYWORDS

digital clinical measures; academic research; funding; biosensor; digital measures; digital health; health outcomes

Introduction

Digital clinical measures are health outcomes or physiological characteristics of an individual's health, wellness, or condition that are collected digitally with a sensor [1]. Digital sensing products enable rapid assessment of health outcomes and support remote and longitudinal monitoring of patients with chronic diseases under daily living conditions [2-5]. During the COVID-19 pandemic, the utility of digital sensor technologies in clinical research [6], clinical care [7], and public health [8,9] have become even more apparent.

In recent years, digital clinical measures have drawn substantial interest from industry, government agencies, academia, and nonprofit institutions, as digital sensing tools, including consumer products and medical devices, are becoming increasingly popular. Consumer products such as smartwatches and smartphones have become part of daily life for many Americans. These have emerged as popular and multipurpose real-time physiological monitoring products capable of measuring sleep and stress in addition to the more traditional actigraphy and heart rate monitoring. In 2020, 26% of Americans owned a smartwatch [10] and 72% of Americans owned a smartphone [11], with annual sales of over US $70 billion [12]. Apart from consumer products, digital sensing products have demonstrated their efficacy as medical devices, both in clinical and remote home monitoring settings to continuously assess vital signs [13], pulmonary congestion in patients with heart failure [14,15], blood and interstitial glucose in patients with diabetes [3], and more.

To support the development and assessment of digital consumer products and medical devices, the volume of academic research has increased across the total product life cycle of digital clinical measures [16]. However, academic contributions to advancing the safe, effective, ethical, and equitable use of digital clinical measures are poorly characterized and, we hypothesize, underfunded. Trust in digital clinical measures is limited, and engaging the academic community is essential to ensure that the field evolves to be worthy of public trust.

For these reasons, a multi-stakeholder group of experts collaborating on The Playbook [1], a precompetitive collaborative of experts in digital health convened by the Digital Medicine Society (DiMe), set out to investigate the nature of academic research related to digital clinical measures. DiMe is a nonprofit professional society dedicated to advancing digital medicine to optimize health [17]. In this systematic review, we explore the representation of subtypes of academic research on digital clinical measures and compare and contrast the funding support for these subareas of research. This systematic review aims to describe the nature of academic research into digital clinical measures, identify areas of focus and gaps, and explore how and whether funding plays a role. With these findings, we hope to establish an integrated and coordinated effort across academia, academic partners, and academic funders to ensure that the expertise within the field is harnessed to ensure that the rapidly expanding domain of digital clinical measures is established as an evidence-based field worthy of our trust.

Methods

Screening

We conducted a systematic search of peer-reviewed literature indexed in PubMed and published between January 1, 2019, and February 24, 2021. For the purposes of this review, we did not restrict the scope of our search to any single digital clinical measure or area of academic research. A multi-stakeholder team of clinical, academic, technical, and operational experts developed the search terms (Multimedia Appendix 1), inclusion criteria (Textbox 1), and selection of data to be extracted from the final publications (Table 1). A biomedical librarian supported the development of the search terms.

Textbox 1. Inclusion criteria adopted to enable the identification of clearly defined academic research related to digital clinical measures.

Research lead

US-led research (ie, ≥50% US-based authors)

Academic research

Academic research with at least one US-based academic researcher (industry-only research articles were excluded)

Article types

Peer-reviewed journals and full-length conference articles (systematic reviews, meta-analyses, editorials, opinion pieces, case reports, and case studies were excluded)

Sensing modality

All portable biometric monitoring technologies (BioMeTs) that rely upon a biometric sensor, such as microphones and accelerometers [16]. BioMeTs are connected digital medicine products that process data captured by mobile sensors using algorithms to generate measures of behavioral or physiological function. (Note: smartphone apps are excluded if they do not rely upon a biometric sensor.)

Publication date

Papers published between January 1, 2019, and February 24, 2021

Table 1. Data fields extracted from identified academic research.

Field	Definition	Allowed values
Title	N/A[a]	Free text
Authors	Last name, first name	Free text
Author affiliation	N/A	Free text
Journal	Name	Free text
Year	N/A	2019, 2020, 2021
DOI	Digital object identifier: a unique alphanumeric string used to identify content and provide a persistent link to the manuscript's online location.	Free text
Nature of academic research	Academic research measured here by the publication of peer-reviewed journals and full-length conference articles by study teams that include researchers from either a university or academic institute and society or nonprofit foundation.	Verification, analytical validation, measure identification, clinical validation, security, ethics, data rights and governance, usability and utility (human factors/behavioral economics), standards, usability and utility (data visualization), economic feasibility, operations (care), operations (research design), operations (research analysis), and operations (data)
Digital clinical measure	Health outcomes or physiological characteristics of an individual's health, wellness, or condition that are collected digitally with a sensor [1]	Biochemical, movement and activity, physiological (electrical, mechanical, optics and imaging)
Funding sources	Funding information	Government, industry, independent foundation, and unfunded

[a]N/A: not applicable.

Following the PubMed search, we conducted a multistep review process to screen articles for inclusion following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [18]. First, we used natural language processing (ie, a custom built Python script; provided in Multimedia Appendix 2 and available on the digital biomarker discovery pipeline [19]) to select papers based on the "Research Lead" and "Academic Research" criteria (Textbox 1). We also excluded articles with "Review" in the title in this step. Second, two of our three trained analysts (authors MMHS, KR, and AB) independently reviewed each publication title against the inclusion criteria. Third, each remaining abstract was reviewed by two of the three analysts (MMHS, KR, and AB) to determine whether the article met our inclusion criteria (Textbox 1). When there was disagreement between two reviewers during either the title or abstract review phase, the decision whether to advance a publication was resolved by the third analyst. Finally, two of three analysts (MMHS, KR, and AB) reviewed the full text of each of the publications that passed the abstract screening stage, with the involvement of a third analyst to settle interrater disagreements (approximately 15% of papers reviewed), to establish the final list of publications for inclusion. The list of articles excluded from the full-text screening process is given in Multimedia Appendix 3.

Data Extraction and Categorization

Following the screening phase, seven analysts (authors MMHS, KR, AB, AVK, AF, YJ, and WKW) extracted data from the articles included in the data extraction phase and categorized each publication as described in Table 1. The articles were categorized according to the following three criteria: nature of academic research, category of digital clinical measures, and source of funding.

The categories to subgroup the "nature of academic research" included verification, analytical validation, measure identification, clinical validation, security, ethics, data rights and governance, usability and utility (human factors and behavioral economics), standards, usability and utility (data visualization), economic feasibility, operations (care), operations (research design), operations (research analysis), and operations (data).

The categories to subgroup "digital clinical measures" included biochemical, movement and activity, and physiological (electrical, mechanical, and optics and imaging).

"Funding sources" were subgrouped by government, industry, independent foundation, and unfunded. Articles with missing funding information were categorized as unfunded. The details of these categories are given in Table 2.

Table 2. Categories for data extraction.

Category	Definitions	Reference
Nature of academic research		
Verification	Evaluates and demonstrates the performance of a sensor technology within a BioMeT[a], and the sample-level data it generates, against a prespecified set of criteria	[16]
Analytical validation	Evaluates the performance of the algorithm, and the ability of this component of the BioMeT to measure, detect, or predict physiological or behavioral metrics	[16]
Clinical validation	Evaluates whether a BioMeT acceptably identifies, measures, or predicts a meaningful clinical, biological, physical, functional state, or experience in the stated context of use (which includes a specified population)	[16]
Measure identification	Research studies to identify key variables from the information extracted from digital sensors, to support decision-making	[20]
Security	Research studies to assess the risks associated with digital clinical measures and taking necessary measures for information security	[21]
Data rights and governance	Research studies to assess the data access, privacy, and sharing (following the FAIR[b] guiding principle)	[22]
Ethics	Research studies to ensure equity and justice during every step of the development and deployment of digital clinical measures (eg, reduce health disparities or racial injustice)	[23]
Usability and utility (human factors/behavioral economics)	Research studies to investigate human factors associated with digital clinical measures (eg, how usable, useful, or unobtrusive a digital clinical measure can be for an end user). It involves surveys from the participants on user experience.	[24]
Standards	Involves standardization of the data extracted from digital clinical measures for interoperability	[25]
Usability and utility (data visualization)	Involves data visualization/result presentation for all end uses	[24]
Economic feasibility	Research studies to investigate economic feasibility of a digital clinical measure	[26]
Operations (care)	Involves clinicians and economists to design clinical workflow and corresponding evaluation that is typically done for a clinical trial	[27]
Operations (research design)	Involves clinicians and biostatisticians to design a research study and execution plan, which is typically done for a clinical trial via power analysis and statistical analysis plan	[28]
Operations (research analysis)	Involves analyzing data from digital clinical measures (eg, data analyst or data scientists)	[29]
Operations (data)	Involves monitoring data and metadata from digital clinical measures (eg, bioinformatics)	[30]
Digital clinical measures		
Biochemical	Senses biochemicals (eg, sweat sensor or continuous glucose monitors)	[31]
Movement and activity	Tracks movement and activity (eg, step count or actigraph)	[31]
Physiological (electrical)	Senses electrical signals related to physiological phenomena (eg, electrocardiography, electroencephalography, electromyography, bioimpedance, electrodermal activity, or electrooooculography)	[32-34]
Physiological (mechanical)	Senses mechanical signals related to physiological phenomena (eg, phonocardiography, speech, lung sounds, joint acoustic emission, seismocardiography, or ballistocardiography)	[35,36]
Physiological (optics and imaging)	Senses optical signals related to physiological phenomena (eg, photoplethysmography, camera for blood volume pulse, or bioradar)	[37]
Funding sources		
Government	US Government funding agencies	[38]
Industry	Pharma, tech, and medical device industry	[38]
Independent foundation	Universities, private nonprofits, societies, and independent associations	[38]
Unfunded	Investigator initiated with no funding sources explicitly stated	

[a]BioMeT: biometric monitoring technology.

[b]FAIR: Findable, Accessible, Interoperable, and Reusable.

For the data extraction process, each publication was reviewed by at least three of the seven analysts (MMHS, KR, AB, AVK, AF, WKW, and YJ). Each publication was assigned to one or more categories of a particular criterion as a result of two or more votes for a particular category for each publication. This method of subgrouping was used to reduce the impact of

individual analyst subjectivity at this stage. Following the initial categorization, articles falling into the government funding subgroup were further categorized by US government agency (ie, National Institutes of Health [NIH], National Science Foundation [NSF], Department of Defense [DOD], Veteran Affairs [VA], National Aeronautics and Space Administration [NASA], Department of Energy [DOE], and "Other"). The "Other" category constitutes government funding sources that were listed for just one article in our pool. Articles with NIH funding were further subgrouped by NIH institutes and centers [39]. Data was standardized after extraction by the five analysts (MMHS, KR, AVK, AF, and AB), and the details of this process are presented in Multimedia Appendix 4.

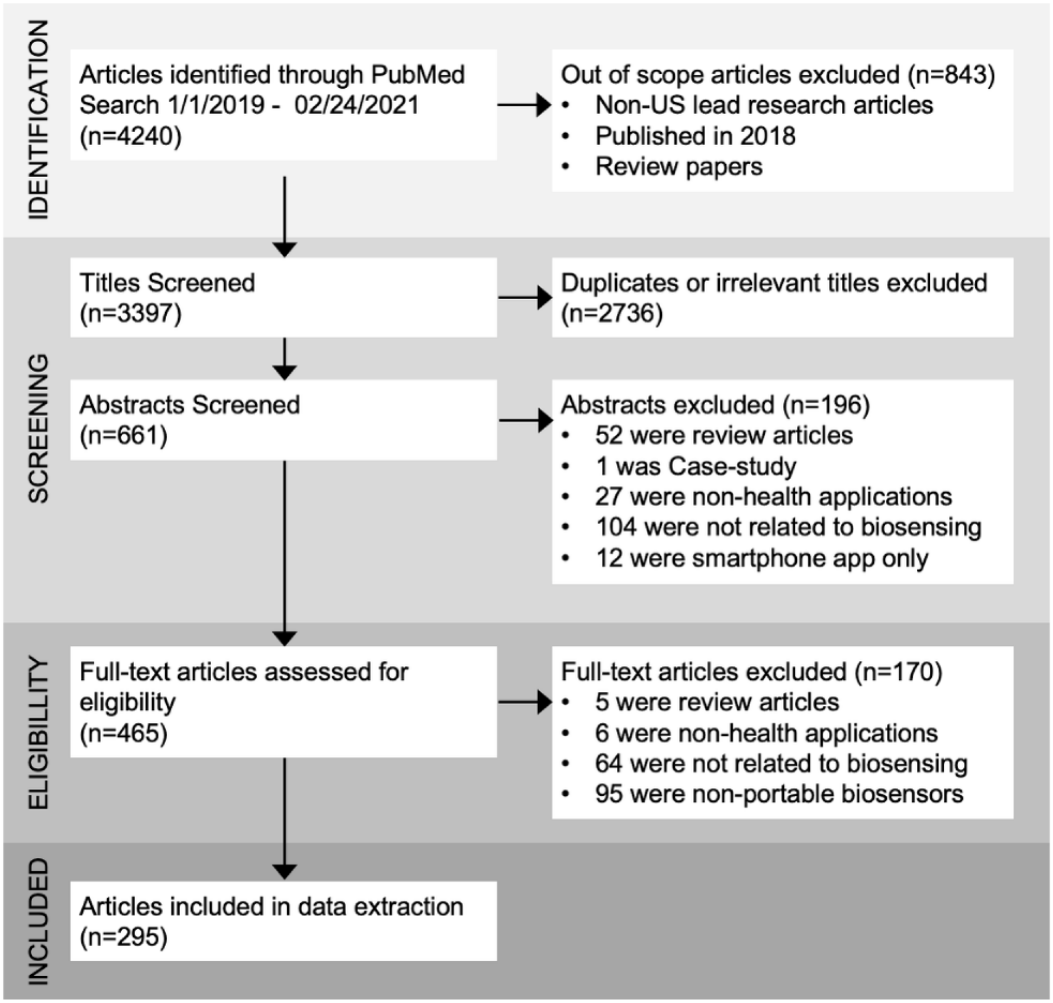
Following the data extraction process, we performed Pearson chi-square tests with one categorical variable to determine whether the representation of academic research studies varies significantly within the following categories: academic research, digital sensors, and funding sources. We assumed equal representation for all categories as the null hypothesis. In this work, we considered *P* values less than .05 to be statistically significant. The statistical tests and data visualization were performed using Python 3.8.5 (Python Software Foundation) on the Spyder-integrated development environment 4.1.5.

Results

Screening

Our initial search on PubMed retrieved 4240 articles (Figure 1). With our custom built Python script, we excluded 843 articles from this initial list based on research lead, publication year, and publication type. Of the 3397 identified articles for subsequent screening, we excluded over 75% (n=2736) after title screening and a further 30% (n=196) after abstract screening, based on our inclusion criteria. The majority of the excluded articles were not related to biosensing, were review articles, or explored nonhealth applications. Following the abstract screening, a total of 465 articles were included in the full-text review, during which we further excluded 170 (37%) articles on the basis of our inclusion criteria. At this stage, articles were mainly excluded because the sensors being studied were nonportable biosensors or because they covered topics unrelated to biosensing; Multimedia Appendix 3 lists the articles excluded in this phase. Data for further analysis was extracted from the remaining 295 articles.

Figure 1. Article screening process and diagram following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) review methodology.

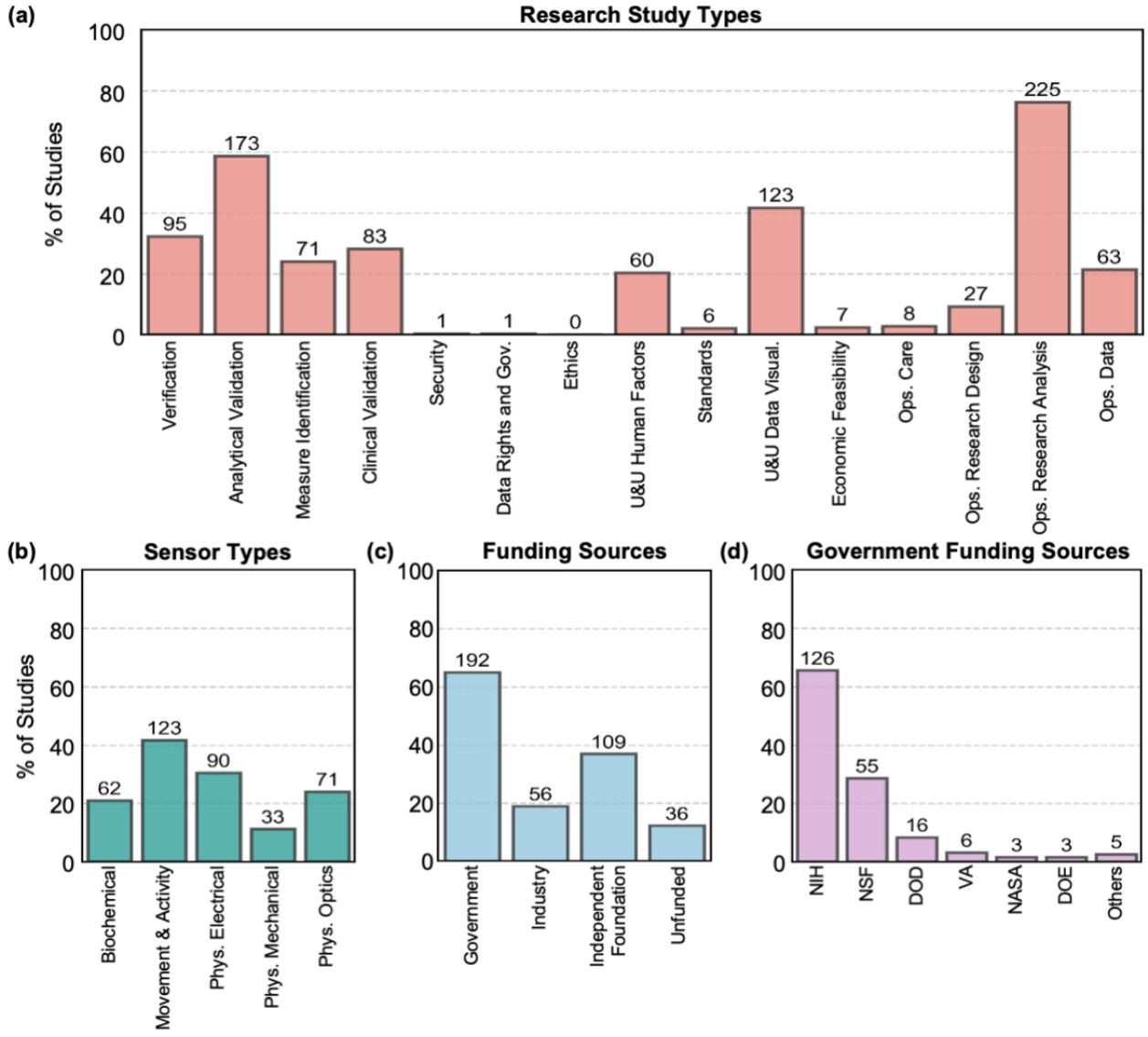


Data Categorization

The 295 articles used for analysis were categorized by research study type, sensor type, and funding source, including broad US government funding sources and specific NIH funding sources (Figure 2). The list of the 295 articles included in the data extraction process with their corresponding categories is located in Multimedia Appendix 4. We observed statistically significant differences (Pearson one-sample chi-square test $P<.001$), indicating unequal distribution of research studies across subcategories for all three overarching categories: academic research, digital sensors, and funding sources. Nearly 76% (n=225) of the studies evaluated were conducted in operations research analysis (Figure 2a). Analytical validation (n=173, 59%), usability and utilities (data visualization; n=123, 42%), verification (n=93, 32%), and clinical validation (n=83, 28%) were other commonly represented study types. On the contrary, ethics (n=0), security (n=1), and data rights and governance (n=1) were uncommon study types. Research on standards (n=6), economic feasibility (n=7), and operations care (n=8) were also uncommon in this pool of articles.

Categorization by sensor types (Figure 2b) revealed that movement and activity were the most commonly studied sensors in the article pool (n=123, 42%), followed by physiological (electrical) sensors (n=90, 31%), physiological (optics and imaging) sensors (n=71, 24%), biochemical sensors (n=62, 21%), and physiological (mechanical) sensors (n=33, 11%). For those studies evaluating movement and activity sensors, actigraphy and activity monitors with wearable accelerometers were the most commonly studied sensors.

Figure 2. Distribution of articles across (a) research study types, (b) different sensing modalities, (c) different funding sources, and (d) different government funding agencies. The bars are showing the percentage of studies, with 100% equivalent to 295 papers included in the data extraction process for (a-c) and 100% equivalent to 192 papers with government funding for (d). The text on top of the bars in all the plots showing the actual number of articles per category. Others in (d): state governments, National Institute of Justice, US Department of Agriculture, National Institute of Food and Agriculture. One article can be grouped into multiple categories for (a-d). DOD: Department of Defense; DOE: Department of Energy; Gov: governance; NASA: National Aeronautics and Space Administration; NIH: National Institutes of Health; NSF: National Science Foundation; Ops: operations; Phys: physiological; U&U: usability and utility; Visual: visualization; VA: Veteran Affairs.



Studies categorized by funding source (Figure 2c) indicated that government agencies are funding the majority (n=192, 65%) of academic research on digital clinical measures, followed by independent foundations (n=109, 37%) and industry (n=56, 19%). Interestingly, more than 1 in 10 digital clinical measures studies (n=36, 12%) was unfunded. Of these unfunded studies, 22 articles explicitly stated that the research team did not receive any external funding, and 14 did not include a statement on funding.

For studies receiving government funding, the NIH was the most frequent contributor in terms of the number of articles funded—66% (n=126) of the studies with government funding were funded by the NIH (Figure 2d). The NSF was the second most frequent government funder of research on digital clinical measures (n=55, 29%), followed by DOD (n=16, 8%), VA (n=6, 3%), NASA (n=3, 2%), DOE (n=3, 2%), and others (the 5 remaining studies were funded by state governments, National Institute of Justice, US Department of Agriculture, or National Institute of Food and Agriculture). Of the 27 institutes and centers at the NIH [39], 24 institutes funded studies on digital clinical measures (Figure 3), indicating widespread interest and applications in this field. The majority of studies were funded by the National Institute of Neurological Disorders and Stroke (n=18, 14% of the studies with NIH funding); the National Institute of Biomedical Imaging and Bioengineering (n=18, 14%); the National Heart, Lung, and Blood Institute (n=17, 13%); and the National Center for Advancing Translational Sciences (n=14, 11%).

Of the articles that reported receiving funding from independent foundations (n=109), 75 (69%) studies received funding from institutional funds at universities, 46 (42%) studies received funding from private nonprofits (eg, Bill and Melinda Gates Foundation or Chan Zuckerberg Initiative), and 6 (6%) received funding from societies and associations (eg, American Heart Association).

To understand whether specific funding types may be driving specific sectors of digital clinical measures research, and where a lack of funding may be contributing to low research output, we explored the distribution of funding across different research study types (Figure 4). Subdividing the research topics by funding type, we found that the proportion of research support from each of the four funding categories was fairly consistent across research areas. The most frequent research and funding combination out of the 295 articles was operations research analysis supported by government funding (n=148, 50%). The second most common combination was analytical validation studies supported by government funding (n=105, 36% of overall studies). Operations research analysis and analytical validation also represent the first- and second-largest sectors of overall digital clinical measures research, respectively (Figure 2a). Interestingly, the third most frequent research and funding combination was not another research category but rather an operations research analysis funded by foundations (n=80, 27% of overall studies), indicating the large overall footprint that operations research analysis occupies in the academic digital clinical measures research space. By contrast, even given the large proportion of government funding, research categorized as analytical validation was also the most likely to be unfunded, with 30 out of the 173 (17%) studies reported as unfunded.

Figure 3. Distribution of articles across different NIH institutes and centers. The bars are showing the percentage of studies, with 100% equivalent to 126 papers with NIH funding. The text on top of the bars showing the actual number of articles per category. One article can be grouped into multiple categories. FIC: Fogarty International Center; NBIB: National Institute of Biomedical Imaging and Bioengineering; NCATS: National Center for Advancing Translational Sciences; NCCIH: National Center for Complementary and Integrative Health; NEI: National Eye Institute; NHLBI: National Heart, Lung, and Blood Institute; NIA: National Institute on Aging; NIAAA: National Institute on Alcohol Abuse and Alcoholism; NICHD: National Institute of Child Health and Human Development; NIDA: National Institute on Drug Abuse; NIDCD: National Institute on Deafness and Other Communication Disorders; NIDCR: National Institute of Dental and Craniofacial Research; NIDDK: National Institute of Diabetes and Digestive and Kidney Diseases; NIEHS: National Institute of Environmental Health Sciences; NIGMS: National Institute of General Medical Sciences; NIH: National Institutes of Health; NIMH: National Institute of Mental Health; NIMHD: National Institute on Minority Health and Health Disparities; NINDS: National Institute of Neurological Disorders and Stroke; NLM: National Library of Medicine; OAR: Office of AIDS Research; OD: Office of Dietary Supplements.

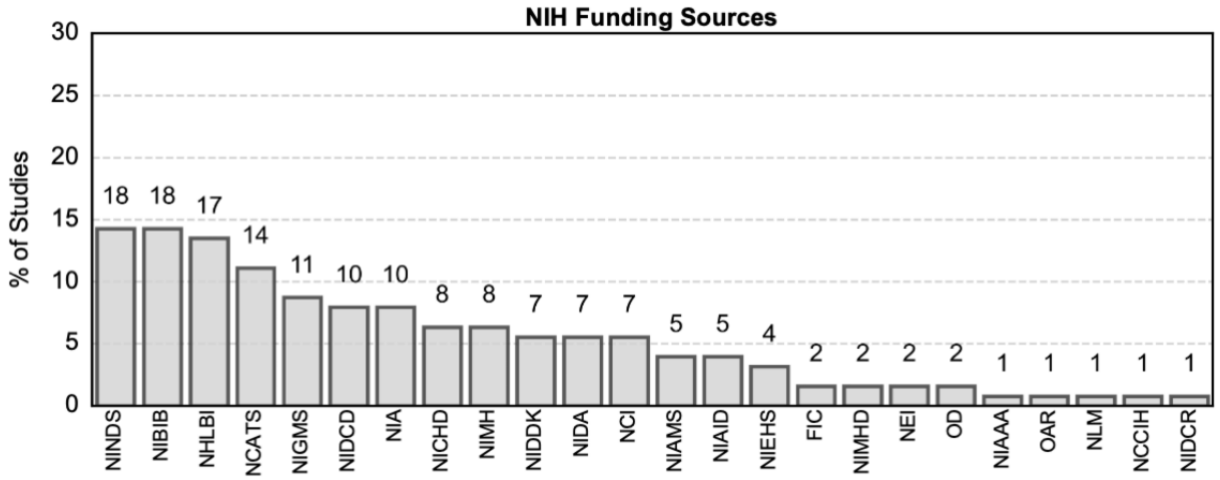
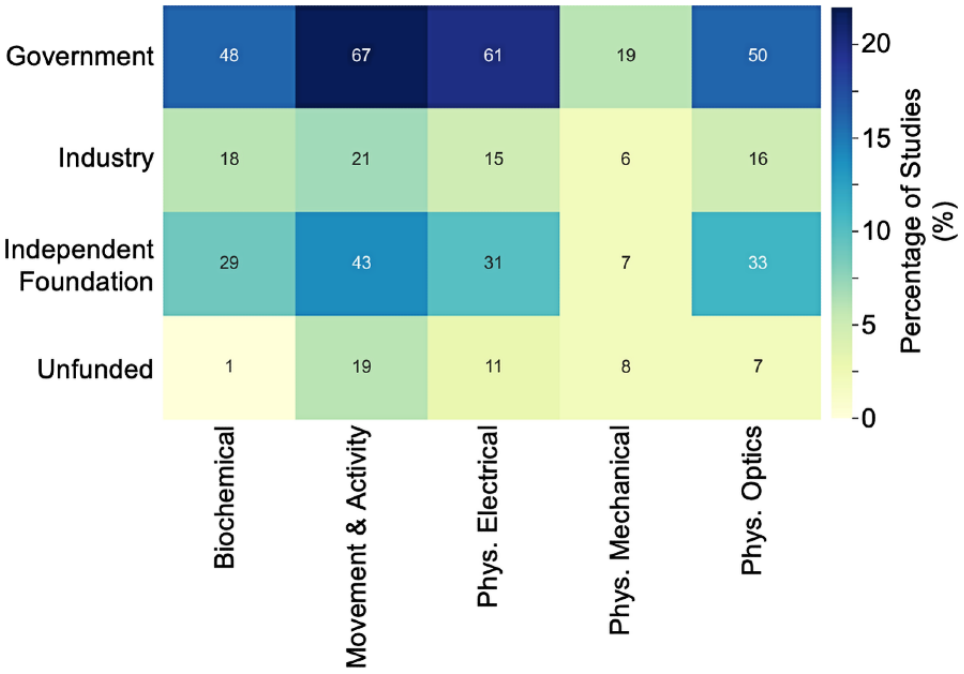


Figure 4. Distribution of funding sources across different research study types, with the heat map color showing the percentage of studies (100% equivalent to 295 papers included in the data extraction process) and the text in each cell showing the actual number of articles per category. One article can be grouped into multiple categories. Gov: governance; Ops: operations; U&U: usability and utility; Visual: visualization.

Similar to the previous analysis subdividing research topics by funding type, we sought to understand whether the volume of literature surrounding particular types of digital sensors is related to funding. Therefore, we subdivided the articles in the different digital sensor categories from Figure 2b by funding type. We found that the most frequent digital sensor and funding combination was movement and activity sensors funded by government agencies (n=67, 23% of overall studies; Figure 5). Government funding also supported the majority of research into physiological (electrical) sensors (n=61, 21%), physiological (optics and imaging) sensors (n=50, 17%), and biochemical sensors (n=48, 16%). By contrast, even given the large proportion of government funding, digital sensors categorized as movement and activity were the second most likely to be unfunded, with 15% (n=19) of the 123 studies on movement and activity sensors being unfunded. Biochemical sensors were the least likely to be studied without funding (n=1, 1% of studies that used biochemical sensors), whereas studies into physiological (mechanical) sensors were most the commonly unfunded, with 8 of 33 (24%) studies reported as unfunded.

Figure 5. Distribution of funding sources across different digital sensor types, with the heat map color showing the percentage of studies (100% equivalent to 295 papers included in the data extraction process) and the text in each cell showing the actual number of articles per category. One article can be grouped into multiple categories. Phys: physiological.



Discussion

Principal Findings

In this systematic review, we describe the nature of academic research related to digital clinical measures and the distribution of funding across different types of academic research and sensing modalities.

Verification, analytical validation, and clinical validation studies [16] are, together, the most frequently published study types in this review. As verification, analytical validation, and clinical validation is foundational to establishing whether a digital clinical measure is fit-for-purpose [16], these findings indicate that academic research supporting the development and evaluation of digital clinical measures is appropriate for a nascent field. However, the paucity of published studies examining the security, data rights and governance, ethics, standards, and economic feasibility of digital sensing products are alarming given the rapid growth and adoption of digital clinical measures [40]. The risks of harm to individuals from unauthorized access to data arising from inadequate security, misuse of data due to poor data rights and governance, and inequities arising from the development and deployment of digital clinical measures without sufficient consideration of the ethical implications are substantial [21,23,41,42]. It is imperative that academic investigators skilled in these areas are motivated and funded to pursue a systematic evaluation of the current state of affairs and to propose best practices to ensure that digital clinical measures fulfill their promise without causing harm to individuals or populations.

Research studies examining the usability and utility of digital sensing products are relatively common compared to publications reporting research into security, data rights and governance, and economic feasibility, which ought to trend together [24]. This is not to say that usability and utility of digital clinical measures is overstudied, but rather suggests that research into these other characteristics of digital sensing products is lagging. Similarly, the number of publications reporting measure identification is relatively low compared to research into the development and deployment of these same measures. This may be cause for concern if we cannot be certain that digital clinical measures being developed have already been determined to be clinically relevant and grounded in aspects of health that patients and clinicians care most about [20]. As we strive to increase the patient focus and efficiency of health care, it is critical that we are separating signals from noise and not advancing digital clinical measures that offer little value to individual patients and the health care system.

Research into the operational aspects of deploying digital clinical measures is the largest single study type identified by our review. Although digital clinical measures cannot add value unless they are successfully operationalized during routine clinical care and in clinical trials, focusing academic research on deployment without first ensuring that the digital clinical measures are fit-for-purpose and trustworthy leaves the entire field of digital health at risk of collecting vast swaths of data that, at best, are of no value and, at worst, could cause harm. During the rapid acceleration of digital clinical measurements, research into the selection and development of high-quality measures and tools must be a primary focus of academic research in this new field.

Research related to movement and activity sensors are most common when we parse the article pool by sensor type. This finding is consistent with other literature where digital measures of activity have been found to be most commonly used to answer clinical questions [43]. Movement and activity sensors also inform the majority of digital end points used by the industry in medical product development [40]. Physiological (electrical),

physiological (optical and imaging), and biochemical sensors are well represented in this review, which is consistent with the recent growth in the use of portable electrocardiograms, photoplethysmography, and continuous glucose monitoring, respectively.

Our review indicates that government agencies and independent foundations are funding most of the academic research studies related to digital clinical measures. Industry funding was relatively low, and this is likely due to our definition of academic studies that excludes studies that only have industry-affiliated authors without academic research partnerships. Of the government agencies, the NIH is funding most of the academic research studies, which is consistent with previous research examining funding of US biomedical research [38]. The distribution of funding across different NIH institutes and centers demonstrates that certain therapeutic areas might be getting more funds compared to others. However, we have not extracted information on funding distribution across different therapeutic areas, as it was out of scope for this systematic review. Future work should explore which therapeutic areas are more likely to receive funding and which areas are least funded. Of the independent foundation–funded research studies, institutional funds at universities are funding the majority on digital clinical measures as compared with private nonprofits and public charities, which is also consistent with the literature [38].

After operational research, analytical validation is the most common government-funded study type in digital clinical measurement. This is critically important as analytical validation includes examination of algorithmic bias [16], which must be an area of focus given research findings that digital sensing products may not perform equally well across different skin tones, among other factors [44,45]. However, although the total number of government-funded analytical validation publications is high, analytical validation studies are also the most likely to be unfunded (n=30, 17%), suggesting that academic researchers are pursuing analytical validation studies even when funding may not exist. This work is to be applauded but is not sustainable. Additional funding for analytical validation must be made available to ensure that digital clinical measures are developed equitably.

Although movement and activity sensors are the most used sensors in academic research, these sensors are still the second most likely to be unfunded (n=19, 15%), suggesting that academic researchers are pursuing research into movement and activity sensors even when funding may not exist. This is again praiseworthy but not sustainable, considering the rapid adoption of these sensors in our daily life [10,11] and clinical studies [43]. Sufficient funding is required to ensure the development and deployment of these movement and activity sensors reliably and equitably.

Our review has several limitations. First, we have focused only on academic research led by US-based academic researchers. Future research should expand beyond the United States to examine trends in academic research into digital clinical measures globally. Second, we searched only one database (PubMed) to retrieve articles for this review. PubMed only indexes research related to life sciences and biomedicine [46]. As digital medicine is a highly interdisciplinary field, many relevant studies may not have been captured in our review. For example, sensor verification studies may be published in traditional engineering journals that are not indexed by PubMed. Future studies will be enhanced by the use of multiple databases across disciplines. In addition, as an emerging interdisciplinary field, we must strive to reference the complete corpus of relevant literature, not only those publications familiar to us in our individual disciplines. Finally, the subjective nature of the review and data extraction process may hinder repeatability, and we attempted to mitigate this risk using innovative methods such as using a majority voting system and using natural language processing to automate the initial screening phase.

This review reports the current state of academic research on the rapidly expanding and highly promising field of digital clinical measures. Substantial work is being done in areas such as validation and operations, with a paucity of research in other areas like security and ethics. Future studies should investigate why critical research into the safe, effective, ethical, and equitable advancement of digital clinical measures is largely absent from the published literature. Both academic researchers and funding agencies should focus on the subareas of academic research on digital clinical measures that are underrepresented and relatively underfunded to ensure that funding priorities adequately reflect the evidentiary needs of the field.

Conclusion

Academic research related to digital clinical measures is not keeping pace with the rapid expansion and adoption of digital sensing products. Although substantial foundational research validating the performance of digital clinical measures is being conducted, academic studies of security, data rights and governance, economic feasibility, ethics, and standards necessary to advance the field are lagging. These areas must be bolstered to minimize the growing chasm between the promised benefits of digital clinical measures and their potential risks. As expected, research funding appears to be associated with increased research publications. An integrated and coordinated effort is required across academia, academic partners, and academic funders to establish the field of digital clinical measures as an evidence-based field worthy of our trust.

Acknowledgments

The authors would like to thank Sarah Park, librarian for engineering and computer science at Duke University, for her support in the development of the search terms. This study was funded, in part, by the Digital Medicine Society. This project has been made possible, in part, by grant 2020-218599 from the Chan Zuckerberg Initiative Doner-Advised Fund, an advised fund of Silicon Valley Community Foundation.

Authors' Contributions

JCG and JD conceived the study. MMHS, JCG, TH, JP, SC, JC, and JD participated in the study's design and coordination, and drafted the search terms and inclusion criteria. MMHS ran the search on PubMed. MMHS, KR, and AB participated in the screening phase. MMHS, KR, AB, AVK, AF, YJ, and WKW participated in the data extraction phase. MMHS, KR, AB, AVK, and AF standardized the extracted data. MMHS carried out the data analysis and created visual representations. MMHS, JCG, and JD drafted the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

TH is employed by Activinsights Ltd. The other authors declare no conflicts of interest.

Multimedia Appendix 1
PubMed search terms.
[DOCX File , 18 KB - jmir_v23i9e29875_app1.docx]

Multimedia Appendix 2
Python code.
[DOCX File , 17 KB - jmir_v23i9e29875_app2.docx]

Multimedia Appendix 3
List of articles excluded after full-text review.
[XLSX File (Microsoft Excel File), 45 KB - jmir_v23i9e29875_app3.xlsx]

Multimedia Appendix 4
List of included papers with categories.
[XLSX File (Microsoft Excel File), 186 KB - jmir_v23i9e29875_app4.xlsx]

References

1. The Playbook From Digital Medicine Society (DiMe). URL: https://playbook.dimesociety.org/ [accessed 2021-04-19]
2. Singhal A, Cowie MR. The role of wearables in heart failure. Curr Heart Fail Rep 2020 Aug;17(4):125-132 [FREE Full text] [doi: 10.1007/s11897-020-00467-x] [Medline: 32494944]
3. Bruen D, Delaney C, Florea L, Diamond D. Glucose sensing for diabetes monitoring: recent developments. Sensors (Basel) 2017 Aug 12;17(8):1866 [FREE Full text] [doi: 10.3390/s17081866] [Medline: 28805693]
4. Li X, Dunn J, Salins D, Zhou G, Zhou W, Schüssler-Fiorenza Rose SM, et al. Digital health: tracking physiomes and activity using wearable biosensors reveals useful health-related information. PLoS Biol 2017 Jan;15(1):e2001402 [FREE Full text] [doi: 10.1371/journal.pbio.2001402] [Medline: 28081144]
5. Iqbal F, Lam K, Joshi M, Khan S, Ashrafian H, Darzi A. Clinical outcomes of digital sensor alerting systems in remote monitoring: a systematic review and meta-analysis. NPJ Digit Med 2021 Jan 08;4(1):7. [doi: 10.1038/s41746-020-00378-0] [Medline: 33420338]
6. Hirten RP, Danieletto M, Tomalin L, Choi KH, Zweig M, Golden E, et al. Use of physiological data from a wearable device to identify SARS-CoV-2 infection and symptoms and predict COVID-19 diagnosis: observational study. J Med Internet Res 2021 Feb 22;23(2):e26107 [FREE Full text] [doi: 10.2196/26107] [Medline: 33529156]
7. Eisenkraft A, Maor Y, Constantini K, Goldstein N, Nachman N, Levy R, et al. Trajectories of key physiological parameters in COVID-19 patients using continuous remote monitoring and health AI. Res Square. Preprint posted online on December 15, 2020. [doi: 10.21203/rs.3.rs-122775/v1]
8. Quer G, Radin JM, Gadaleta M, Baca-Motes K, Ariniello L, Ramos E, et al. Wearable sensor data and self-reported symptoms for COVID-19 detection. Nat Med 2021 Jan;27(1):73-77. [doi: 10.1038/s41591-020-1123-x] [Medline: 33122860]
9. Mishra T, Wang M, Metwally AA, Bogu GK, Brooks AW, Bahmani A, et al. Pre-symptomatic detection of COVID-19 from smartwatch data. Nat Biomed Eng 2020 Dec;4(12):1208-1220. [doi: 10.1038/s41551-020-00640-6] [Medline: 33208926]
10. Smartwatch penetration 2020. Statista. URL: https://www.statista.com/statistics/1107874/access-to-smartwatch-in-households-worldwide/ [accessed 2021-04-21]
11. Smartphone penetration in the US 2021. Statista. URL: https://www.statista.com/statistics/201183/forecast-of-smartphone-penetration-in-the-us/ [accessed 2021-04-21]
12. Statista. URL: https://www.statista.com/topics/2711/us-smartphone-market/ [accessed 2021-04-21]
13. Dias D, Paulo Silva Cunha J. Wearable health devices-vital sign monitoring, systems and technologies. Sensors (Basel) 2018 Jul 25;18(8):2414 [FREE Full text] [doi: 10.3390/s18082414] [Medline: 30044415]

14. Abraham WT, Adamson PB, Bourge RC, Aaron MF, Costanzo MR, Stevenson LW, CHAMPION Trial Study Group. Wireless pulmonary artery haemodynamic monitoring in chronic heart failure: a randomised controlled trial. Lancet 2011 Feb 19;377(9766):658-666. [doi: 10.1016/S0140-6736(11)60101-3] [Medline: 21315441]
15. Amir O, Ben-Gal T, Weinstein JM, Schliamser J, Burkhoff D, Abbo A, et al. Evaluation of remote dielectric sensing (ReDS) technology-guided therapy for decreasing heart failure re-hospitalizations. Int J Cardiol 2017 Aug 01;240:279-284 [FREE Full text] [doi: 10.1016/j.ijcard.2017.02.120] [Medline: 28341372]
16. Goldsack J, Coravos A, Bakker J, Bent B, Dowling A, Fitzer-Attas C, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). NPJ Digit Med 2020;3:55. [doi: 10.1038/s41746-020-0260-4] [Medline: 32337371]
17. Digital Medicine Society (DiMe). URL: https://www.dimesociety.org/ [accessed 2021-04-21]
18. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
19. Bent B, Wang K, Grzesiak E, Jiang C, Qi Y, Jiang Y, et al. The digital biomarker discovery pipeline: an open-source software platform for the development of digital biomarkers using mHealth and wearables data. J Clin Transl Sci 2020 Jul 14;5(1):e19 [FREE Full text] [doi: 10.1017/cts.2020.511] [Medline: 33948242]
20. Manta C, Patrick-Lake B, Goldsack JC. Digital measures that matter to patients: a framework to guide the selection and development of digital measures of health. Digit Biomark 2020;4(3):69-77 [FREE Full text] [doi: 10.1159/000509725] [Medline: 33083687]
21. Carmody S, Coravos A, Fahs G, Hatch A, Medina J, Woods B, et al. Building resilient medical technology supply chains with a software bill of materials. NPJ Digit Med 2021 Feb 23;4(1):34. [doi: 10.1038/s41746-021-00403-w] [Medline: 33623135]
22. McGraw D, Mandl KD. Privacy protections to encourage use of health-relevant digital data in a learning health system. NPJ Digit Med 2021 Jan 04;4(1):2. [doi: 10.1038/s41746-020-00362-8] [Medline: 33398052]
23. Sieck CJ, Sheon A, Ancker JS, Castek J, Callahan B, Siefer A. Digital inclusion as a social determinant of health. NPJ Digit Med 2021 Mar 17;4(1):52. [doi: 10.1038/s41746-021-00413-8] [Medline: 33731887]
24. Coravos A, Doerr M, Goldsack J, Manta C, Shervey M, Woods B, et al. Modernizing and designing evaluation frameworks for connected sensor technologies in medicine. NPJ Digit Med 2020;3:37. [doi: 10.1038/s41746-020-0237-3] [Medline: 32195372]
25. Interoperability in healthcare. Healthcare Information and Management Systems Society. 2020. URL: https://www.himss.org/resources/interoperability-healthcare [accessed 2021-04-21]
26. McNamee P, Murray E, Kelly MP, Bojke L, Chilcott J, Fischer A, et al. Designing and undertaking a health economics study of digital health interventions. Am J Prev Med 2016 Nov;51(5):852-860. [doi: 10.1016/j.amepre.2016.05.007] [Medline: 27745685]
27. Shaw J, Agarwal P, Desveaux L, Palma DC, Stamenova V, Jamieson T, et al. Beyond "implementation": digital health innovation and service design. NPJ Digit Med 2018;1:48. [doi: 10.1038/s41746-018-0059-8] [Medline: 31304327]
28. Finelli L, Narasimhan V. Leading a digital transformation in the pharmaceutical industry: reimagining the way we work in global drug development. Clin Pharmacol Ther 2020 Oct;108(4):756-761 [FREE Full text] [doi: 10.1002/cpt.1850] [Medline: 32294230]
29. Doherty A, Jackson D, Hammerla N, Plötz T, Olivier P, Granat MH, et al. Large scale population assessment of physical activity using wrist worn accelerometers: The UK Biobank Study. PLoS One 2017;12(2):e0169649 [FREE Full text] [doi: 10.1371/journal.pone.0169649] [Medline: 28146576]
30. Badawy R, Hameed F, Bataille L, Little MA, Claes K, Saria S, et al. Metadata concepts for advancing the use of digital health technologies in clinical research. Digit Biomark 2019;3(3):116-132 [FREE Full text] [doi: 10.1159/000502951] [Medline: 32175520]
31. Dunn J, Runge R, Snyder M. Wearables and the medical revolution. Per Med 2018 Sep;15(5):429-448 [FREE Full text] [doi: 10.2217/pme-2018-0044] [Medline: 30259801]
32. Mirvis D, Goldberger A. Chapter 13: Electrocardiography. In: Heart Disease. A Textbook of Cardiovascular Medicine. Philadelphia, PA: WB Saunders; 2001:82-128.
33. Cybulski G, Strasz A, Niewiadomski W, Gąsiorowska A. Impedance cardiography: recent advancements. Cardiol J 2012;19(5):550-556. [doi: 10.5603/cj.2012.0104] [Medline: 23042327]
34. Jackson A, Bolger D. The neurophysiological bases of EEG and EEG measurement: a review for the rest of us. Psychophysiology 2014 Nov;51(11):1061-1071. [doi: 10.1111/psyp.12283] [Medline: 25039563]
35. Kovács F, Horváth C, Balogh AT, Hosszú G. Fetal phonocardiography--past and future possibilities. Comput Methods Programs Biomed 2011 Oct;104(1):19-25. [doi: 10.1016/j.cmpb.2010.10.006] [Medline: 21146247]
36. Inan OT, Migeotte P, Park K, Etemadi M, Tavakolian K, Casanella R, et al. Ballistocardiography and seismocardiography: a review of recent advances. IEEE J Biomed Health Inform 2015 Jul;19(4):1414-1427. [doi: 10.1109/JBHI.2014.2361732] [Medline: 25312966]

37. Kamshilin AA, Nippolainen E, Sidorov IS, Vasilev PV, Erofeev NP, Podolian NP, et al. A new look at the essence of the imaging photoplethysmography. Sci Rep 2015 May 21;5:10494. [doi: 10.1038/srep10494] [Medline: 25994481]
38. Dorsey ER, de Roulet J, Thompson JP, Reminick JI, Thai A, White-Stellato Z, et al. Funding of US biomedical research, 2003-2008. JAMA 2010 Jan 13;303(2):137-143 [FREE Full text] [doi: 10.1001/jama.2009.1987] [Medline: 20068207]
39. List of NIH institutes, centers, and offices. National Institutes of Health. 2015. URL: https://www.nih.gov/institutes-nih/list-nih-institutes-centers-offices [accessed 2021-04-21]
40. DiMe's Library of Digital Endpoints. Digital Medicine Society (DiMe). URL: https://www.dimesociety.org/communication-education/library-of-digital-endpoints/ [accessed 2021-04-20]
41. Safety first: the nuances of health technology that can hurt your patients. Digit Med Society (DiMe). URL: https://www.dimesociety.org/safety-first-the-nuances-of-health-technology-that-can-hurt-your-patients/ [accessed 2021-04-22]
42. Foschini L, Goldsack J, Continella A, Wang YX. When fitness data becomes research data, your privacy may be at risk. MobiHealthNews. URL: https://www.mobihealthnews.com/author/luca-foschini-jennifer-goldsack-andrea-continella-and-yu-xiang-wang [accessed 2021-04-22]
43. Perry B, Herrington W, Goldsack JC, Grandinetti CA, Vasisht KP, Landray MJ, et al. Use of mobile devices to measure outcomes in clinical research, 2010-2016: a systematic literature review. Digit Biomark 2018;2(1):11-30 [FREE Full text] [doi: 10.1159/000486347] [Medline: 29938250]
44. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 2020;3:18. [doi: 10.1038/s41746-020-0226-6] [Medline: 32047863]
45. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial bias in pulse oximetry measurement. N Engl J Med 2020 Dec 17;383(25):2477-2478 [FREE Full text] [doi: 10.1056/NEJMc2029240] [Medline: 33326721]
46. PubMed. URL: https://pubmed.ncbi.nlm.nih.gov/ [accessed 2021-04-21]

Abbreviations

DiMe: Digital Medicine Society
DOD: Department of Defence
DOE: Department of Energy
NASA: National Aeronautics and Space Administration
NIH: National Institutes of Health
NSF: National Science Foundation
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
VA: Veteran Affairs

Edited by R Kukafka; submitted 23.04.21; peer-reviewed by C Manta, Y Man; comments to author 28.06.21; revised version received 02.07.21; accepted 12.08.21; published 15.09.21.

Please cite as:
Shandhi MMH, Goldsack JC, Ryan K, Bennion A, Kotla AV, Feng A, Jiang Y, Wang WK, Hurst T, Patena J, Carini S, Chung J, Dunn J
Recent Academic Research on Clinically Relevant Digital Measures: Systematic Review
J Med Internet Res 2021;23(9):e29875
URL: https://www.jmir.org/2021/9/e29875
doi:10.2196/29875
PMID:34524089

Review

Effects of Message Framing on Cancer Prevention and Detection Behaviors, Intentions, and Attitudes: Systematic Review and Meta-analysis

Abidan Ainiwaer[1], PhD; Shuai Zhang[1], PhD; Xiayiabasi Ainiwaer[2], BM; Feicheng Ma[1], MSc

[1]School of Information Management, Wuhan University, Wuhan, China

[2]School of Public Health, Fudan University, Shanghai, China

Corresponding Author:
Feicheng Ma, MSc
School of Information Management, Wuhan University
No 299 Bayi Road, Wuchang District
Wuhan
China
Phone: 86 86 13507119710
Email: fchma@whu.edu.cn

Abstract

Background: With the increasing health care burden of cancer, public health organizations are increasingly emphasizing the importance of calling people to engage in long-term prevention and periodical detection. How to best deliver behavioral recommendations and health outcomes in messaging is an important issue.

Objective: This study aims to disaggregate the effects of gain-framed and loss-framed messages on cancer prevention and detection behaviors and intentions and attitudes, which has the potential to inform cancer control programs.

Methods: A search of three electronic databases (Web of Science, Scopus, and PubMed) was conducted for studies published between January 2000 and December 2020. After a good agreement achieved on a sample by two authors, the article selection (κ=0.8356), quality assessment (κ=0.8137), and data extraction (κ=0.9804) were mainly performed by one author. The standardized mean difference (attitude and intention) and the odds ratio (behaviors) were calculated to evaluate the effectiveness of message framing (gain-framed message and loss-framed message). Calculations were conducted, and figures were produced by Review Manager 5.3.

Results: The title and abstract of 168 unique citations were scanned, of which 53 were included for a full-text review. A total of 24 randomized controlled trials were included, predominantly examining message framing on cancer prevention and detection behavior change interventions. There were 9 studies that used attitude to predict message framing effect and 16 studies that used intention, whereas 6 studies used behavior to examine the message framing effect directly. The use of loss-framed messages improved cancer detection behavior (OR 0.76, 95% CI 0.64-0.90; *P*=.001), and the results from subgroup analysis indicated that the effect would be weak with time. No effect of framing was found when effectiveness was assessed by attitudes (prevention: SMD=0.02, 95% CI –0.13 to 0.17; *P*=.79; detection: SMD=–0.05, 95% CI –0.15 to 0.05; *P*=.32) or intentions (prevention: SMD=–0.05, 95% CI –0.19 to 0.09; *P*=.48; detection: SMD=0.02, 95% CI –0.26 to 0.29; *P*=.92) among studies encouraging cancer prevention and cancer detection.

Conclusions: Research has shown that it is almost impossible to change people's attitudes or intentions about cancer prevention and detection with a gain-framed or loss-framed message. However, loss-framed messages have achieved preliminary success in persuading people to adopt cancer detection behaviors. Future studies could improve the intervention design to achieve better intervention effectiveness.

(J Med Internet Res 2021;23(9):e27634) doi:10.2196/27634

KEYWORDS

gain framing; loss framing; attitude; intention; behaviors; cancer prevention; cancer detection

Introduction

Background

Cancer accounts for 1 in 6 deaths globally. The number of cancer cases worldwide may increase by 60%, especially in low-income and middle-income countries where the increase may be as high as 81%, as estimated by the latest World Cancer Report [1]. This trend will undoubtedly lead to socioeconomic pressure and a shortage of medical resources. The current clinical research on cancer medicine mainly focuses on finding the cause of cancer and preventing the spread of cancer cells in the early stage [2,3]. However, these findings can be helpful only if people are sufficiently aware of adopting healthy behavior (ie, prevention and detection). This is a unique form of a health crisis that requires intensive communicative efforts. Therefore, messages need to be carefully designed in cancer communication to achieve positive health outcomes.

Message framing is an effective technique to change health behavior [4-10]. Health messaging attempts to change people's attitudes, intentions, or behaviors toward a specific health topic by emphasizing the expected benefits of undergoing specific health behaviors (ie, gain-framed messaging) or the possible loss if specific health behaviors are not done (ie, loss-framed messaging), to persuade people to follow healthy guidelines. According to the theory of reasoned action, attitude (or intention) is an essential direct predictor of behavior [11-13]. Moreover, attitudes, intentions, and behaviors are the standard measures of the effectiveness of health messages [4,14].

Health behaviors include disease prevention and detection behavior [4,6]. The former aims to avoid illness or deterioration, while the latter aims to reflect the presence or absence of risk. Gain-framed messaging fulfills the promise of a safer and more certain disease prevention measure in terms of disease prevention, so it is more effective than loss-framed messaging [4,15]. O'Keefe and Jensen [16] also showed the advantage of the gain-framed messaging in disease prevention, but they indicate that it does not apply to skin cancer [17,18]. On the other hand, in terms of disease detection, it was thought that the loss-framed messaging showed a higher persuasive effect [4,15]. However, the meta-analysis results show no significant difference between gain and loss framing [6,16]. Thus, apart from demonstrating the contribution of loss-framed messaging to breast cancer detection [17], there has been no definitive conclusion regarding cancer and message framing.

Objectives

The concept of message framing is an essential strategy in health promotion, but research on how to optimally frame cancer prevention and detection messaging is scarce. Furthermore, while the effectiveness of message framing in cancer prevention and detection remains unclear, there is still an increasing number of related studies that combine message framing with other variables (ie, color) to verify the strengthening or weakening of the framing effect. Therefore, a thorough examination of the effectiveness of the message framing in the context of cancer is needed.

This review mainly aims to systematically summarize the characteristics of the relevant intervention studies and then pool the effect sizes from the relevant studies to quantify the effects of those interventions on attitude, intention, and behavior change. The findings of this review can provide recommendations for researchers and clinicians to design effective messages in cancer communication.

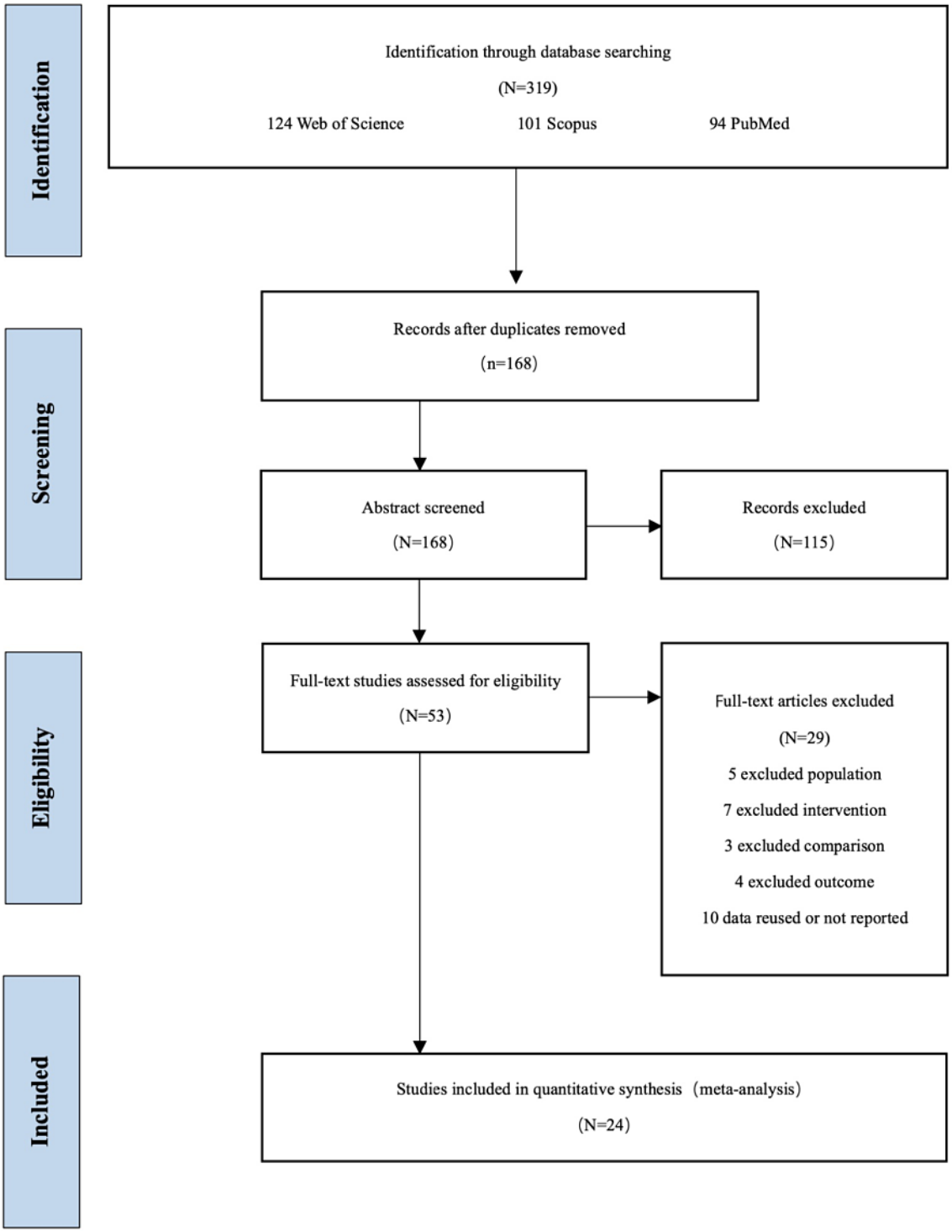
Methods

Search Strategies

This review was conducted and is reported according to the PRISMA (preferred reporting items for systematic reviews and meta-analyses) guidelines. As mentioned in the initial registration (registration ID CRD42021252658), a series of structured electronic searches were performed in three English databases, including Web of Science, Scopus, and PubMed, focusing on the effectiveness of message framing related to a cancer topic and dated from January 1, 2000, to December 31, 2020. The procedures guiding article inclusion are presented in the flow chart in Figure 1. Example search terms are as follows: goal fram*, loss fram*, message fram*, goal fram*, health message, health infor*AND attitude, intention, behavio*, behavio*intention, behavio* chang* AND cancer, screen*, prevention, detection, AND specific validated database filters for randomized controlled trials

All articles identified in the search strategy were exported into reference management software (version 2.48.0; Mendeley) for duplicate checking and further screening. The reference lists of eligible articles were further reviewed to identify other relevant studies. Relevant reviews that emerged from the search strategy were checked for any additional studies. Grey literature (ie, working papers, unpublished studies, conference proceedings or abstracts, and dissertations) was not considered eligible. The preliminary retrieval process was taken as the total literature volume combining the reviews independently completed by two authors. The authors selected a sample of eligible studies after achieving good agreement (κ=0.8356), with the remainder being selected by one author.

Figure 1. Preferred reporting items for systematic reviews and meta-analysis flowchart.



Study Inclusion and Exclusion Criteria

Studies were included if they met the following criteria: (1) participants (not cancer patient or survivors) of both genders and any age range exposed to either the loss-framed or the gain-framed cancer prevention or detection (not treatment) message in experimental research; (2) interventions that have been delivered via either emphasizing the expected benefits of taking specific health behaviors (gain-framed message) or the possible loss of not taking specific health behaviors (loss-framed message); (3) the gain-framed and loss-framed message must come in pairs as a comparison and related to a specific cancer topic, not cancer in general; (4) the effectiveness of cancer-related messages was measured in terms of change in attitude (by scale), intention (by scale), and behavior (by yes or no), and the analysis of primary data was reported, and sufficient quantitative data was provided for estimating the total effect size. In addition, only randomized controlled trials (RCTs) were

included. Finally, published studies that repeatedly used the same data were excluded, leaving only one study that reported the most complete results.

Data Extraction

Data from a sample of eligible studies was extracted by two authors, and after they achieved good agreement (κ=0.9804), the remainder was extracted by one author. The main framework of the extraction criteria was drawn by the authors altogether. The following information was extracted: (1) basic study characteristics including the first author, publication date, country, and funding of the study; (2) participant characteristics including sample size, age, ethnic, and gender ratio (female); (3) intervention characteristics including cancer type, research setting (lab or not), message contents, message resources, message delivery channel, intervention duration, and underpinning theories; (4) outcome measures including measurement of the outcomes; and (5) main results including intervention completion ratio and converted effect size (standardized mean difference [SMD] and odds ratio [OR]).

Bias Assessments

The risk of bias for RCTs was independently assessed by two authors (κ=0.8137) using the Cochrane Collaboration tool for assessing bias [19]. Differences of opinion were discussed and agreed upon by the two authors.

Strategy for Data Synthesis and Meta-Analysis

All meta-analyses were performed with Review Manager 5.3 (version 5.3; The Cochrane Collaboration) [20]. This study performed meta-analyses on the three message effectiveness indicators (attitude, intention, and behavior). Among them, for the two continuous variables (attitude and intention), the SMD of the gain-frame (as the intervention group) and the loss-frame (as the control group) are calculated to represent an effect size. Moreover, two subgroups of cancer prevention and cancer detection were designed separately and analyzed under each indicator. For the two-category variable (behavior), the OR of the gain-frame (as the intervention group) and the loss-frame (as the control group) are calculated to represent an effect size. In addition, two subgroups of cancer detection within 6 months (<6 months) or more than 6 months (>6 months) after the experiment were designed for comparison and analysis of the impact of the timing of the framing effect on cancer detection behavior. All meta-analyses used random-effects models. I^2 statistics are used to determine heterogeneity and measure the degree of inconsistency. I^2 values are 25%, 50%, and 75%, corresponding to low, medium, and high levels of heterogeneity, respectively. For high heterogeneity, the source of heterogeneity was investigated by conducting a subgroup analysis to explore potential moderators and demonstrate why heterogeneity existed. We used the funnel plot to observe whether there is publication bias and used the Egger regression line to confirm possible publication bias further.

Results

Study Characteristics

The literature search identified a total of 319 studies extracted by two authors. After deduplication, a total of 168 studies' titles and abstracts were screened, with 53 remaining for full-text screening. Among those, 24 articles met the inclusion criteria, and 4 studies met the inclusion criteria but did not meet the exclusion criteria. The authors attempted to obtain the research data of 8 studies that met the first two inclusion criteria but failed to provide usable data by contacting their corresponding authors via email and received 2 replies [21,22]. Due to the oversize issue, the main content of each study is summarized in Multimedia Appendix 1. Of the 24 studies used in the final analysis, 13 were conducted by the United States [21,23-34], 2 by the United Kingdom [35,36], South Korea [22,37], and China [38,39], and 1 by Australia [40], Ireland [41], Italy [42], Singapore [43], and South Africa [44]. Among the studies used, 5 were funded by the government or institute grant [22,23,25,26,38], 2 by the institute and the center [24,29], 2 by the institute and universities [30,34], 2 by universities [28,37], and 1 by the health charity and behavioral insights project [35]. The remaining 12 studies did not report receiving any funding [21,27,31-33,36,39-44].

The total sample size was 11,637, ranging from n=85 (intervention group) to n=752 (control group). The participants' average age ranged from 12 to 91 years. Because some cancers are associated with a particular sex, 2 studies notably recruited all-male participants [32,42], while another 10 studies recruited all-female participants [21,23,24,26,30,34,35,37,39,44]. Also, there was a range of different cancer topics targeted during the study interventions, including 9 studies for skin cancer [22,25,27,31,33,36,40,41,45], 2 for colorectal cancer [28,38], and 1 for lung cancer [29].

Most of the experiments were conducted online, through email [27-29,40,43,44], a phone call [23] or text message [35], and other methods like webpage URLs [31,37]. There are also a small number of experiments that were done in a laboratory setting using traditional printed materials [21,22,25,32,33,41,42,45]. Messages were delivered in a variety of ways, including text only [27,28,36,44], text plus images or graphics [26,29,38,40], and video [24,30,34,39,43]. The majority of studies were based on prospect theory [21,22,24,27,29,32,34,37,41,43-45], and some of them used theory of planned (or reasoned) behavior [28,31,38,39], self-affirmation theory [36], and the health belief model [30,42]. There are also some studies that did not use theories or models [23,25,35]. The majority of studies focused on attitude [28,31,32,37-40,42,43] and intention [21,22,25-29,33,36,37,39-43,45], whereas 6 studies used behavior to examine the message framing effect directly [23,24,30,34,35,44].

Bias and Heterogeneity Assessments

As shown in Figures 2 and 3, the risk of bias assessment indicated that the included articles were relatively high in quality. For the publication bias examination, the funnel plot

was used for preliminary identification and the Egger regression line to confirm the possible publication bias further.

Figure 2. Risk of bias summary of the individual studies (k=24).

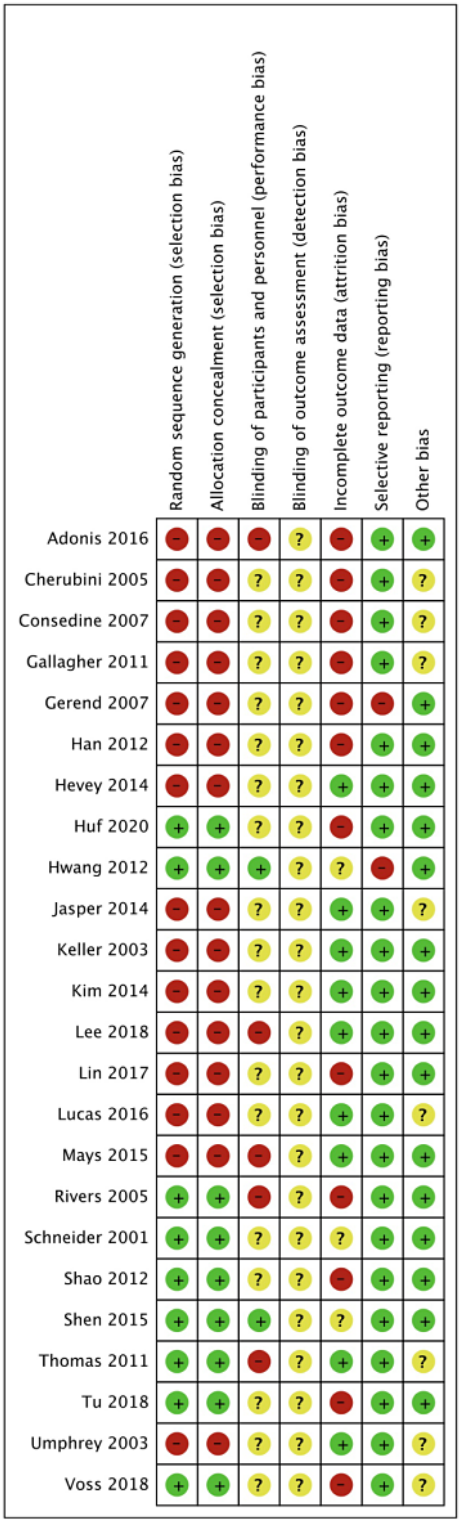
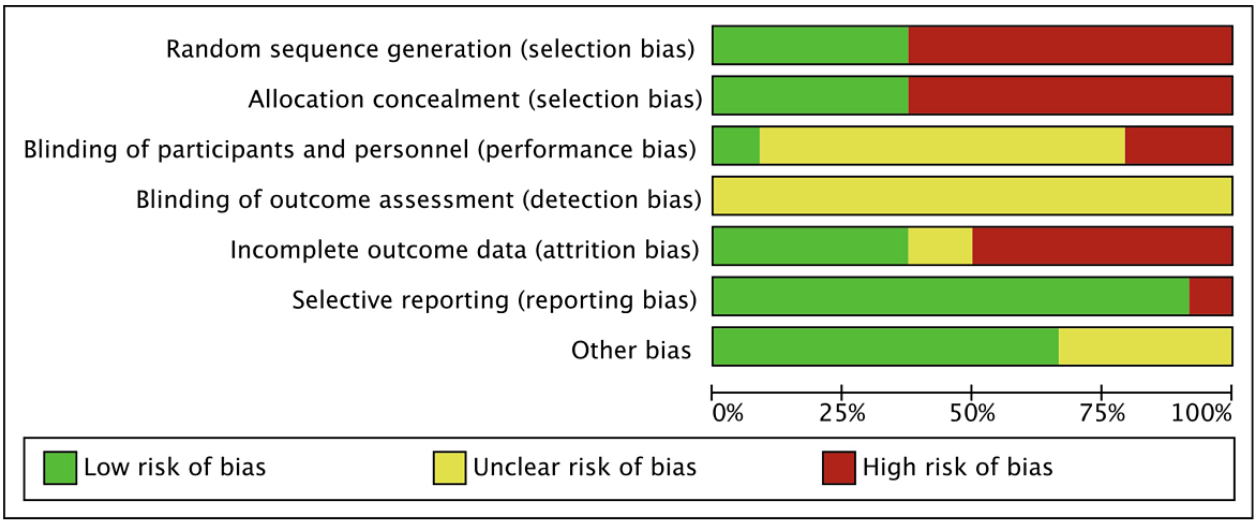


Figure 3. Risk of bias graph.



Intervention Effectiveness

In the following sections, the synthesized results regarding attitude-related, intention-related, and behavior-related outcomes are introduced individually. The main results are visualized in the forest plot.

Attitude-Related Outcomes

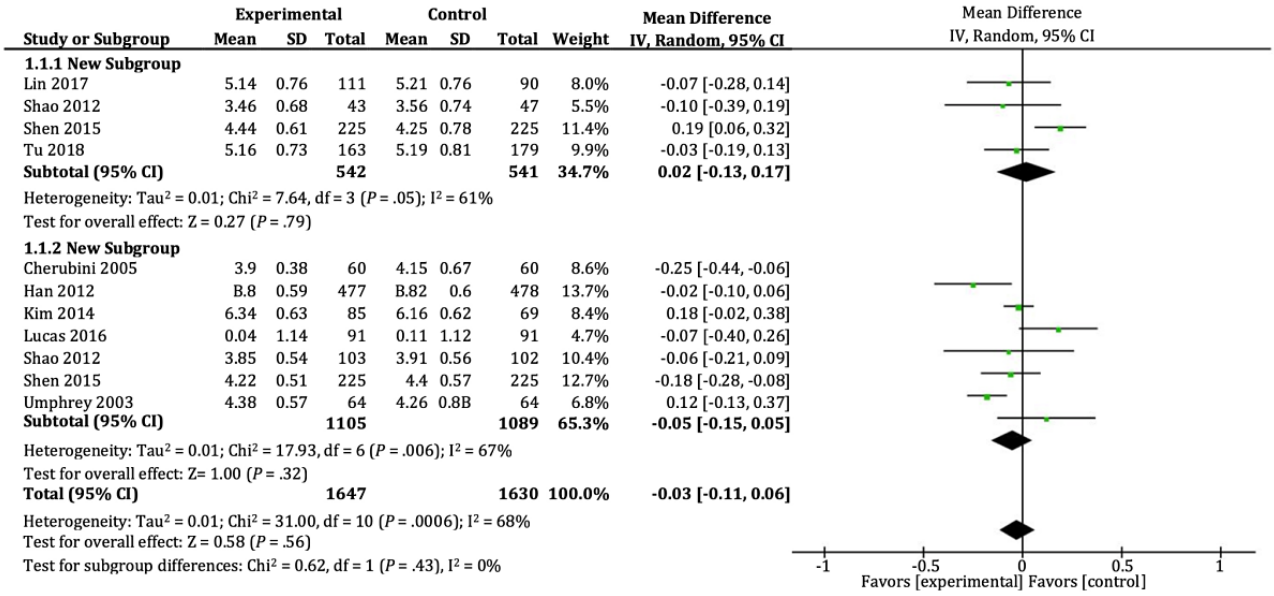
Attitudes are commonly defined as "a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor." In this study, an attitude refers to the tendency towards engaging in a particular behavior or object targeted by a study's intervention.

The attitude was used to indicate the persuasiveness of cancer-related messages in 9 studies. The relevant data involved a total of 3277 participants, including 1647 in the gain group and 1630 in the loss group. As seen in Figure 4, no significant difference was found between the two groups in either cancer prevention attitude (SMD=0.02, 95% CI –0.13 to 0.17; *P*=.79) or cancer detection attitude (SMD=–0.05, 95% CI –0.15 to 0.05; *P*=.32). There was medium heterogeneity (I^2=68%; *P*<.001) across the trials.

The medium heterogeneity can be attributed to the study by Kim [43] and Shen [31]. Participants of both genders were involved in Kim [43], which focused on breast cancer, while the other studies all invited gender-specific participants who were more likely to be affected by the cancer of interest. In addition, Shen [31] invited students to participate in the experiment, and women accounted for 69% of participants. As undergraduates may be more sensitive to information about skin cancer because they are more focused on appearance rather than because they perceive health risks, excluding these 2 studies did decrease the heterogeneity (I^2=0%, *P*=.11), but no significant difference still existed between the gain and loss groups in either cancer prevention intention (SMD=–0.05, 95% CI –0.17 to 0.06; *P*=.37) or cancer detection intention (SMD=–0.05, 95% CI –0.15 to 0.04; *P*=.26).

The funnel plot (Multimedia Appendix 2) demonstrates relative symmetry for the studies of cancer detection or prevention attitude, which suggests no publication bias existed and the above analysis was reliable.

Figure 4. Forest plots of cancer detection attitude and cancer prevention attitude.



Intention-Related Outcomes

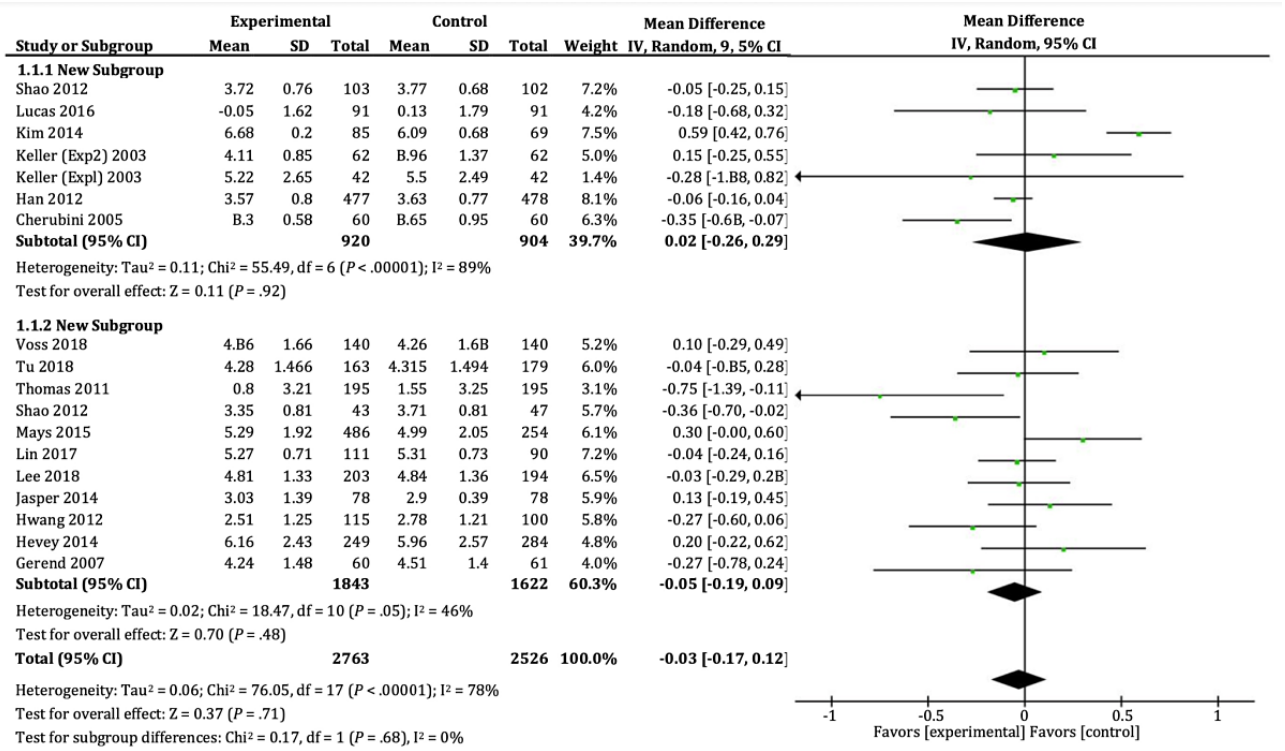
Intention refers to a state of wanting, planning, or expecting to act in a given way. It can be general (ie, intending to mammogram) or specific (i.e., intending to mammogram in 3 months). In this study, the difference between intentions and other related cognitions such as expectations or willingness was not distinguished.

The intention was used to indicate the persuasiveness of cancer-related messages in 16 studies. The relevant data involved a total of 5289 participants, including 2763 in the gain group and 2526 in the loss group. As shown in Figure 5, no significant difference was found between the 2 groups in either cancer prevention intention (SMD=–0.05, 95% CI –0.19 to 0.09; *P*=.48) or cancer detection intention (SMD=0.02, 95% CI –0.26 to 0.29; *P*=.92). There was large heterogeneity (I^2=78%; *P*<.001) across the trials.

Because the large heterogeneity can also be attributed to the study by Kim [43], excluding this study did decrease the heterogeneity (I^2=34%, *P*=.09). In addition, the heterogeneity is due to the gender characteristics of the participants, as mentioned above. However, there is still no significant difference between the gain and loss groups in either cancer prevention intention (SMD=–0.05, 95% CI –0.19 to 0.09; *P*=.48) or cancer detection intention (SMD=–0.08, 95% CI –0.17 to 0.01; *P*=.08).

The funnel plot (Multimedia Appendix 3) demonstrates relative symmetry for the studies of cancer detection or prevention intention except for Kim [43], which suggests that there existed no publication bias after excluding this study.

Figure 5. Forest plots of cancer detection intention and cancer prevention intention.



Behavior-Related Outcomes

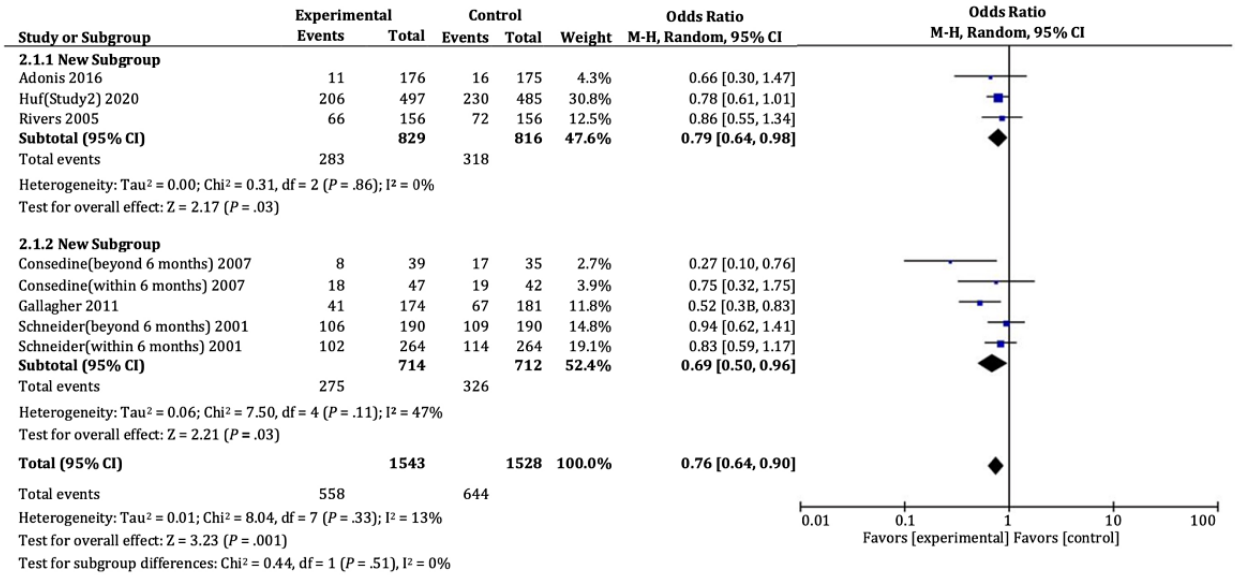
In most cases, behavioral measures are obtained by researchers using self-reported data from participants through follow-up surveys. However, it does not rule out the existence of special access to behavioral data (eg, a health insurance database record).

Behavior was used to indicate the persuasiveness of messages about cancer detection in 6 studies. The relevant data involved a total of 3071 participants, including 1543 in the gain group and 1528 in the loss group. It should be mentioned that cancer detection behavior was measured twice in Schneider et al [30] and Consedine et al [23]. According to Figure 6, loss-framed messages were significantly more likely than gain-framed ones to persuade people into engaging in both cervical cancer detection behavior (OR 0.79, 95% CI 0.64-0.98; *P*=.03) and breast cancer detection behavior (OR 0.69, 95% CI 0.50-0.96; *P*=.03), with no heterogeneity (I^2=13%; *P*=.33).

There was often a time interval between cancer detection behavior and the experiment in which the participants read the framed messages. For the five measurements of cancer detection behavior that were performed within 6 months after the experiments, loss-framed messages were still significantly more persuasive than gain-framed ones (OR 0.76, 95% CI 0.61-0.95; *P*=.01), with no heterogeneity (I^2=0%; *P*=.82). However, the intervention of loss framing became ineffective beyond 6 months (OR 0.73, 95% CI 0.49-1.11; *P*=.14), with medium heterogeneity (I^2=59%; *P*=.09).

The funnel plot (Multimedia Appendix 4) demonstrates asymmetry for the studies of cancer detection behavior, from which one cannot tell whether publication bias existed. However, the Egger linear regression analysis suggests there was no publication bias (*P*=.09).

Figure 6. Forest plots of cervical cancer detection behavior and breast cancer detection behavior.



Relationships Between Behavior and Attitude or Intention

Given that significant results were obtained for cervical or breast cancer detection behavior, 4 studies of detection attitude or intention focusing on either of these two cancers were extracted to examine the relationships between behavior and attitude or intention. Excluding Kim [43] as a source of heterogeneity, the relevant data involved a total of 1284 participants, including 641 in the gain group and 643 in the loss group. No significant difference was found between the two groups in cancer detection attitude or intention (SMD=–0.05, 95% CI –0.14 to 0.03; *P*=.20), with no heterogeneity (I^2=0.0%; *P*=.76). That is, as far as cervical and breast cancers are concerned, cancer detection behavior was not necessarily preceded by cancer detection attitude or intention. The funnel plot demonstrates relative symmetry for the studies of cancer detection attitude or intention, which suggests no publication bias existed, and the above analysis was reliable.

Discussion

Principal Findings

This study meta-analyzed 24 empirical studies on the persuasive effect of message framing in cancer prevention and detection. By subdividing the convincing effect into three dimensions, attitude, intention, and behavior, the results show that gain-framed and loss-framed messages have no significant differences in people's attitudes and intentions to change their cancer prevention or detection behavior. However, for cancer detection behavior, especially breast and cervical cancer detection, loss-framed messaging is more persuasive than gain-framed messaging, and this effect gradually weakens over time.

This meta-analysis did not identify any advantages of gain-framed messaging. Previous studies have proposed that gain-framed messaging is more persuasive than loss-framed in disease prevention, especially preventative behavior [6,16]. Since no relevant data on prevention behavior in the cancer-related literature are included in this meta-analysis, it cannot further verify whether the predecessors' benefits in disease prevention behavior can be equally applicable to the cancer topic. The results also negated the comparative advantage of loss framing in disease detection. There are several possible reasons for this contradictory conclusion. First, the previous meta-analysis did not distinguish between the three measurement indicators, so the results were affected by a large number of measurement indicators [16]. Second, the amount of cancer-related literature included in the previous meta-analysis is small, so the conclusions drawn do not apply to all cancer topics [6]. Therefore, it is necessary to verify whether these existing conclusions can be applied to all health topics based on previous studies. Moreover, because of different health topics, the focus also varies. Some emphasize daily prevention (such as using sunscreen), while some require regular detection (such as a colonoscopy).

This meta-analysis also supports some of the results of previous studies. Neither gain-framed nor loss-framed messaging can exert the framing effect of changing people's attitudes or intentions [6]. Our meta-analysis includes multiple cancer themes, and it also expands the positive impact of the loss framing found in breast cancer detection [46].

One interesting finding of this meta-analysis is that the message framing effect is time-limited. Others have questioned the duration of the framing effect in explaining the declining return rate [34]. On the other hand, this meta-analysis evaluates a larger scale of data and finds that the framing effect diminishes over time. Therefore, future research could further investigate how often health information stimulation can exert the framing effect and determine the effect of repeating the same framing.

Another notable result of this study is the disconnect between people's attitudes, intentions, and behavior. When reorganizing data related to attitudes and intention for breast and cervical cancer prevention and detection, these results also show that attitudes and intentions cannot effectively predict behavior,

indicating that there may indeed be a disconnect between people's attitudes, intentions, and behavior. This conclusion has also been confirmed by previous studies [6,47]. The ultimate purpose of any framed message, but especially a health message, is to promote a specific behavior. In the research process, paying too much attention to the influence of other variables besides behavior may outweigh the gains. Further verification is worthwhile to measure whether subjective attitudes and intentions to target behaviors can replace or predict the subsequent behavior after reading the framed message. The exploration in this area cannot stop at testing latent variables such as attitude and intention, and it should return its focus to "behavior" itself [48].

Limitations

As with all meta-analyses, this study was limited by the included literature measurement indicators and provided data. For example, the measurement data of attitude and intention included in the literature far exceed behavior-related data. Cancer detection behavior data were obtained only due to the lack of actual behavioral data related to cancer prevention. Also, the data are limited to two types of cancer (breast cancer and cervical cancer). Its generalizability also needs to be further verified based on more relevant research. Due to the low email response rate (2 out of 8), this meta-analysis excluded some studies meeting the inclusion criteria.

Conclusions

This meta-analysis shows that gain framing and loss framing have no significant difference in attitudes and intentions about cancer prevention and detection. On the other hand, loss framing can promote breast cancer and cervical cancer detection better than gain framing. However, the effect of the loss framing gradually weakens over time. Therefore, when constructing a message to promote cancer detection, more considerations can be given to loss framing's promised short-term effects. People can be screened for related cancers quickly by emphasizing the possible risk of not performing cancer detection.

Acknowledgments

AA acknowledges financial support from the National Natural Science Foundation of China (grant 71661167007).

Authors' Contributions

AA and FM conceptualized the study design. AA and SZ collected and analyzed the data. AA, SZ, and XA interpreted the results and wrote the manuscript. AA and SZ revised the manuscript. All authors have read and approved the final draft of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed records of eligible studies.
[DOCX File , 25 KB - jmir_v23i9e27634_app1.docx]

Multimedia Appendix 2
Funnel plot depicting the publication bias of attitude-related studies.
[PNG File , 45 KB - jmir_v23i9e27634_app2.png]

Multimedia Appendix 3
Funnel plot depicting the publication bias of intention-related studies.
[PNG File , 40 KB - jmir_v23i9e27634_app3.png]

Multimedia Appendix 4
Funnel plot depicting the publication bias of behavior-related studies.
[PNG File , 37 KB - jmir_v23i9e27634_app4.png]

References

1. Wild C, Weiderpass E, Bernard WS. World Cancer Report: Cancer research for Cancer Prevention, 2020. International Agency of Research on Cancer, World Health Organization. URL: https://publications.iarc.fr/586 [accessed 2021-01-10]
2. Anderson RL, Balasas T, Callaghan J, Coombes RC, Evans J, Hall JA, Cancer Research UKCancer Therapeutics CRC Australia Metastasis Working Group. A framework for the development of effective anti-metastatic agents. Nat Rev Clin Oncol 2019 Mar;16(3):185-204 [FREE Full text] [doi: 10.1038/s41571-018-0134-8] [Medline: 30514977]
3. Piberger AL, Bowry A, Kelly RDW, Walker AK, González-Acosta D, Bailey LJ, et al. PrimPol-dependent single-stranded gap formation mediates homologous recombination at bulky DNA adducts. Nat Commun 2020 Nov 17;11(1):5863 [FREE Full text] [doi: 10.1038/s41467-020-19570-7] [Medline: 33203852]

4. Rothman AJ, Salovey P. Shaping perceptions to motivate healthy behavior: the role of message framing. Psychol Bull 1997 Jan;121(1):3-19. [doi: 10.1037/0033-2909.121.1.3] [Medline: 9000890]
5. Akl EA, Oxman AD, Herrin J, Vist GE, Terrenato I, Sperati F, et al. Framing of health information messages. Cochrane Database Syst Rev 2011 Dec 07(12):CD006777. [doi: 10.1002/14651858.CD006777.pub2] [Medline: 22161408]
6. Gallagher KM, Updegraff JA. Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. Ann Behav Med 2012 Feb;43(1):101-116. [doi: 10.1007/s12160-011-9308-7] [Medline: 21993844]
7. Graham AL, Fang Y, Moreno JL, Streiff SL, Villegas J, Muñoz RF, et al. Online advertising to reach and recruit Latino smokers to an internet cessation program: impact and costs. J Med Internet Res 2012 Aug 27;14(4):135-147 [FREE Full text] [doi: 10.2196/jmir.2162] [Medline: 22954502]
8. Jiang T, Wu X, Chen Y, Wang Y. The effects of message framing on online health headline selection of female users: A moderation of approach/avoidance motivation. Int J Med Inform 2021 Apr;148:104397. [doi: 10.1016/j.ijmedinf.2021.104397] [Medline: 33503558]
9. Niemiec R, Jones MS, Mertens A, Dillard C. The effectiveness of COVID-related message framing on public beliefs and behaviors related to plant-based diets. Appetite 2021 May 14;165:105293. [doi: 10.1016/j.appet.2021.105293] [Medline: 33992747]
10. Xu X, Yang M, Zhao YC, Zhu Q. Effects of message framing and evidence type on health information behavior: the case of promoting HPV vaccination. AJIM 2020 Dec 07;73(1):63-79. [doi: 10.1108/ajim-02-2020-0055]
11. Triandis HC. Values, attitudes, and interpersonal behavior. Nebr Symp Motiv 1980;27:195-259. [Medline: 7242748]
12. Vezich I, Katzman P, Ames D, Falk E, Lieberman M. Modulating the neural bases of persuasion: why/how, gain/loss, and users/non-users. Soc Cogn Affect Neurosci 2017 Feb 01;12(2):283-297 [FREE Full text] [doi: 10.1093/scan/nsw113] [Medline: 27521303]
13. Matkovic J, Clemens KS, Faasse K, Geers AL. Handwashing Message Type Predicts Behavioral Intentions in the United States at the Beginning of the Global COVID-19 Pandemic. Front Public Health 2021;9:583491 [FREE Full text] [doi: 10.3389/fpubh.2021.583491] [Medline: 34026698]
14. Ballard AM, Davis A, Hoffner CA. The Impact of Health Narratives on Persuasion in African American Women: A Systematic Review and Meta-Analysis. Health Commun 2021 May;36(5):560-571. [doi: 10.1080/10410236.2020.1731936] [Medline: 32122156]
15. Rothman A, Bartels R, Wlaschin J, Salovey P. The strategic use of gain- and loss-framed messages to promote healthy behavior: How theory can inform practice. Journal of Communication 2006;56(SUPPL):S202-S220. [doi: 10.1111/j.1460-2466.2006.00290.x]
16. O'keefe DJ, Jensen JD. Chapter 1: The Advantages of Compliance or the Disadvantages of Noncompliance? A Meta-Analytic Review of the Relative Persuasive Effectiveness of Gain-Framed and Loss-Framed Messages. Communication Yearbook 2006 Jan;30(1):1-43. [doi: 10.1207/s15567419cy3001_1]
17. O'Keefe DJ, Jensen JD. The relative persuasiveness of gain-framed and loss-framed messages for encouraging disease prevention behaviors: a meta-analytic review. J Health Commun 2007;12(7):623-644. [doi: 10.1080/10810730701615198] [Medline: 17934940]
18. O'Keefe DJ, Wu D. Gain-framed messages do not motivate sun protection: a meta-analytic review of randomized trials comparing gain-framed and loss-framed appeals for promoting skin cancer prevention. Int J Environ Res Public Health 2012 Jun;9(6):2121-2133 [FREE Full text] [doi: 10.3390/ijerph9062121] [Medline: 22829794]
19. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928 [FREE Full text] [Medline: 22008217]
20. Sterne JAC, Egger M, Sutton AJ. Meta-analysis software. In: Egger M, Davey-Smith G, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. London, United Kingdom: BMJ Publishing Group; 2001:336-346.
21. Gerend MA, Shepherd JE. Using message framing to promote acceptance of the human papillomavirus vaccine. Health Psychol 2007 Nov;26(6):745-752. [doi: 10.1037/0278-6133.26.6.745] [Medline: 18020847]
22. Hwang Y, Cho H, Sands L, Jeong S. Effects of gain- and loss-framed messages on the sun safety behavior of adolescents: the moderating role of risk perceptions. J Health Psychol 2012 Sep;17(6):929-940. [doi: 10.1177/1359105311428536] [Medline: 22186772]
23. Consedine NS, Horton D, Magai C, Kukafka R. Breast screening in response to gain, loss, and empowerment framed messages among diverse, low-income women. J Health Care Poor Underserved 2007 Aug;18(3):550-566. [doi: 10.1353/hpu.2007.0057] [Medline: 17675713]
24. Gallagher KM, Updegraff JA, Rothman AJ, Sims L. Perceived susceptibility to breast cancer moderates the effect of gain- and loss-framed messages on use of screening mammography. Health Psychol 2011 Mar;30(2):145-152 [FREE Full text] [doi: 10.1037/a0022264] [Medline: 21401248]
25. Jasper JD, Woolf J, Christman SD. Responding to framed health messages: different strokes for different (handedness) folks. Psychol Health 2014;29(6):671-686. [doi: 10.1080/08870446.2014.881999] [Medline: 24428603]
26. Keller P, Lipkus I, Rimer B. Affect, Framing, and Persuasion. Journal of Marketing Research 2018 Oct 10;40(1):54-64. [doi: 10.1509/jmkr.40.1.54.19133]

27. Lee MJ, Kang H. Designing Skin Cancer Prevention Messages: Should We Emphasize Gains or Losses? Message Framing, Risk Type, and Prior Experience. Am J Health Promot 2018 May;32(4):939-948. [doi: 10.1177/0890117117729584] [Medline: 28946758]
28. Lucas T, Hayman LW, Blessman JE, Asabigi K, Novak JM. Gain versus loss-framed messaging and colorectal cancer screening among African Americans: A preliminary examination of perceived racism and culturally targeted dual messaging. Br J Health Psychol 2016 May;21(2):249-267. [doi: 10.1111/bjhp.12160] [Medline: 26333060]
29. Mays D, Niaura RS, Evans WD, Hammond D, Luta G, Tercyak KP. Cigarette packaging and health warnings: the impact of plain packaging and message framing on young smokers. Tob Control 2015 Mar;24(e1):e87-e92 [FREE Full text] [doi: 10.1136/tobaccocontrol-2013-051234] [Medline: 24420310]
30. Schneider TR, Salovey P, Apanovitch AM, Pizarro J, McCarthy D, Zullo J, et al. The effects of message framing and ethnic targeting on mammography use among low-income women. Health Psychol 2001 Jul;20(4):256-266. [doi: 10.1037//0278-6133.20.4.256] [Medline: 11515737]
31. Shen L, Mercer Kollar L. Testing Moderators of Message Framing Effect. Communication Research 2013 Jul 01;42(5):626-648 [FREE Full text] [doi: 10.1177/0093650213493924]
32. Umphrey L. The effects of message framing and message processing on testicular self‐examination attitudes and perceived susceptibility. Communication Research Reports 2003 Mar;20(2):97-105. [doi: 10.1080/08824090309388805] [Medline: 12064616]
33. Voss RP, Corser R, McCormick M, Jasper JD. Influencing health decision-making: A study of colour and message framing. Psychol Health 2018 Jul;33(7):941-954. [doi: 10.1080/08870446.2018.1453509] [Medline: 29667448]
34. Rivers SE, Salovey P, Pizarro DA, Pizarro J, Schneider TR. Message framing and pap test utilization among women attending a community health clinic. J Health Psychol 2005 Jan;10(1):65-77. [doi: 10.1177/1359105305048556] [Medline: 15576501]
35. Huf S, Kerrison RS, King D, Chadborn T, Richmond A, Cunningham D, et al. Behavioral economics informed message content in text message reminders to improve cervical screening participation: Two pragmatic randomized controlled trials. Prev Med 2020 Oct;139:106170. [doi: 10.1016/j.ypmed.2020.106170] [Medline: 32610059]
36. Thomas K, Hevey D, Pertl M, Ní CS, Craig A, Maher L. Appearance matters: the frame and focus of health messages influences beliefs about skin cancer. Br J Health Psychol 2011 May;16(Pt 2):418-429. [doi: 10.1348/135910710X520088] [Medline: 21489067]
37. Han K, Jo S. Does culture matter?: a cross-national investigation of women's responses to cancer prevention campaigns. Health Care Women Int 2012;33(1):75-94. [doi: 10.1080/07399332.2011.630117] [Medline: 22150267]
38. Lin C, Yeh W. How Does Health-Related Advertising with a Regulatory Focus and Goal Framing Affect Attitudes toward Ads and Healthy Behavior Intentions? Int J Environ Res Public Health 2017 Dec 04;14(12):1507 [FREE Full text] [doi: 10.3390/ijerph14121507] [Medline: 29207544]
39. Tu Y, Lin Y, Fan L, Tsai T, Wang H. Effects of Multimedia Framed Messages on Human Papillomavirus Prevention Among Adolescents. West J Nurs Res 2019 Jan;41(1):58-77. [doi: 10.1177/0193945918763873] [Medline: 29560818]
40. Shao W. Framing and Efficacy: The Effect of Regulatory Fit on Skin Cancer Prevention and Detection. Journal of Nonprofit and Public Sector Marketing 2012 Jul;24(3):161-180. [doi: 10.1080/10495142.2012.705176]
41. Hevey D, Dolan M. Approach/avoidance motivation, message framing and skin cancer prevention: a test of the congruency hypothesis. J Health Psychol 2014 Aug;19(8):1003-1012. [doi: 10.1177/1359105313483154] [Medline: 23682057]
42. Cherubini P, Rumiati R, Rossi D, Nigro F, Calabrò A. Improving Attitudes Toward Prostate Examinations by Loss‐Framed Appeals. Journal of Applied Social Psychology 2006;35(4):732-742. [doi: 10.1111/j.1559-1816.2005.tb02144.x]
43. Kim HJ. The impacts of vicarious illness experience on response to gain- versus loss-framed breast cancer screening (BCS) messages. Health Commun 2014;29(9):854-865. [doi: 10.1080/10410236.2013.807903] [Medline: 24266759]
44. Adonis L, Paramanund J, Basu D, Luiz J. Framing preventive care messaging and cervical cancer screening in a health-insured population in South Africa: Implications for population-based communication? J Health Psychol 2017 Sep;22(11):1365-1375. [doi: 10.1177/1359105316628735] [Medline: 26888327]
45. Hoffner C, Ye J. Young adults' responses to news about sunscreen and skin cancer: the role of framing and social comparison. Health Commun 2009 Apr;24(3):189-198. [doi: 10.1080/10410230902804067] [Medline: 19415551]
46. O'Keefe D, Jensen J. The Relative Persuasiveness of Gain-framed and Loss-framed Messages for Encouraging Disease Detection Behaviors: A Meta analytic Review. Journal of Communication (2) 2009:296-316. [doi: 10.1111/j.1460-2466.2009.01417.x]
47. Webb TL, Sheeran P. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychol Bull 2006 Mar;132(2):249-268. [doi: 10.1037/0033-2909.132.2.249] [Medline: 16536643]
48. Baumeister RF, Vohs KD, Funder DC. Psychology as the Science of Self-Reports and Finger Movements: Whatever Happened to Actual Behavior? Perspect Psychol Sci 2007 Dec;2(4):396-403. [doi: 10.1111/j.1745-6916.2007.00051.x] [Medline: 26151975]

Abbreviations

OR: odds ratio
PRISMA: preferred reporting items for systematic reviews and meta-analyses
RCT: randomized controlled trial
SMD: standardized mean difference

Edited by R Kukafka; submitted 01.02.21; peer-reviewed by J Makin, X He; comments to author 16.04.21; revised version received 08.06.21; accepted 27.07.21; published 16.09.21.

Please cite as:
Ainiwaer A, Zhang S, Ainiwaer X, Ma F
Effects of Message Framing on Cancer Prevention and Detection Behaviors, Intentions, and Attitudes: Systematic Review and Meta-analysis
J Med Internet Res 2021;23(9):e27634
URL: https://www.jmir.org/2021/9/e27634
doi:10.2196/27634
PMID:34528887

Review

Mobile Applications in Clinical and Perioperative Care for Anesthesia: Narrative Review

Sabrina Pan[1], BS; Lisa Qia Rong[2], MD

[1]Weill Cornell Medicine, New York, NY, United States

[2]Department of Anesthesiology, Weill Cornell Medicine - New York-Presbyterian Hospital, New York, NY, United States

Corresponding Author:
Lisa Qia Rong, MD
Department of Anesthesiology
Weill Cornell Medicine - New York-Presbyterian Hospital
525 E 68th St
New York, NY, 10065
United States
Phone: 1 212 746 2962
Email: lir9065@med.cornell.edu

Abstract

Background: The increasing use of smartphones by providers and patients alike demonstrates that digital health utilizing mobile applications has the potential to transform perioperative care and education in anesthesia.

Objective: This literature review describes the current scope of the use of mobile applications in anesthesiology.

Methods: Literature was searched using PubMed, Scopus, and clinicaltrials.gov for articles published from January 1, 2010, through April 1, 2020. Only English language studies were included. Articles were included if they examined the use of a mobile health application in the setting of anesthesia or the perioperative (immediate preoperative, intraoperative, and postoperative) period. Studies were excluded if they explored video interventions or did not examine the feasibility or efficacy of the mobile app.

Results: We included 29 articles, and three areas of clinical functionality were identified: patient-centered care (preoperative, intraoperative, and postoperative), systems-based improvement, and medical education. Several studies demonstrate the feasibility and reliability of mobile apps in these areas, but many are only tested for efficacy in simulated environments or with small patient samples

Conclusions: Mobile health applications show promise in improving communication between anesthesiologists, improving workflow efficiency, enhancing medical education, and reducing hospital costs. However, there is a need for validation and improvement before full implementation by the provider, patients, and hospital systems. Future studies are needed to demonstrate meaningful health outcomes to create guidelines and recommendations specific to the application of mobile technology to health care.

(J Med Internet Res 2021;23(9):e25115) doi:10.2196/25115

KEYWORDS

mobile applications; mHealth; perioperative medicine; anesthesia

Introduction

Mobile health (mHealth)—digital health technologies using mobile phones, tablets, and wearables to improve health outcomes—have the potential to rapidly change the face of medicine, health care, and medical education. By 2019, 81% of Americans owned a smartphone, a significant increase from 35% in 2011 [1]. Additionally, there has been an increased level of interest in using mobile apps to monitor health in the US population, with 45,028 mHealth related applications available for download in Apple's App Store in 2019, up from 28,343 in 2015 [2]. New advances in mobile apps provide opportunities for innovative patient care, real-time data delivery, and patient-provider engagement in anesthesiology.

Previous reviews in anesthesia have discussed commercially available digital health apps provided in the Apple App Store or Google Play market [3]. This review, however, describes recent clinically tested digital health apps found in scientific

literature and how these tools may be integrated into anesthesiologists' clinical practice.

This review aimed to describe the current scope of mobile application use within anesthesiology. We classified apps found in recent scientific literature into areas of clinical functionality that were deduced from the analysis of the included studies. We also described the limitations of mobile application studies and suggested future directions.

Methods

Overview

A review was conducted to answer the following research question: What mHealth applications in anesthesiology have been described in the published literature, and what are their utility and future potential in clinical practice?

Database Search

A complete literature search was conducted with the librarian's assistance using PubMed, Scopus, and clinicaltrials.gov for articles published from January 1, 2010, through April 1, 2020. The search terms were "mobile applications" OR "mobile health" OR "digital health" OR "mobile app" OR "iOS app" OR "mHealth" OR "smartphone" AND "anesthesiology" OR "anesthesia" OR "anesthetic" OR "anesthesiology" OR "anesthesia" OR "anesthetic. The search included titles, abstracts, and keywords for all databases.

Article Selection

Title, abstract, and full text of studies were reviewed by the first author (SP), and the senior author (LR) resolved any ambiguities. Only English language studies were included. Articles were included if they examined the use of an mHealth application in the setting of anesthesia or perioperative evaluation. Studies were excluded if they explored video interventions or did not examine the feasibility or utility of the mobile app.

Data Collection and Summary

Each article was assessed by design, population characteristics, methods, area of clinical functionality, and outcomes. A qualitative analysis was then conducted to summarize the depth and breadth of mHealth apps that have been utilized within anesthesiology. Major clinical functionalities were deduced from this qualitative analysis, and each study was then classified into one of these categories. We then described the limitations of the data from these studies and provided a basis for the future direction of mHealth in the field.

Results

Our search returned 80 studies on PubMed, 141 on Scopus, and 5 studies on clinicaltrials.gov, yielding a total of 226 studies, of which 49 were duplicate studies. The titles and abstracts of the remaining 177 articles were screened. We identified 66 that meet the criteria, and their full text was reviewed, after which 29 articles were selected for a qualitative review.

Within the literature, we identified three areas of clinical functionality: patient-centered care, systems-based improvement, and medical education. Patient-centered care was further subdivided into preoperative, intraoperative, and postoperative monitoring. Certain mobile applications within systems-based improvement were also subcategorized into medication safety or adverse event reporting or guideline compliance. An overview of each study's characteristics, methods, and outcomes can be found in Table 1.

Table 1. Characteristics and outcomes of included studies.

Study	Design	Sample, N	Sample characteristics	Category	Methods	Results
Cumino et al 2017 [4]	RCT[a]	84	Pediatric (4-8 years) undergoing elective surgeries	Patient-centered care: Preoperative	Four groups: control (verbal anesthesia), informed (parent given an information leaflet about the anesthetic procedure), smartphone (with a child in holding area), and combined (smartphone or informed). Primary endpoint: m-YPAS[b] in holding area and OR[c]	The OR anxiety in the control group was higher (76.2%; *P*=.001) than in the other groups: informed group (38.1%), smartphone group (23.3%), and combined (19.0%).
Lee et al 2013 [5]	RCT	120	Pediatric (1-10 years) surgical patients	Patient-centered care: Preoperative	Midazolam (M group), smartphone app (S), or midazolam & smartphone (SM). Primary outcome: m-YPAS before and after the intervention.	Group S score was significantly lower than group M (*P*=.01, group SM was significantly lower than group M (*P*<.01), and group SM was significantly lower than group S (*P*<.01).
Rubin et al 2019 [6]	Prospective	86	Adult surgical patients	Patient-centered care: Preoperative	An app was designed to administer the Duke Activity Status Index and the 6-minute walk test. Linear regression was performed to estimate the distance walked during the 6MWT from the number of steps measured by the app.	Steps measured by app and research-grade pedometer demonstrated intraclass correlation of 0.87 (0.79-0.92; *P*<.001). Overall model fit was r^2=.72 for the distance estimation algorithm.
Barrachina et al 2017 [7]	Cross-sectional	20	Intensive care unit patients	Patient-centered care: Intraoperative	Analysis of snapshots from patient monitor and photos using Capstesia app were assessed for concordance of PPV[d], CO[e], and dP/dt[f]	Intraclass correlation coefficient or PPV, CO and max dP/dt were 0.991 (95% CI 0.98-0.99), 0.966 (0.96-0.97) and 0.96 (0.95-0.97), respectively.
Carvalho et al 2019 [8]	Prospective	22	Patients with surgery requiring NMB[g]	Patient-centered care: Intraoperative	NMB grade assessed intraoperative care with TOF[h] ratios obtained by a Stimpod accelerometer versus the new app	For 142 time points, there was no significant difference between the two methods (*P*=.78). However, insufficient data that the app can diagnose a TOF higher than 0.
Desebbe et al 2016 [9]	Prospective	2100	N/A[i]	Patient-centered care: Intraoperative	The simulator is used to display arterial waveforms on a computer screen. Data was obtained with different sweep speeds (6 and 12 mm/s) and randomly generated PPV values, pulse pressures, and vitals. Each metric was recorded 5 times at an arterial height scale X1 and 5 times at an arterial height scale X3. Primary outcome: Reproducibility of PPV_{app}[j] and PPV_{man}[k].	The precision error of PPV_{app} and PPV_{man} was 10% (7%-14%) and 6% (3%-10%), respectively. PPV_{app} shows acceptable accuracy with PPV_{man} when at least three pictures are taken to average PPV_{app} at scale X1 (upper limit of the 95% CI of the measurement error<12%).

Study	Design	Sample, N	Sample characteristics	Category	Methods	Results
Joosten et al 2019 [10]	Cross-sectional	40	Abdominal surgery patients	Patient-centered care: Intraoperative	PPV_{CAP}[l] compared with SVV_{PC}[m] at postinduction, preincision, postincision, the end of surgery, and during every hypotensive episode. PPV and SVV are classified into no fluids, gray zone, and fluid administration. Primary outcome: the overall agreement between PPV and SVV and agreement in the fluid administration category	549 pairs of PPV-SVV data were obtained. The overall agreement of PPV_{CAP} with SVV_{PC} was 79% (κ=0.55), demonstrating moderate agreement with only 1% of all measurements resulting in opposite clinical decisions regarding fluid administration.
Joosten et al 2019 [11]	Prospective	57	Elective coronary artery bypass graft patients	Patient-centered care: Intraoperative	The ability of PPV_{CAP} or PPV_{PC}[n] to predict fluid responsiveness after infusion of 5 ml/kg of colloid. Primary outcome: overall agreement of PPV_{CAP} and PPV_{PC} as well as CO_{CAP}[o] and CO from CO_{TD}[p]	No difference in the ability of PPV_{CAP} and PPV_{PC} to predict fluid responsiveness (AUROC[q] 0.74, 95% CI 0.60-0.84 vs AUROC 0.68, 95% CI 0.54-0.80; *P* =.30). CO_{CAP} did not correlate well with CO_{TD}.
Chiu et al 2019 [12]	RCT	156	Procedural patients, age 19-75 years,	Patient-centered care: Preoperative	Completed VAS-100[r] and NRS-11[s] for pain using paper versus smartphone VAS-100 and NRS-11. Primary outcome: correlation at various time points	Panda VAS-100 and original tool correlated strongly at emergence (r =.93) and upon discharge (r =.94); Panda NRS-11 correlated strongly with the original at emergence (r =.93) and upon discharge (r =.96)
Dahlberg et al 2017 [13]	RCT	1027	>17 y/o undergoing day surgery	Patient-centered care: Postoperative	Patients answered SwQoR[t] daily for 14 days: smartphone app versus standard care (control). Primary outcome: cost-effectiveness.	Net savings of €4.77 (approximately US $5.65) per patient w/ intervention. No difference in SwQoR between the two groups.
Dahlberg et al 2019 [14]	RCT	494	>17 years undergoing day surgery	Patient-centered care: Postoperative	Patients randomized to RAPP[u] daily for 14 days via app versus standard care (no follow-up).	62% of contacts made by patients were on postoperative days 1 to 7 and 38% on postoperative days 8 to 14. Demonstrated follow-up should be more long-term.
Highland et al 2019 [15]	RCT	50	Military surgery patients	Patient-centered care: Postoperative	Mobile app (mCare) group vs control telephone group to complete Defense and Veterans Pain Rating Scale, nerve block questions, satisfaction surveys, and system usability survey after surgery.	No difference in initial response rates between app and control. On day 8, 64% of mCare group completed follow-up versus 48% of the control group (*P*=.29).

Study	Design	Sample, N	Sample characteristics	Category	Methods	Results
Jaensson et al 2017 [16]	RCT	997	Adult day surgery patients	Patient-centered care: Postoperative	SwQoR daily for 14 days via app intervention (RAPP) or standard care (control). Primary outcome: SwQoR on postoperative days 7 and 14	Global SwQoR score was significantly lower (better recovery) in the RAPP group compared to the control group on day 7 (28.23 vs 34.87; $P<.001$) and day 14 (20.12 vs 21.90; $P=.002$).
Ke et al 2019 [17]	Qualitative study	15	Obstetric anesthesiologists (9) and cesarean section patients (15)	Patient-centered care: Postoperative	Structured phone or in-person interviews with patients and anesthesiologists to assess care after a cesarean section and to solicit feedback on a prototype mobile app for postoperative cesarean section care	App focusing on patient education and self-monitoring. 80% of patients interacted with the app >2x. Most accessed resources were controlling pain, an overview about days after surgery, and key contacts.
Soh et al 2019 [18]	RCT	42	Gastric cancer surgical patients	Patient-centered care: Postoperative	Alarm on the app every 60 min for 2 days with nurse dashboard versus app without alarms or dashboard. Primary outcome: ISI[v] to assess the frequency of ISI use.	ISI was higher in the test group versus control but not significant (113.5 vs 93.2; $P=.22$). Active coughing showed significantly higher performance in the test group (107.8) compared with the control (94.8).
Sun et al 2014 [19]	Prospective	62	Pediatric patients (4-18 years)	Patient-centered care: Postoperative	Assessed pain with both paper FPS-R[w] and CAS[x] assessments as well as app-based (Panda) versions post-surgery. Primary outcome: correlation between scores.	Panda FPS-R scores correlated strongly with the original tool ($r > .93$). Panda CAS scores correlated strongly with the original CAS scores at both time points ($r>0.87$); mean pain scores were higher (up to plus 0.47) with Panda than with the original.
Warren-Stomburg et al 2016 [20]	Mixed	83	Adult surgical patients	Patient-centered care: Postoperative	Part I: paper-based questionnaire to identify patients' attitudes towards follow-up techniques. Part II: feasibility test of a mobile app for follow-up.	42.2% prefer to respond by the mobile app in part I, but in part II, adherence to answering questions in the app was only 27.2%. Patients > 60 yrs prefer paper follow-up while patients <40 prefer telemedicine ($P=.001$).
Baumann et al 2019 [21]	Randomized crossover	74	Resident and attending anesthesiologists	Systems-based: Medication safety	Compared simulations with and without the app. Primary outcome: the probability of administering the accurate dosage.	The probability of "accurate" rated dosage was 77.7 (70.9-84.5%) in control versus 94 (90-97.8%) with the app.
Gorges et al 2019 [22]	Feasibility study	N/A	N/A	Systems-based	telePORT app designed with input from anesthesiology assistants for design features and work-domain analysis. Primary outcome: Usage patterns quantified.	telePORT is used more for help requests (approximately 4.5 per day) than team messaging (approximately 1 per day). OR monitoring was frequently utilized (34%). Loss of wireless connectivity was a barrier.

Study	Design	Sample, N	Sample characteristics	Category	Methods	Results
Haffey et al 2013 [23]	Retrospective	23	Opioid conversion apps	Systems-based: Medication safety	Apps with opioid dose conversion abilities identified. Dose calculations of seven commonly used opioid switches were compared between apps.	Of 23 apps, 52% had no stated medical professional involvement. There is a significant difference in mean conversion output for hydromorphone between apps with and without medical professional involvement (0.2256 vs 0.2536; *P*=.038).
Jabaley et al 2018 [24]	Cross-sectional	13,846	Anesthesiologists	Systems-based: Medication safety	A free mobile app (Anesthesiologist) deployed a 10-question survey about sugammadex use and related adverse events.	About 50% of anesthesia providers had access to sugammadex and were given a survey. Anaphylaxis rates are estimated to be between 0.005%-0.098%. 22.7% reported adverse drug reactions.
Lane et al 2012 [25]	Feasibility	40	Anesthesiologists	Systems-based	Development of VigiVU, a mobile app for OR awareness and communication, at Vanderbilt University Medical Center followed by beta testing done with a group of 40.	VigiVU push notifications to iPhone were faster than pager (mean 18 sec, SD 8.2). All beta users continued to use the app for the benefits of situational awareness in up to 4 ORs.
Lelaidier et al 2017 [26]	RCT	52	Anesthesia residents	Systems-based: Guideline compliance	Two simulated crises, with and without MAX app. Primary outcome: technical performance during crisis defined by the ESC[y]. Secondary outcome: nontechnical performance	Mean technical performance was higher in the MAX group vs. control group: 81.6%, SD 11.9) versus 58.6%, SD 10.8; *P*<.001 and nontechnical 33.7%, SD 4.4 versus 30.9%, SD 4.9 points; *P*<.001.
McEvoy et al 2016 [27]	RCT	259	Anesthesiologists	Systems-based: Guideline compliance	A 20-question test regarding clinical scenarios related to ASRA[z] guidelines using ASRA Coag app versus any resource except the app (control). Primary outcome: test score	App group (92.4%, SD 6.6) scored higher than control (68.0%, SD 15.8; *P*<.001). App use increased the odds of selecting correct answers (7.8, 95% CI 5.7-10.7).
Rothman et al 2013 [28]	Cross-sectional	N/A	N/A	Systems-based	Message sent with VigiVu versus paging system. Primary outcome: transmission and receipt times calculated as their differences.	Mean latencies <1 sec for iPad and iPod devices and <4 sec for iPhone. Service performed better than third party paging systems (Aquis paging system had 0.6% incidence of prolonged message delivery, >100 sec)
Rubin et al 2017 [29]	Retrospective	N/A	Clinical anesthesia personnel	Systems-based: Medication Safety	Two years before intervention used as the baseline rates of adverse event reporting. Primary outcome: monthly reporting and same-day adverse event reporting compared before and after app implementation.	Median reported 12 events for the first year, 14 for the second year, and 20 after the introduction of the mobile app (*P*=.01). The rate of same-day reporting increased by 10% after the introduction of the app (*P*=.048).
Burstein et al 2018 [30]	Feasibility	N/A	N/A	Medical education	An app that to teaches Bier block.	It was implemented in 2015 with data collection until 2021.

Study	Design	Sample, N	Sample characteristics	Category	Methods	Results
De Oliveira et al 2013 [31]	RCT	20	Medical students	Medical education	Students were randomized to intervention (iLarynx app for 30 min) versus control (no iLarynx access). Primary outcome: the time required to advance fiberscope from the mouth up to the carina on a mannequin. Failed if carina not seen in <120 sec.	80% of the control group failed versus 20% in the intervention group (P=.01). 24 failed attempts in the control group and 4 in the iLarynx group (P<.005).
Linganna et al 2020 [32]	RCT	18	Anesthesiology residents	Medical education	Traditional intraoperative teaching of transesophageal echocardiography (control) vs access to EchoEducator app. Primary outcome: score increase from preintervention to postintervention assessment.	Intervention group demonstrated a greater increase in score; (plus 19.19%, 95% CI 4.14%-34.24%; P=.02) compared to control.

[a]RCT: randomized controlled trial.

[b]m-YPAS: modified Yale Perioperative Anxiety Scale.

[c]OR: operating room.

[d]PPV: pulse pressure variation.

[e]CO: cardiac output.

[f]dP/dt: max slope of pressure curve.

[g]NMB: neuromuscular block.

[h]TOF: train of four.

[i]N/A: not applicable.

[j]PPV_{app}: pulse pressure variation by app.

[k]PPV_{man}: pulse pressure variation by manual calculation.

[l]PPV_{CAP}: pulse pressure variation using Capstesia.

[m]SVV_{PC}: stroke volume variation using pulse contour analysis.

[n]PPV_{PC}: pulse pressure variation using pulse contour analysis.

[o]CO_{CAP}: cardiac output using Capstesia.

[p]CO_{TD}: cardiac output using thermal dilution.

[q]AUROC: area under the receive operating characteristic

[r]VAS-100: visual analog scale-100.

[s]NRS-11: numeric rating scale-11.

[t]SwQoR: Swedish Quality of Recovery.

[u]RAPP: recovery assessment by phone points.

[v]ISI: incentive spirometer index.

[w]FPS-R: faces pain scale-revised.

[x]CAS: color analog scale.

[y]ESC: European Society for Cardiology.

[z]ASRA: American Society of Regional Anesthesia and Pain Medicine.

Discussion

Patient-Centered Care

Within anesthesia, the perioperative surgical home has arisen as a model that encompasses the goals of patient-centered care [33]. Mobile apps will play a large role in this area as they can directly engage patients before they arrive at the hospital and after their procedures. By providing on-the-go education, monitoring, and behavioral interventions, the anesthesia team can reduce patient anxiety preoperatively, increase communication through real-time mobile interactions perioperatively, and improve compliance with perioperative instructions.

Preoperative Interventions

Mobile apps provide an opportunity to decrease anxiety without additional sedatives in the pediatric population. Preoperative anxiety in pediatric populations is estimated to be as high as 50% [34]. Lee et al randomized 120 children between ages 1 and 10 (ASA I, II) who were undergoing elective surgery to receive intravenous midazolam (M), behavioral intervention through smartphone applications (eg, Soundtouch interactive, Pororo Sticker Book, Angry Bird) tailored to the child's developmental status and preferences (S), or both (SM). While they found that anxiety using the modified Yale Preoperative Anxiety Scale (m-YPAS) was lower post-intervention for all three groups (M group: mean 52.8, SD 11.8 vs mean 41.0, SD 7.0; S group: mean 59.2, SD 17.6 vs mean 36.4, SD 7.3; SM group: mean 58.3, SD 17.5 vs mean 26.0, SD 3.4), the S group had lower anxiety levels relative to the M group ($P<.01$), and the SM group had the lowest level of anxiety ($P<.01$) [5].

In another randomized controlled trial (RCT) of 84 pediatric patients ages 4 to 8, the children's median anxiety levels were lower for patients in the smartphone group (parents verbally informed of the anesthetic procedure and child received smartphone app) versus the control group (parent verbally informed about the anesthetic procedure only). The smartphone group had significantly lower m-YPAS scores compared to the control group, measured in the operating room (OR) before anesthesia induction (55.0 vs 23.4; $P<.001$) [4]. The ability to tailor mobile app choices to a child's developmental stage demonstrates how mobile technology can be customizable for individual patients.

Another novel use for smartphone apps in the preoperative area is to improve dynamic communications using a centralized platform such as Listeo+. The Listeo+ app provides personalized information to patients, reminders for preoperative recommendations, and a channel for various provider communications. Listeo+ is currently being tested in a multicenter RCT to evaluate the percentage of compliance to preoperative recommendations, thereby decreasing the rate of surgery cancellations [35].

Mobile apps, such as a preoperative functional capacity app, can assist with preoperative risk assessment, even before the patient enters the health care facility. At the University of Chicago, the Step Test app integrates the Duke Activity Status Index (DASI) with the 6-minute walk test (6MWT) [6]. Step Test uses voice prompts and Apple's CMPedometer to derive calculated step counts for the 6MWT. Rubin et al demonstrated that Step Test's estimated steps exhibited good agreement with a research-grade pedometer in a cohort of 78 patients (intraclass correlation coefficient of 0.87; $P<.001$). The app facilitates efficient administration of DASI and 6MWT and provides immediate data to providers without occupying clinic staff. Future applications may include improving functional capacity through patient optimization before procedures. In summary, emerging mHealth apps demonstrate their potential to transform preoperative care by decreasing anxiety and providing coordinated, efficient care. Hopefully, in the future, providing an interface to optimize patient health status before surgery to improve outcomes.

Intraoperative Monitoring

Digital mHealth apps have also been tested as tools to enhance intraoperative monitoring. An ideal mobile app for intraoperative monitoring should be easy-to-use, accurate, noninvasive, cost-effective, and reliable. Capstesia (version 1.1.6; Galenic App SL) is an inexpensive and attractive smartphone app used to calculate pulse pressure variations (PPV) and cardiac output (CO) from a picture taken of the arterial waveform on a monitor screen. In a simulated environment comparing 408 pairs of PPV readings, Capstesia's PPV demonstrated acceptable accuracy compared with the manual PPV if at least 3 photos of the waveform were taken (measurement error <12%). Accuracy was improved if it was averaged across 5 photos [9] but is less practical in the OR setting. Additional validation studies showed that Capstesia had a percentage of error of 20% for PPV and 13.8% for CO among 20 patients in the intensive care unit [7]. However, when the app was studied in 57 patients undergoing elective cardiac surgeries, its calculated PPV only weakly predicted fluid responsiveness (sensitivity of 73%, 95% CI 0.54-0.92, and a specificity of 74%, 95% CI 0.48-0.90) [11]. When comparing Capstesia's PPV to stroke volume variation from an uncalibrated pulse wave analysis monitor, Joosten et al found that there was 79% overall agreement between the two, with a kappa coefficient of 0.55 [10]. These results demonstrate the promise and difficulties of creating a mobile app for real-world clinical intraoperative settings.

Digital mHealth apps have also shown promise in intraoperative neuromuscular blockade (NMB) assessment. Often, recovery of NMB is done subjectively due to the lack of accelerators [36,37] and can result in residual weakness and respiratory complications [38]. In a sample of 22 patients, Carvalho et al demonstrated a strong correlation between train-of-four ratios obtained by a standard accelerometer against an Android app used with the phone attached to the patient's hand (R=0.98). The app also had a small mean difference (0.0004, 95% limits of agreement +/ 0.12) against standard accelerometry [8]. Thus, the digital app is a feasible way to assess NMB. However, further studies are needed to demonstrate efficacy before the widespread application of mobile apps as monitoring devices in the OR.

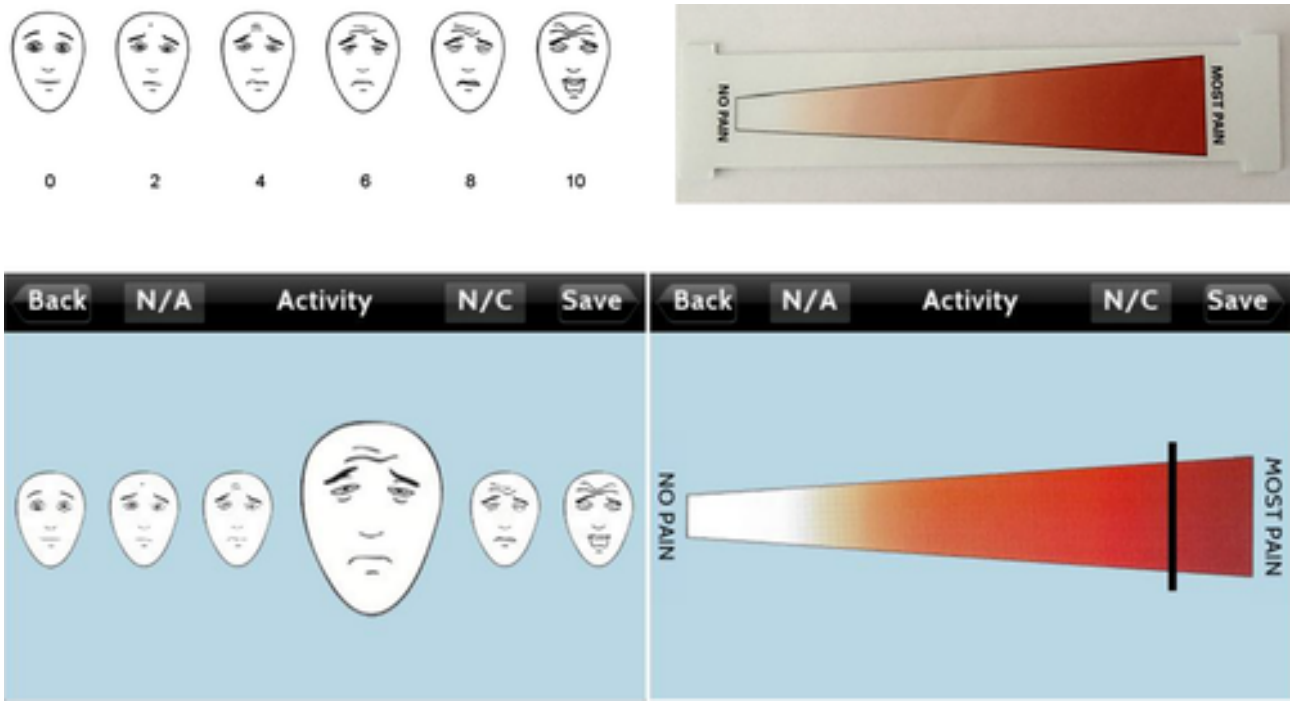
Postoperative Monitoring

There is an increasing focus on patient-reported outcome measures (PROM), alongside traditional biomarkers, in evaluating clinical care [39]. Mobile apps have emerged as an exciting new tool to track PROMs such as symptoms, pain, and satisfaction with health care delivery beyond the inpatient stay. Apps have been tested in clinical trials demonstrating initial feasibility and cost-effectiveness compared to the standard of care [13,15,16].

One important measure that mobile apps help track is poorly managed acute postoperative pain, which can be related to increased risk of psychological morbidity, decreased quality of life (QoL), and chronic postsurgical pain [40]. The application Pain Assessment using a Novel Digital Application (Panda) contains electronic versions of the faces pain scale revised (FPS-R) and color analog scale (CAS). Figure 1 demonstrates both the original FPS-R and the electronic version seen on the

app. In a prospective study of 62 patients, the app version correlated strongly with original scores at various postoperative time points (FPS-R: r>0.93; CAS: r>0.87) [19] and was preferable to the paper version (81% of FPS-R and 76% of CAS participants). Therefore, digital apps may improve adherence to self-reported pain. Additionally, in a study of 156 adult patients undergoing procedures commonly associated with postsurgical pain, Chiu et al found that the app version of the numeric rating scale 11 pain scores were equivalent to that of the paper version when used in adult patients following emergence from anesthesia and at discharge from the postanesthesia care unit [12]. The use of Panda in both adults and pediatric patients demonstrates the app's potential to be used in a variety of patient populations.

Figure 1. Original faces pain scale-revised (top) and adapted electronic faces pain scale-revised for Panda app (bottom) (reproduced with permission from John Wiley and Sons) [19].



Another postoperative mobile app was built for patient-initiated inquiry to identify complications earlier and improve patients' QoL after surgery. An RCT of 1027 day-surgery patients examined postoperative follow-up with a smartphone app known as recovery assessment by phone points (RAPP). The authors found that there were lower health care costs associated with smartphone app follow-up compared to standard of care follow-up (€60.69 vs €37.29; P=.008 with a mean difference of €23.66). However, after the intervention costs were included, net savings were only €4.77 per patient [13]. Patients using the RAPP also reported better global Swedish Quality of Recovery compared to patients using paper assessments on postoperative day 7 (28.23, SD 29.97 vs 34.87, SD 30.68; P<.001) and postoperative day 14 (20.12, SD 26.19 vs 21.90, SD 22.40; P=.002) [16]. The study also helped determine that most patient inquiries (62%) occurred 1 to 7 days postsurgery, most commonly related to the surgical wound (43/119, 36%) and pain (33/119, 28%) [14].

mHealth apps can also incentivize patients to improve respiratory mechanics postoperatively. The Go-breath app was tested in 42 patients who received general anesthesia undergoing a robotic or laparoscopic surgery or laparotomy to encourage the improved use of incentive spirometry (IS). Although it is a nice concept, patients who used the Go-breath app did not score statistically higher on the incentive spirometer index (frequency of IS use over two days) compared to patients who did not (113.5 vs 93.2; P=.22) [18].

Mobile apps can also provide a platform for multidisciplinary care in the postcesarean section period. For example, in a feasibility study including 15 patients of an app designed with obstetric anesthesiologists, 80% of patients used the app at least twice to self-monitor for complications and access pain control resources, thus increasing the amount of patient monitoring for common postcesarean complications and pain [17].

Systems-Based Improvement

For anesthesiologists, smartphone apps are a convenient and accessible way to improve patient safety and quality of care [41,42] through medication safety, adverse event rates, guideline adherence, and responsiveness to emergencies.

For example, immediate response to OR emergencies is of utmost importance, with patient safety and outcomes on the line. Anesthesiologists at Vanderbilt have built an extensive iOS platform known as VigiVu to improve responsiveness through real-time OR videos, vital signs, anesthetic interventions, voice and text communication, and electronic medical record access [25]. Rothman et al demonstrated the superior reliability of the app's push notifications compared to traditional paging systems—only 0.03% of iPhone notification latencies were over 100 seconds compared to 0.6% for third-party paging latencies [28]. Similarly, anesthesiologists

at the University of British Columbia developed an iPhone app, telePORT, to support team-based communication and real-time OR monitoring. Initial results showed that telePORT was successfully integrated for OR help requests (4.5 requests per day) and OR monitoring, representing 34% of app visits [22]. The use of smartphone apps in these situations increases patient safety by prioritizing emergency situations and proper resource allocation. Additionally, data collection from patient monitoring and outcomes in mobile apps can also be utilized for both quality improvement and clinical research in the future.

Medication Safety and Adverse Event Reporting

Mobile apps have also been used to increase adverse event reporting through transparent and convenient processes that help further a culture of safety and quality. An iOS and Android adverse event–reporting app was developed at the University of Chicago (see figure 1 in Rubin et al [29] for an example of the iOS interface used to report events). After implementation, median monthly reporting rates for all providers increased from 12 to 20 (P<.001), with same-day reporting increasing by 10% during the intervention period (P=.048) [29]. Similarly, apps have shown efficacy in monitoring global adverse events regarding the administration of newer drugs. Jabaley et al repurposed a calculator app (Anesthesiologist) to distribute a survey regarding adverse drug reactions associated with sugammadex administration. Using the mobile app, the investigators gathered data from 2770 anesthesia providers that had experience administering the drug across 119 countries [24], demonstrating that an app may be used for crowdsourcing and surveilling new drugs. The app survey found that anaphylaxis rates were estimated to be between 0.005%-0.098%, and 22.7% of survey responders reported witnessing any sugammadex-related adverse drug reactions such as bradycardia and incomplete reversal of NMB [24].

Finally, anesthesiologists have historically used smartphones to calculate drug dosages. Baumann et al compared the probability of administering accurate medication dosages in emergency simulations with a dosage calculator app versus without it [21]. The probability of an accurate dose administration was higher in the app group compared to the control group (94%, SD 90-97.8 vs 77.7%, SD 70.9-84.5) [21]. Apps are often used for medication conversion in the outpatient pain management setting as well. For example, there are many opioid conversion applications designed to reduce medication errors. However, these apps should be used cautiously. In a study examining 23 opioid apps, only 50% provided direct references to sources for their conversion ratios, and over 50% had no documented medical professional involvement [23]. The conversion from 1 mg of oral morphine to oral hydromorphone was statistically different between apps with medical professional involvement versus those without (0.24 vs. 0.25; P=.04), demonstrating that there may be varying degrees of reliability for these mobile apps.

Guideline Compliance

Built-in smartphone-based decision support tools can similarly improve adherence to anesthesia guidelines to reduce technical errors. Numerous reports have found that poor application of guideline standards results in worse patient safety outcomes [43-45]. However, guidelines are numerous, lengthy, and cumbersome to follow. In an RCT across eight institutions, 259 anesthesiologists completed a test requiring the application of the American Society of Regional Anesthesia and Pain Medicine (ASRA) guidelines to clinical situations. The intervention group used an electronic decision support tool, ASRA Coags, which was programmed with the latest guidelines and decision logic, and the control group used any other resource. The authors found that the intervention group had a significantly higher score on a clinical scenario knowledge test (92.4 vs 68.0; P<.001) compared to the control group [27], irrespective of training. Together, these studies show the potential for mobile apps to act as a platform for the rapid application of complex guidelines in clinical situations, potentially reducing adverse events, preventing errors, and improving the quality of care.

Mobile apps have also been developed to improve performance in high-stress situations in lieu of posters, flow charts, and checklists. For example, MAX, a handheld cognitive aid app, was used as an intervention in a RCT of 52 anesthesia residents. Results demonstrated that residents performed better technically, as rated by independent observers when using MAX in simulated crises compared to without MAX use (81.6%, SD 11.9 vs 58.6%, 10.8; P<.001). The app also improved leadership (P=.003), problem-solving (P<.001), and resource-using (P=.006) as assessed by the Ottawa Global Rating scale [26].

Medical Education

The landscape of medical education is changing from the traditional model of classroom teaching to models such as the flipped classroom and asynchronous learning. Mobile technology enables asynchronous learning—a time and location-independent learning model—with greater access to online modules, podcasts, and videos. For example, smartphone apps have now been developed for point-of-care learning of regional anesthetic procedures such as the Bier Block [30]. These apps may provide an all-in-one learning center with videos, cognitive aids, dose calculators, self-timers, and evidence-based references readily available at the learner's convenience [30]. Additionally, education apps use gamification or game-design elements to increase user motivation. Linganna et al developed EchoEducator, a mobile app with transesophageal echocardiography image-based content, and tested it against traditional intraoperative teaching amongst 18 anesthesiology residents over two weeks (please see figure 2 in Linganna et al [32] for examples of questions testing pathology and structures as well as the feedback the resident receives). The app group had a greater knowledge increase than the control group (+19.19%, 95% CI 4.14-34.24; P=.02) in an assessment based on the perioperative transesophageal echocardiography exam [32]. Residents also reported that they would recommend the app to others because of its content customizability and convenience.

Another application of mobile apps in medical education is through simulation. For example, iLarynx utilizes the iPhone's built-in accelerometer to mimic hand movements for fiberoptic intubation. When tested in 20 novice medical students, 80% of students in the standard training group had at least one failed attempt (>120 seconds) of visualizing the carina compared to

20% of students in the iLarynx group (P=.01). There was also continued group improvement in the iLarynx group but not in the standard training group [31].

Current Obstacles and Future Directions

Despite the many functions that mobile technology could fulfill in anesthesia, various obstacles, including low patient participation, privacy and security, provider responsibility, and shifting workplace norms, ultimately impede their comprehensive integration into the clinical environment. For example, preoperative and postoperative monitoring interventions depend on patient interaction and motivation. While there are some reports of high adherence to mobile app interventions [15], other studies show a poorer response rate, especially among the elderly—a large hurdle in app development [20]. Intraoperative apps need to be assessed for ease of use, reliability, and accuracy in the OR and not distract anesthesiologists. Monitoring apps, although promising, need more clinical validation across a range of populations.

More evidence is needed to demonstrate mobile apps can meaningfully improve clinical outcomes moving forward. Large-scale, prospective studies with outcomes data are necessary to provide sufficient evidence for widespread implementation. Investigators should also consider using the mHealth evidence reporting and assessment guidelines, which address the complex nature of mobile technology research [46].

Limitations

Similar to all narrative reviews, this review is limited by the lack of appraisal criteria for included studies. Additionally, only mobile apps derived from the scientific literature have been included. Thus, there are likely mobile applications apps not studied in the literature that were omitted. We have performed a qualitative review of the search and included only studies that fit into the categories we created, thus excluding other studies.

There are additional limitations to the mHealth apps described. Many are often inadequately tested clinically, lack efficacy data, and do not adhere to standard guidelines. Similarly, there are limited guidelines and protocols on how to use the data collected by these apps, despite their potential for engaging a wide population. Future research should explore if patient engagement through smartphone apps may increase efficiency, patient satisfaction, and outcomes during the perioperative period.

Conclusion

With the rapid uptake of smartphones amongst patients and clinicians alike, mobile applications will likely play a larger role within anesthesiology in the future. mHealth apps have novel roles in improving patient care, efficiency, and intraoperative monitoring during surgery. Mobile apps have also been shown to facilitate systems-wide change by creating a culture of improving patient quality and safety. However, as the field of anesthesiology moves forward into the digital health space, demonstrating feasibility is not sufficient; clinicians must critically evaluate mobile app study protocols and rigor. Despite studies with smaller populations and simulated environments, this review still finds emerging evidence that mHealth applications have the potential to significantly improve communication between anesthesiologists, improve workflow efficiency, enhance medical education, and reduce hospitals' costs in the perioperative arena.

Conflicts of Interest

None declared.

References

1. Mobile Fact Sheet. Pew Research Center. Washington D.C; 2019. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2020-06-15]
2. Appfigures. Number of mHealth apps available in the Apple App Store from 1st quarter 2015 to 1st quarter.: Statista URL: https://www.statista.com/statistics/779910/health-apps-available-ios-worldwide/ [accessed 2020-06-15]
3. Rothman BS, Gupta RK, McEvoy MD. Mobile Technology in the Perioperative Arena. Anesthesia & Analgesia 2017;124(3):807-818. [doi: 10.1213/ane.0000000000001858]
4. Cumino D, Vieira J, Lima L, Stievano L, Silva R. Mathias LAST. Smartphone-based behavioural intervention alleviates children's anxiety during anaesthesia induction: A randomised controlled trial. European Journal of Anaesthesiology (EJA) 2017;34(3):A. [doi: 10.1097/eja.0000000000000589]
5. Lee J, Jung H, Lee G, Kim H, Park S, Woo S. Effect of behavioral intervention using smartphone application for preoperative anxiety in pediatric patients. Korean J Anesthesiol 2013 Dec;65(6):508-518 [FREE Full text] [doi: 10.4097/kjae.2013.65.6.508] [Medline: 24427456]
6. Rubin D, Dalton A, Tank A, Berkowitz M, Arnolds D, Liao C, et al. Development and Pilot Study of an iOS Smartphone Application for Perioperative Functional Capacity Assessment. Anesth Analg 2020 Sep;131(3):830-839. [doi: 10.1213/ANE.0000000000004440] [Medline: 31567326]
7. Barrachina B, Cobos RFMN, Mardones NFCA, Castañeda A, Vinuesa C. Assessment of a smartphone app (Capstesia) for measuring pulse pressure variation: agreement between two methods: A Cross-sectional study. Eur J Anaesthesiol 2017 Feb;34(2):75-80. [doi: 10.1097/EJA.0000000000000569] [Medline: 28002069]
8. Carvalho H, Verdonck M, Berghmans J, Poelaert J. Development and validation of an android-based application for anaesthesia neuromuscular monitoring. J Clin Monit Comput 2019 Oct 16;33(5):863-870. [doi: 10.1007/s10877-018-0224-2] [Medline: 30446866]

9. Desebbe O, Joosten A, Suehiro K, Lahham S, Essiet M, Rinehart J, et al. A Novel Mobile Phone Application for Pulse Pressure Variation Monitoring Based on Feature Extraction Technology. Anesthesia & Analgesia 2016;123(1):105-113. [doi: 10.1213/ane.0000000000001282]
10. Joosten A, Jacobs A, Desebbe O, Vincent J, Sarah S, Rinehart J, et al. Monitoring of pulse pressure variation using a new smartphone application (Capstesia) versus stroke volume variation using an uncalibrated pulse wave analysis monitor: a clinical decision making study during major abdominal surgery. J Clin Monit Comput 2019 Oct 03;33(5):787-793. [doi: 10.1007/s10877-018-00241-4] [Medline: 30607806]
11. Joosten A, Boudart C, Vincent J, Vanden Eynden F, Barvais L, Van Obbergh L, et al. Ability of a New Smartphone Pulse Pressure Variation and Cardiac Output Application to Predict Fluid Responsiveness in Patients Undergoing Cardiac Surgery. Anesthesia & Analgesia 2019;128(6):1145-1151. [doi: 10.1213/ane.0000000000003652]
12. Chiu LYL, Sun T, Ree R, Dunsmuir D, Dotto A, Ansermino JM, et al. The evaluation of smartphone versions of the visual analogue scale and numeric rating scale as postoperative pain assessment tools: a prospective randomized trial. Can J Anaesth 2019 Jun 22;66(6):706-715. [doi: 10.1007/s12630-019-01324-9] [Medline: 30796700]
13. Dahlberg K, Philipsson A, Hagberg L, Jaensson M, Hälleberg-Nyman M, Nilsson U. Cost-effectiveness of a systematic e-assessed follow-up of postoperative recovery after day surgery: a multicentre randomized trial. Br J Anaesth 2017 Nov 01;119(5):1039-1046 [FREE Full text] [doi: 10.1093/bja/aex332] [Medline: 29077819]
14. Dahlberg K, Jaensson M, Nilsson U. "Let the patient decide" - Person-centered postoperative follow-up contacts, initiated via a phone app after day surgery: Secondary analysis of a randomized controlled trial. Int J Surg 2019 Jan;61:33-37 [FREE Full text] [doi: 10.1016/j.ijsu.2018.11.022] [Medline: 30503602]
15. Highland K, Tran J, Edwards H, Bedocs P, Suen J, Buckenmaier C. Feasibility of App-Based Postsurgical Assessment of Pain, Pain Impact, and Regional Anesthesia Effects: A Pilot Randomized Controlled Trial. Pain Med 2019 Aug 01;20(8):1592-1599. [doi: 10.1093/pm/pny288] [Medline: 30726985]
16. Jaensson M, Dahlberg K, Eriksson M, Nilsson U. Evaluation of postoperative recovery in day surgery patients using a mobile phone application: a multicentre randomized trial. Br J Anaesth 2017 Nov 01;119(5):1030-1038 [FREE Full text] [doi: 10.1093/bja/aex331] [Medline: 29077818]
17. Ke JXC, George RB, Wozney L, Chorney JL. Patient-centred perioperative mobile application in Cesarean delivery: needs assessment and development. Can J Anaesth 2019 Oct 13;66(10):1194-1201. [doi: 10.1007/s12630-019-01392-x] [Medline: 31087270]
18. Soh JY, Lee SU, Lee I, Yoon KS, Song C, Kim NH, et al. A Mobile Phone–Based Self-Monitoring Tool for Perioperative Gastric Cancer Patients With Incentive Spirometer: Randomized Controlled Trial. JMIR Mhealth Uhealth 2019 Feb 19;7(2):e12204. [doi: 10.2196/12204]
19. Sun T, West N, Ansermino JM, Montgomery CJ, Myers D, Dunsmuir D, et al. A smartphone version of the Faces Pain Scale-Revised and the Color Analog Scale for postoperative pain assessment in children. Paediatr Anaesth 2015 Dec 28;25(12):1264-1273. [doi: 10.1111/pan.12790] [Medline: 26507916]
20. Warren-Stomberg M, Jacobsson J, Brattwall M, Jildenstål P. At-home monitoring after surgery/anaesthesia - a challenge. J Eval Clin Pract 2016 Dec 01;22(6):882-886. [doi: 10.1111/jep.12551] [Medline: 27134050]
21. Baumann D, Dibbern N, Sehner S, Zöllner C, Reip W, Kubitz JC. Validation of a mobile app for reducing errors of administration of medications in an emergency. J Clin Monit Comput 2019 Jun 2;33(3):531-539. [doi: 10.1007/s10877-018-0187-3] [Medline: 30073445]
22. Görges M, West NC, Petersen CL, Ansermino JM. Development and Implementation of the Portable Operating Room Tracker App With Vital Signs Streaming Infrastructure: Operational Feasibility Study. JMIR Perioper Med 2019 Aug 05;2(2):e13559 [FREE Full text] [doi: 10.2196/13559] [Medline: 33393912]
23. Haffey F, Brady RRW, Maxwell S. A comparison of the reliability of smartphone apps for opioid conversion. Drug Saf 2013 Feb;36(2):111-117. [doi: 10.1007/s40264-013-0015-0] [Medline: 23322549]
24. Jabaley CS, Wolf FA, Lynde GC, O'Reilly-Shah VN. Crowdsourcing sugammadex adverse event rates using an in-app survey: feasibility assessment from an observational study. Ther Adv Drug Saf 2018 Jul;9(7):331-342 [FREE Full text] [doi: 10.1177/2042098618769565] [Medline: 30034775]
25. Lane JS, Sandberg WS, Rothman B. Development and implementation of an integrated mobile situational awareness iPhone application VigiVU™ at an academic medical center. Int J Comput Assist Radiol Surg 2012 Sep 7;7(5):721-735. [doi: 10.1007/s11548-012-0683-8] [Medline: 22484718]
26. Lelaidier R, Balança B, Boet S, Faure A, Lilot M, Lecomte F, et al. Use of a hand-held digital cognitive aid in simulated crises: the MAX randomized controlled trial. Br J Anaesth 2017 Nov 01;119(5):1015-1021 [FREE Full text] [doi: 10.1093/bja/aex256] [Medline: 29028930]
27. McEvoy M, Hand WR, Stiegler MP, DiLorenzo AN, Ehrenfeld JM, Moran KR, et al. A Smartphone-based Decision Support Tool Improves Test Performance Concerning Application of the Guidelines for Managing Regional Anesthesia in the Patient Receiving Antithrombotic or Thrombolytic Therapy. Anesthesiology 2016 Jan;124(1):186-198 [FREE Full text] [doi: 10.1097/ALN.0000000000000885] [Medline: 26513023]

28. Rothman BS, Dexter F, Epstein RH. Communication Latencies of Apple Push Notification Messages Relevant for Delivery of Time-Critical Information to Anesthesia Providers. Anesthesia & Analgesia 2013;117(2):398-404. [doi: 10.1213/ane.0b013e318299a7f7]
29. Rubin DS, Pesyna C, Jakubczyk S, Liao C, Tung A. Introduction of a Mobile Adverse Event Reporting System Is Associated With Participation in Adverse Event Reporting. Am J Med Qual 2019 Jun 25;34(1):30-35. [doi: 10.1177/1062860618781920] [Medline: 29938520]
30. Burstein B, Bretholz A. A Novel Smartphone App to Support Learning and Maintaining Competency With Bier Blocks for Pediatric Forearm Fracture Reductions: Protocol for a Mixed-Methods Study. JMIR Res Protoc 2018 Dec 21;7(12):e10363 [FREE Full text] [doi: 10.2196/10363] [Medline: 30578214]
31. De Oliveira GS, Glassenberg R, Chang R, Fitzgerald P, McCarthy RJ. Virtual airway simulation to improve dexterity among novices performing fibreoptic intubation. Anaesthesia 2013 Oct;68(10):1053-1058 [FREE Full text] [doi: 10.1111/anae.12379] [Medline: 23952805]
32. Linganna RE, Patel SJ, Ghofaily LA, Mackay EJ, Spelde AE, Zhou EY, et al. Pilot Study Suggests Smartphone Application Knowledge Improves Resident Transesophageal Echocardiography Knowledge: A Randomized Controlled Trial. J Cardiothorac Vasc Anesth 2020 Aug;34(8):2126-2132. [doi: 10.1053/j.jvca.2019.12.051] [Medline: 32035748]
33. Kain ZN, Vakharia S, Garson L, Engwall S, Schwarzkopf R, Gupta R, et al. The Perioperative Surgical Home as a Future Perioperative Practice Model. Anesthesia & Analgesia 2014;118(5):1126-1130. [doi: 10.1213/ane.0000000000000190]
34. Davidson AJ, Shrivastava PP, Jamsen K, Huang GH, Czarnecki C, Gibson MA, et al. Risk factors for anxiety at induction of anesthesia in children: a prospective cohort study. Paediatr Anaesth 2006 Sep;16(9):919-927. [doi: 10.1111/j.1460-9592.2006.01904.x] [Medline: 16918652]
35. Herrera-Usagre M, Santana V, Burgos-Pol R, Oliva JP, Sabater E, Rita-Acosta M, et al. Effect of a Mobile App on Preoperative Patient Preparation for Major Ambulatory Surgery: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2019 Jan 16;8(1):e10938 [FREE Full text] [doi: 10.2196/10938] [Medline: 30664480]
36. Naguib M, Kopman AF, Lien CA, Hunter JM, Lopez A, Brull SJ. A Survey of Current Management of Neuromuscular Block in the United States and Europe. Anesthesia & Analgesia 2010;111(1):110-119. [doi: 10.1213/ane.0b013e3181c07428]
37. Söderström CM, Eskildsen KZ, Gätke MR, Staehr-Rye AK. Objective neuromuscular monitoring of neuromuscular blockade in Denmark: an online-based survey of current practice. Acta Anaesthesiol Scand 2017 Jul 02;61(6):619-626. [doi: 10.1111/aas.12907] [Medline: 28573656]
38. Bulka C, Terekhov MA, Martin B, Dmochowski R, Hayes R, Ehrenfeld J. Nondepolarizing Neuromuscular Blocking Agents, Reversal, and Risk of Postoperative Pneumonia. Anesthesiology 2016 Oct;125(4):647-655 [FREE Full text] [doi: 10.1097/ALN.0000000000001279] [Medline: 27496656]
39. Weldring T, Smith SMS. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). Health Serv Insights 2013;6:61-68 [FREE Full text] [doi: 10.4137/HSI.S11093] [Medline: 25114561]
40. Shipton EA. The Transition from Acute to Chronic Post Surgical Pain. Anaesth Intensive Care 2011 Sep 01;39(5):824-836. [doi: 10.1177/0310057x1103900506]
41. Ozdalga E, Ozdalga A, Ahuja N. The smartphone in medicine: a review of current and potential use among physicians and students. J Med Internet Res 2012 Sep 27;14(5):e128 [FREE Full text] [doi: 10.2196/jmir.1994] [Medline: 23017375]
42. Nabors C, Peterson SJ, Aronow W, Sule S, Mumtaz A, Shah T, et al. Mobile physician reporting of clinically significant events-a novel way to improve handoff communication and supervision of resident on call activities. J Patient Saf 2014 Dec;10(4):211-217. [doi: 10.1097/PTS.0b013e31829952ff] [Medline: 25408999]
43. McEvoy MD, Field LC, Moore HE, Smalley JC, Nietert PJ, Scarbrough SH. The effect of adherence to ACLS protocols on survival of event in the setting of in-hospital cardiac arrest. Resuscitation 2014 Jan;85(1):82-87 [FREE Full text] [doi: 10.1016/j.resuscitation.2013.09.019] [Medline: 24103233]
44. Affronti ML, Schneider SM, Herndon JE, Schlundt S, Friedman HS. Adherence to antiemetic guidelines in patients with malignant glioma: a quality improvement project to translate evidence into practice. Support Care Cancer 2014 Jul 26;22(7):1897-1905 [FREE Full text] [doi: 10.1007/s00520-014-2136-0] [Medline: 24570103]
45. Mearis M, Shega JW, Knoebel RW. Does Adherence to National Comprehensive Cancer Network Guidelines Improve Pain-Related Outcomes? An Evaluation of Inpatient Cancer Pain Management at an Academic Medical Center. Journal of Pain and Symptom Management 2014 Sep;48(3):451-458. [doi: 10.1016/j.jpainsymman.2013.09.016]
46. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]

Abbreviations

6MWT: 6 Minute Walk Test
ASRA: American Society of Regional Anesthesia
CAS: color analog scale
CO: cardiac output

DASI: Duke Activity Status Index
FPS-R: faces pain scale-revised
IS: incentive spirometry
ICC: intraclass correlation coefficients
mHealth: mobile health
mYPAS: modified Yale Preoperative Anxiety Scale
NMB: neuromuscular blockade
OR: operating room
PPV: pulse pressure variation
PROM: patient-reported outcome measures
QoL: quality of life
RAPP: recovery assessment by phone points
RCT: randomized controlled trial

Edited by R Kukafka; submitted 22.10.20; peer-reviewed by AS Poncette, C Reis; comments to author 08.12.20; revised version received 28.12.20; accepted 14.06.21; published 17.09.21.

Please cite as:
Pan S, Rong LQ
Mobile Applications in Clinical and Perioperative Care for Anesthesia: Narrative Review
J Med Internet Res 2021;23(9):e25115
URL: https://www.jmir.org/2021/9/e25115
doi:10.2196/25115
PMID:34533468

Review

How Self-tracking and the Quantified Self Promote Health and Well-being: Systematic Review

Shan Feng[1], MSc; Matti Mäntymäki[1], PhD; Amandeep Dhir[2], PhD; Hannu Salmela[1], PhD

[1]Department of Management and Entrepreneurship, Turku School of Economics, University of Turku, Turku, Finland

[2]Department of Management, School of Business and Law, University of Agder, Kristiansand, Norway

Corresponding Author:
Matti Mäntymäki, PhD
Department of Management and Entrepreneurship, Turku School of Economics
University of Turku
Rehtorinpellonkatu 3
Turku, 20500
Finland
Phone: 358 504867657
Email: matti.mantymaki@utu.fi

Abstract

Background: Self-tracking technologies are widely used in people's daily lives and health care. Academic research on self-tracking and the quantified self has also accumulated rapidly in recent years. Surprisingly, there is a paucity of research that reviews, classifies, and synthesizes the state of the art with respect to self-tracking and the quantified self.

Objective: Our objective was to identify the state of the art of self-tracking and the quantified self in terms of health and well-being.

Methods: We have undertaken a systematic literature review on self-tracking and the quantified self in promoting health and well-being. After a rigorous literature search, followed by inclusions, exclusions, and the application of article quality assessment protocols, 67 empirical studies qualified for the review.

Results: Our results demonstrate that prior research has focused on 3 stakeholders with respect to self-tracking and the quantified self, namely end users, patients and people with illnesses, and health care professionals and caregivers. We used these stakeholder groups to cluster the research themes of the reviewed studies. We identified 11 research themes. There are 6 themes under the end-user cluster: user motivation and goal setting, usage and effects of self-tracking, continuance intention and long-term usage, management of personal data, rejection and discontinuance, and user characteristics. The patient and people with illnesses cluster contains three themes: usage experience of patients and people with illnesses, management of patient-generated data, and advantages and disadvantages in the clinical context. The health care professional and caregiver cluster contains two themes: collaboration among patients, health care professionals, and caregivers, and changes in the roles of patients and professionals. Moreover, we classified the future research suggestions given in the literature into 5 directions in terms of research designs and research topics. Finally, based on our reflections on the observations from the review, we suggest four future research directions: (1) users' cognitions and emotions related to processing and interpreting the information produced by tracking devices and apps; (2) the dark side of self-tracking (eg, its adverse psychosocial consequences); (3) self-tracking as a societal phenomenon; and (4) systemic impacts of self-tracking on health care and the actors involved.

Conclusions: This systematic literature review contributes to research and practice by assisting future research activities and providing practitioners with a concise overview of the state of the art of self-tracking and the quantified self.

(J Med Internet Res 2021;23(9):e25171) doi:10.2196/25171

KEYWORDS

self-tracking; quantified self; health; well-being; systematic literature review; literature review

Introduction

Self-tracking—also referred to as the quantified self, self-monitoring, lifelogging, personal analytics, and personal informatics [1]—has become pervasive in people's daily lives and increasingly common in health care. For example, health and fitness ranked ninth out of 34 categories on the App Store, accounting for 3.41% of all the available apps [2,3]. In Google Play, health and fitness apps ranked 13th out of 33, accounting for 3.53% of all the available apps [4]. Moreover, the App Store hosted 48,608 apps for health care purposes, and Google Play 47,140 in the third quarter during 2020 [5,6]. Although global smartwatch shipments have been estimated to exceed 100 million units in 2020 and continue to grow in 2021 [7], self-tracking tools have become available for increasing numbers of people across the globe.

Research on self-tracking and the quantified self has proliferated rapidly in recent years [8-14]. This growth calls for review studies that help enrich the knowledge base by classifying and synthesizing prior research and providing directions for future research activities [15,16]. However, dedicated review studies are scant in the most recent literature on self-tracking and the quantified self. Altogether, the extant literature features 8 review studies [17-24], of which 5 have been published in journals [17-19,22,24] and 3 in conference proceedings [20,21,23].

Moreover, some of the prior review studies have not focused exclusively on self-tracking or the quantified self. For example, Paton et al [18] focused on wearable self-tracking devices, social media platforms used by the self-tracking community, and personal health records. West et al [19] conducted a review of self-tracking in the context of patient-generated data, whereas Lentferink et al [24] identified the key components of self-tracking and persuasive eCoaching in automated healthy lifestyle interventions to improve the design of these interventions. Therefore, a review focusing only on the quantified self and self-tracking is needed.

Third, with respect to the scope, prior review studies have typically focused on a certain subset of the literature. For example, Moya et al [20] focused on the adoption and utilization of self-tracking technologies. Almalki et al [17] limited their scope to studies that applied activity theory for health self-quantification. Jiang and Cameron [22] only focused on self-monitoring in the context of chronic disease. Furthermore, most of the extant reviews covered a relatively small number of studies. The number of studies covered by the prior reviews are 43 [21], 32 [24], 28 [23], 26 [17], and 23 [19] respectively, and one of the prior reviews did not clearly mention the number of reviewed studies. Considering these limitations in terms of the scope of prior reviews and the increasing research volume on self-tracking and the quantified self, there is a need for an up-to-date and comprehensive literature review.

To address this gap in the literature, we undertook a systematic literature review (SLR) of the research on self-tracking and the quantified self in terms of health and well-being. The purpose of this study can be summarized as follows: (1) to identify high-quality contributions in the area, (2) to classify the literature based on publication channels, research methods, theoretical backgrounds, and variables used, (3) to synthesize the main research areas and research themes, and (4) to identify future research directions. The final pool of the reviewed articles comprised 67 empirical studies.

We mapped the research subjects of the reviewed articles and identified three main stakeholder groups: end users, patients and people with illnesses, and health care professionals and caregivers. For the end users, the primary usage of self-tracking is for tracking exercise, daily activity levels, and sleep. For the two other stakeholder groups (ie, patients and people with illnesses as well as health care professionals and caregivers), the primary usage of self-tracking is to support the treatment of an illness or other medical conditions. We further used these stakeholder groups for clustering the research themes of the reviewed studies. By mapping the research themes, examining the focal constructs of the prior studies, synthesizing the future research directions, and finally reflecting on the findings and suggesting new research areas, this study aids in building future research efforts [15,16].

The remainder of the paper is structured as follows. The background and related works are discussed next followed by the research methodology and details of the article selection procedure. Then, we report the results of the analysis. Finally, the implications, limitations, and some directions for future research are discussed. Multimedia Appendix 1 presents the article search strategy. Multimedia Appendix 2 contains the quality assessment criteria used for selecting the articles to be included in the review. Multimedia Appendices 3, 4, and 5 contain detailed information about the reviewed articles. Multimedia Appendix 6 presents a summary of the reviewed studies based on their foci and research theme matrices.

Background and Related Works

Self-tracking and the Quantified Self

Self-tracking and the quantified self are not new labels. As early as the 1970s, wearable computers for self-tracking were used as forms of personal surveillance [25]. In 2001, a small number of media practitioners began to use newly available digital technology to track their daily life for designing web 1.0 interfaces [26]. According to Lupton [1], self-tracking "involves practices in which people knowingly and purposively collect information about themselves, which they then review and consider applying in their lives."

As further pointed out by Lupton [1], in addition to self-tracking, there are several other terms—such as lifelogging, personal informatics, and the quantified self—used to describe the practices by which people may seek to monitor their everyday life. Lifelogging can be viewed as the practice of recording information about one's life using digital tools. Personal informatics is a term used mostly in the academic human-computer interaction community [1]. The concept of the "quantified self," originally coined by Wolf and Kelly in 2007, refers to "self-knowledge through numbers" [1]. They used quantitative data as a means or an embodiment of monitoring the elements of everyday life. The term can be further viewed as a collaboration between users and toolmakers [27], a cultural phenomenon involving technology [8], and an

outcome of self-tracking [28]. In fact, the original intention of Wolf and Kelly was to use the term "the quantified self" to describe this digital self-tracking phenomenon [1].

Health and Well-being

According to the World Health Organization, health is explicitly linked with well-being, which is "a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity" [29]. With the development and diffusion of wearable, wireless communication, and cloud computing technologies, self-tracking devices can not only provide data, but also refine them into key performance indicators and produce visualizations of these data. People use self-tracking technologies to self-collect personal data including biological, physical, behavioral, or environmental information [30,31]. Moreover, according to Lupton (2016), the domain of self-tracking today also includes relationships and work productivity. In the scope of self-tracking or the quantified self, health and well-being are the main tracking domains. People who knowingly and purposively track their health and well-being information (such as heart rate, sleep, physical activity, calories, clinical symptoms, stress, and recovery) to review and modify their lives widely use self-tracking devices.

Related Works

Self-tracking and the quantified self have been examined in various disciplines and from various theoretical and methodological premises. For example, the computer science research field has examined wearable augmented reality systems based on walking locomotion analysis [32] and developed algorithms for monitoring sleep [33]. The relevant literature also indicates certain design considerations, such as presenting negative data in a way that does not demotivate users [34]. Communication research has discussed topics such as self-tracking as a communicative phenomenon with social media, the self, and social networks of peers [35]. Furthermore, sociological studies have discussed ways of attributing meaning to data-gathering practices in terms of the quantified self [36]. Prior literature also includes research focusing on the societal and ethical concerns regarding self-tracking [37], including the value of personalized health care. In medicine, prior studies have explored the clinical experience of self-tracking technologies in the context of chronic diseases [38,39] and use of self-tracking devices in rehabilitation [40,41]. Finally, research on information systems has examined themes such as user acceptance of self-tracking, user motivation, and goal attainment related to self-tracking and the quantified self [42-44].

To keep the scope of the study manageable, we deliberately focused on self-tracking and the quantified self, executed via devices, apps, and platforms, and excluded digital diaries and video recordings to focus solely on the role of self-tracking in health and well-being. Moreover, we focused on empirical research where the primary research subjects are humans. Thus, purely technical papers and articles focusing on products, services, and markets were excluded.

Methods

Research Questions

The purpose of the SLR is to determine the current state of research on self-tracking and the quantified self in the domain of health and well-being. To this end, we address 6 specific research questions (RQs) that are presented in Table 1.

Table 1. Research questions.

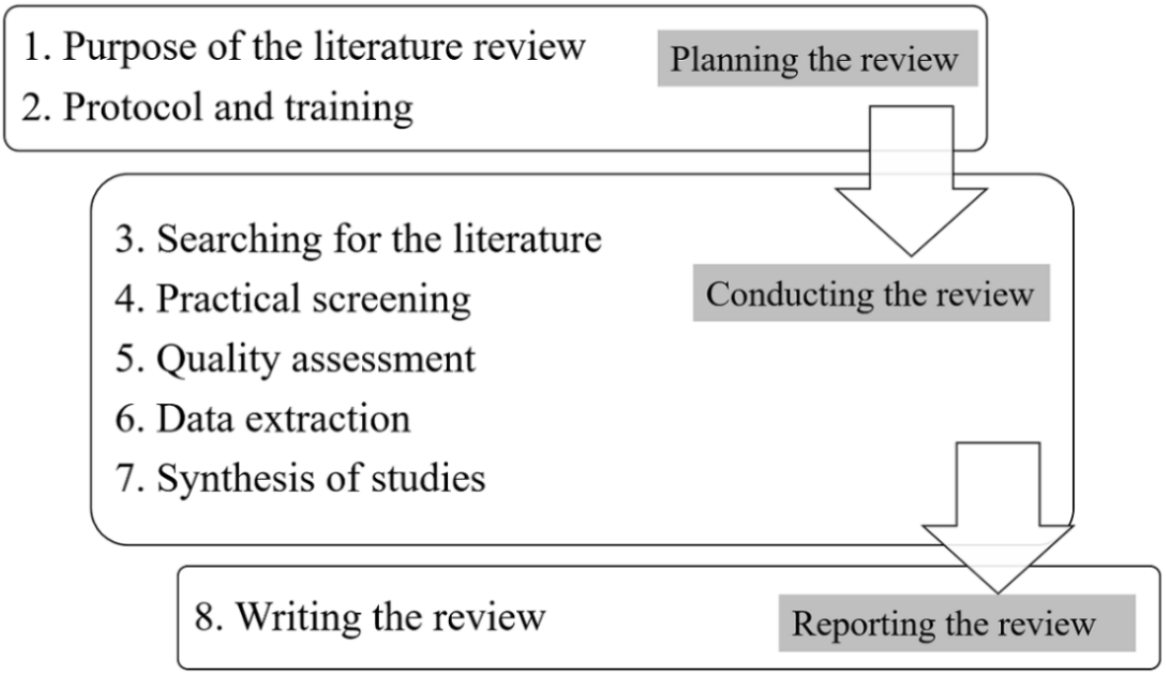
Number	Research question
RQ1[a]	How has the volume of publications on self-tracking and the quantified self in the domain of health and well-being evolved?
RQ2	What are the most important publication channels?
RQ3	What research methods have been used?
RQ4	What theoretical backgrounds and variables have been employed?
RQ5	What recurring research themes can be identified from the literature?
RQ6	What future research directions can be synthesized from the literature?

[a]RQ: research question.

Research Design

SLRs have been developed as a specific method for identifying and synthesizing research findings [45]. They are considered particularly useful to disseminate the key findings of large and complex bodies of research. SLRs employ a transparent and rigorous approach (review protocol) to identify and synthesize all available research findings of sufficient quality concerning a specific research question or subject [46]. According to Victor [47], an SLR differs from a traditional literature review owing to its specific emphasis on the following features: (1) comprehensive coverage of the literature as far as possible; (2) paying careful attention to the quality of the included evidence; (3) taking a clear, systematic approach to data synthesis, and (4) generally following transparent and rigorous processes. In this research, we followed a well-established 8-step review protocol [48]. Figure 1 summarizes the review process.

Figure 1. Review process.



Database and Search Syntax

Articles were obtained from five academic databases: Scopus, Web of Science, PubMed, Taylor & Francis, and the Association for Information Systems eLibrary. The search strategy included keywords such as "self-track*," "self track*," "quantified self," "quantified-self," "self quantif*," and "self-quantif*." We limited the search to journal and conference papers published in English. After conducting the literature search in the 5 databases, we eliminated duplicates. The details of the search strategy are shown in Multimedia Appendix 1.

Inclusion and Exclusion Criteria

Two authors (SF and MM) conducted the inclusion/exclusion procedure, prepared notes, and exchanged information throughout the process. Citation chaining was conducted to further reinforce the comprehensiveness of the article search. Textbox 1 presents the inclusion and exclusion criteria.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria (IC)
IC1: Studies with the main topic (title/abstract/keywords) containing self-tracking or the quantified self
IC2: Studies published in a peer-reviewed journal or scientific conference
IC3: Studies published in English
IC4: Studies with title/keywords/abstract-based screening
IC5: Studies with full-text–based screening
IC6: Studies with humans as primary empirical research subjects
IC7: Studies based on citation chaining
Exclusion criteria (EC)
EC1: Studies with matching title and digital object identifiers
EC2: Studies whose main topic is not self-tracking or the quantified self (digitized self-tracking and self-quantification via mobile devices, apps, and platforms; ignore diary and video recordings)
EC3: Studies where the purpose of the paper is not improving health and well-being
EC4: Studies that are not empirical studies (literature reviews, conceptual papers, technical papers, or editorials)
EC5: Studies in which the primary empirical subjects are not human, such as products, services, or markets
EC6: Studies based on quality assessment

Quality Assessment

Quality assessment is a step in the SLR process conducted to ensure that the results of the review are suitable and impartial by identifying the articles that are not of sufficient quality to be included in the sample pool in an objective and replicable manner [48]. The quality assessment protocol incorporates the research method, structuring, implications (contributions or implications), and limitations, and ranking of the publication channel. The maximum total score from the assessment was 8.5. The criteria for quality assessment are presented in Table 2. To further ensure the high quality of the articles included in the review, we followed the approach proposed by Idri et al [49] and Behera et al [50] and set 50% of the maximum quality score as the threshold for a paper to be included in the in-depth review. Hence, 17 articles that received a total score of less than half of the maximum score (ie, 4.25) were omitted from the review. The quality score of each article can be found in Multimedia Appendix 2.

Table 2. Quality assessment criteria.

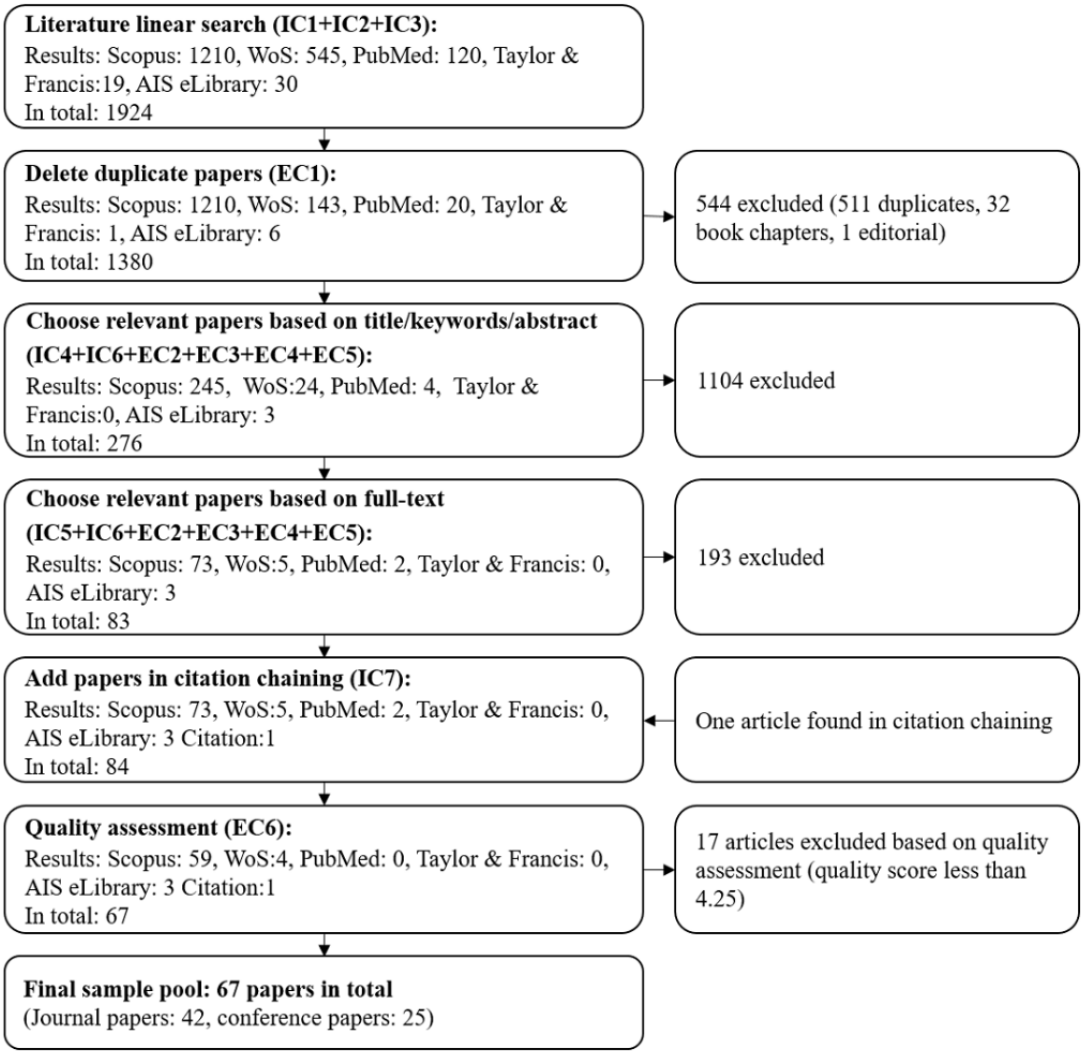
Criterion	Description	References
QA1[a]	The empirical study adopts a qualitative, quantitative, or mixed method approach. The possible answers are mixed method (+2), and quantitative or qualitative research (+1).	[50]
QA2	The study is a fully structured article divided into four basic sections: introduction, methods, results, and discussion. The answers are Yes (+1) and No (+0).	[51]
QA3	The study unequivocally describes the research process in sufficient detail. A quantitative study shows the questionnaire items, a qualitative one the coding and categorization process, and an experimental one the details of the experiment. The answers are Yes (+2), Partially (+1), and No (+0).	[17]
QA4	The study clearly documents the research implications (contributions or implications) and limitations. The answers are Yes (+2), Partially (+1), and No (+0).	[50]
QA5	The study was published in a reliable and recognized publication journal.	[49,52]
	Based on Journal Citation Reports (JCR: an annual publication that provides information about academic journals with impact factor data): journal in the top 25% (Q1 +1.5), in the 25%–50% group (Q2 +1), in the 50%–100% group (Q3 or Q4 +0.5), and no JCR ranking (+0).	
	Conference ranking based on CORE (conference ranking portal): CORE A* or A (+1.5), CORE B (+1), CORE C (+0.5), and no CORE ranking (+0).	

[a]QA: quality assessment.

Pool of Articles Included in the Review

Finally, 67 articles were selected as the final sample, of which 42 were journal articles and 25 articles were published in conference proceedings. Figure 2 presents the pool of articles in each stage of the inclusion and exclusion procedure.

Figure 2. Inclusion and exclusion procedure. AIS: Association for Information Systems; IC: inclusion criterion; EC: exclusion criterion; WoS: Web of Science.



Results

RQs 1 and 2: Publications by Year and Channel

In the pool of articles included in the review, the first was published in 2013. As shown in Figure 3, the number of papers published per year increased from 2013 until 2019. Table 3 presents the most popular publication channels. A considerable proportion of the conference papers appeared in information systems conferences, whereas the journal articles were published mostly in outlets focusing on health.

Figure 3. Articles included in the review by year.

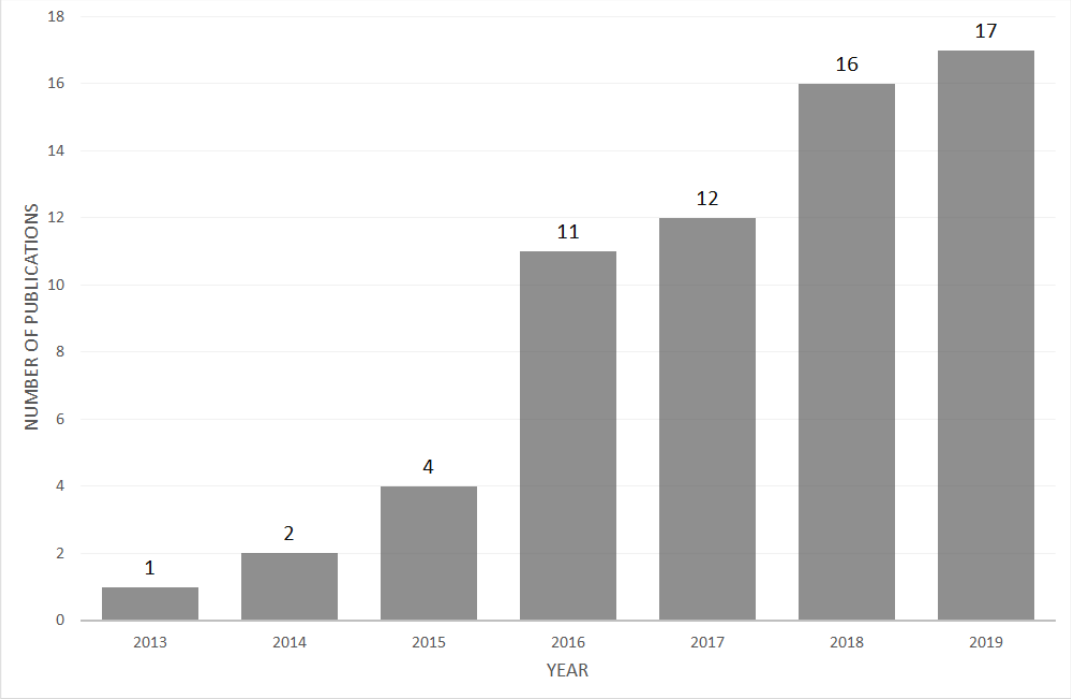


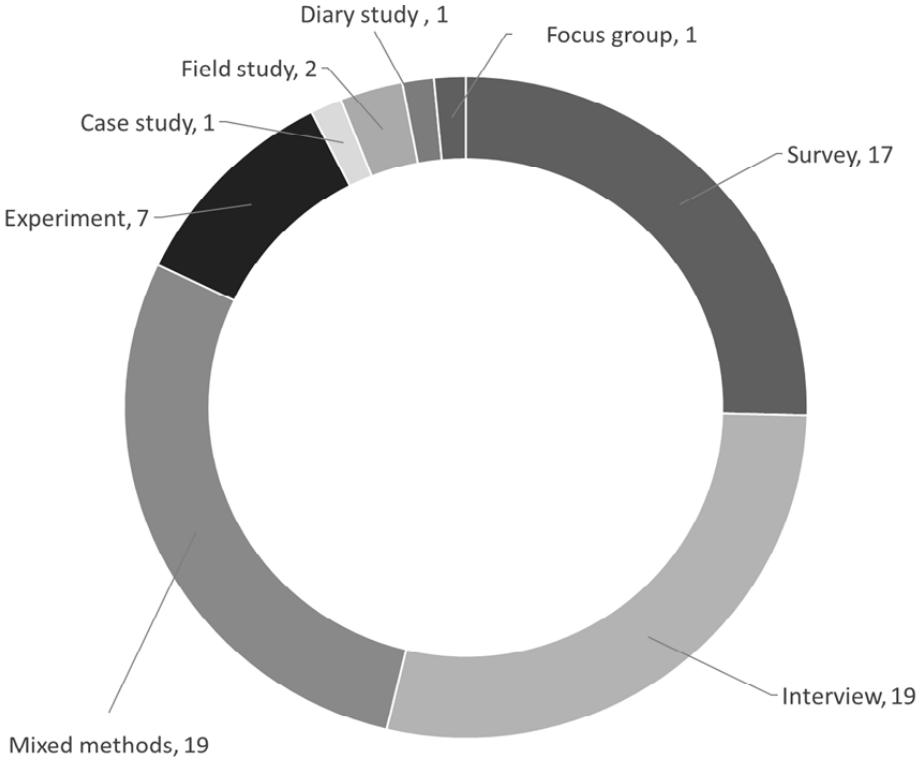
Table 3. Most common publication channels (number of publications greater than 1).

Conference/journal	Number of papers, n (%) (N=67)
Conference	
Conference on Human Factors in Computing Systems	6 (8.96)
Americas Conference on Information Systems	3 (4.48)
Pacific Asia Conference on Information Systems	2 (2.99)
International Conference on Information Systems	2 (2.99)
Other conferences	12 (17.91)
Total	25 (37.31)
Journal	
Computers in Human Behavior	5 (7.46)
JMIR mHealth and uHealth	5 (7.46)
Journal of Medical Internet Research	4 (5.97)
Digital Health	2 (2.99)
Health Sociology Review	2 (2.99)
International Journal of Human-Computer Studies	2 (2.99)
Sociology of Health and Illness	2 (2.99)
Telemedicine and e-Health	2 (2.99)
Other journals	18 (26.87)
Total	42 (62.69)

RQ 3: Distribution of Articles by Study Method

We categorized the pool of studies based on the research methods (see Figure 4). For this purpose, we examined how the empirical data were collected. The 2 most commonly used methodologies were mixed-methods research and interviews. In the mixed-methods research category, combining surveys with interviews was the most frequently used investigative approach. Most of the experimental research was conducted in a longitudinal fashion. Multimedia Appendices 3, 4, and 5 provide more information about the reviewed studies, including theories, methods, and focal constructs.

Figure 4. Pool of articles reviewed by research methods (total number of articles=67).



RQ 4: Theoretical Backgrounds and Variables on Self-tracking and the Quantified Self

In our pool of 67 studies, 39 articles explicitly stated the theoretical foundation on which their studies were based. The remaining articles typically discussed prior literature but did not explicitly build on any theoretical foundation. This includes articles using the grounded theory methodology to build or contribute to theory [38,53-55]. In the aforementioned 39 articles, 18 studies used theories to develop hypotheses and research models to be tested with quantitative methods, whereas in 21 studies, the primary role of the theory was to support qualitative research.

Among the 18 quantitative studies with explicitly stated theoretical backgrounds, the most widely used theoretical lenses were the technology acceptance model (TAM) and self-determination theory. Understandably, these theoretical backgrounds were common among studies focusing on self-tracking behaviors, motivations, goals, and use continuance. Table 4 provides information about the theoretical backgrounds and dependent variables of the studies where the theoretical foundation was explicitly stated. Multimedia Appendices 3, 4, and 5 provide more details about the reviewed articles [56-68].

Table 4. Theoretical backgrounds and dependent variables.

Dependent variable categories	Theory/theories	References
Self-tracking behavior/technology usage		
Habitual web-based fitness community use	Self-determination theory, technology acceptance model	[68]
Number of activities	Social cognitive theory	[56]
Behavior (activity index + sleep index + food index)	Health information technology acceptance model	[57]
Behavior (compliance of self-tracking behavior)	Health information technology acceptance model	[58]
Behavioral intention to use (wearable devices in the workplace)	Self-determination theory, technology acceptance model	[59]
Motivation/goal to use self-tracking technologies		
Classes of motivational designs (gamification, social networking, and self-quantification)	Self-determination theory, social comparison theory, goal-setting theory	[60]
Self-tracking motivation fulfillment	Gimpel five-factor framework of self-tracking motivations	[11]
Continuance intention of self-tracking		
Continuance intention (of using smartwatches)	Net valence framework	[61]
Continuance intention (of using quantified-self technology)	Motivational affordance theory, technology continuance theory	[62]
Continuance intention (of the quantified self)	Trans-theoretical model, expectation-confirmation theory	[63]
Cognitive dissonance when using self-tracking technologies		
Awareness of two inconsistent cognitions, self-tracking usage, and cognitive dissonance	Cognitive dissonance theory	[64-66]
Self-disclosure of personal information		
Self-disclosure (personal information in self-tracking apps)	Stage-based model, five-factor model	[28]
Dependency effect on wearing activity trackers		
Dependency effect (on wearing activity trackers)	Self-determination theory	[10]
Health consciousness, physical and psychological well-being		
Health consciousness, physical health, and psychological well-being	PERMA (positive emotion, engagement, relationships, meaning, accomplishment) model	[67]
Sharing health-tracking records		
Sharing health-tracking records	Health belief model	[27]

In addition to the dependent variables presented in Table 4, we also investigated the independent, mediating, and moderating variables employed in the reviewed studies. Multimedia Appendices 3, 4, and 5 provide more details about the reviewed studies. As depicted in Table 4, the most common dependent variables were self-tracking behavior and device usage, followed by motivation and goal for using self-tracking technologies, and continuance intention of self-tracking. Independent variables in the reviewed studies were normally extracted from theories; for example, perceived ease of use and usefulness were extracted from TAM [57,59,68], 3 strategies for reducing cognitive dissonance from the cognitive dissonance theory [64-66], 5 personality traits from the big-five personality domains [28], and 5-factor motivation (self-entertainment, self-association, self-design, self-discipline, and self-healing) from the Gimpel five-factor framework of self-tracking motivations [11]. Some studies combined theory with self-tracking practices and classified independent variables into more specific categories; for example, Chuah [61] divided perceived benefits into four perspectives, including utilitarian, hedonic, social, and symbolic, and divided perceived risks into privacy-related risks and physical risks. Suh [62] specifically subdivided motivation into hedonic, utilitarian, and eudaemonic (refers to self-fulfillment and self-improvement) motivations. In addition, Hamari [60] specifically described three independent variables: goal focus (outcome and process), goal orientation (proving, avoidance, and mastery), and goal attributes (difficulty and specificity).

In addition to the aforementioned dependent variables, the reviewed literature has covered various other variables. For example, perceived empowerment and personalization had mediating effects between self-tracking technology and advice compliance [69]. Intrinsic motivation was identified a mediator between need for cognitive closure and the dependency effect on self-tracking technologies [10]. When using smartwatches, inspiration and well-being played a mediating role between the perceived benefits and risks, and continuance intention [61]. Perceived threat, perceived usefulness, perceived ease of use mediated consumers' health statuses, beliefs and concerns, subjective norms, and self-efficacy to attitude [57]. Finally, prior research has also explored the moderating effects of demographics such as gender, education, income [70], and BMI [69].

RQ 5: Themes in the Existing Literature

To structure the literature analysis, we developed a classification scheme. To this end, we applied the concept matrix approach presented by Webster and Watson [71]. A concept matrix represents a logical method of defining a set of "concepts" according to which the reviewed articles can be classified.

We first identified the focus of each reviewed article based on its title, abstract, research question/objective, and results. Thereafter, we classified the articles based on the subjects of the empirical research. This led to the formation of three stakeholder groups, namely end users, patients and people with illnesses, and health care professionals and caregivers. The studies adopting the end-user perspective focused on self-tracking related to personal health, well-being, fitness, or sleep. The second, (ie, the patients' perspectives) focused on self-tracking use usage cases related to treatment of illnesses. The third, (ie, the health care professional and caregiver perspectives) focused on treatment of illnesses and medical conditions, but the primary users are the patients' doctors, nurses, or family members. Consequently, we used the perspectives of these stakeholder groups to establish a concept matrix. Finally, we classified the studies into 11 themes that will be discussed in the following section. Multimedia Appendix 6 presents a summary of the reviewed studies based on their foci and research theme matrices.

Stakeholder Group 1: End Users

(1) User Motivation and Goal Setting

With respect to the studies that adopted the end users' perspectives, user motivation constitutes an important stream of research. Most of the research examined the motivation for self-tracking usage [10-12,62,72]. In their research, Attig and Franke [10] divided the motivation for tracking physical activity into intrinsic and extrinsic motivations. The results revealed that people with strong extrinsic motivation may be highly dependent on the use of the tracking devices. Suh [62] classified motivation into hedonic, utilitarian, and eudaemonic. The results illustrated that unlike hedonic motivation, utilitarian and eudaemonic motivations positively impact continuance intention use of self-tracking technologies. Further, Pingo and Narayan [12] suggested that users may experience gratification when they can use tracking devices to control their health status. To clarify the reason for the usage of tracking devices, Gimpel et al [44] employed a survey to develop a five-factor framework of self-tracking motivations. The factors were self-entertainment, self-association, self-design, self-discipline, and self-healing. Gimpel et al [11] also illustrated that only the motivation for self-entertainment increases the wearable self-tracking device usage, which in turn influences the fulfillment of the self-entertainment motivation. Baudier et al [59] also emphasized that self-entertainment and self-design have a positive effect on the behavioral intention to use the health care Internet of Things in the workplace.

In prior research, goal setting and motivation always appeared at the same time. Gordon et al [13] found that users of weight loss–tracking apps preferred to choose achieving goals as their motivation. Self-monitoring motivation and attitude toward weight loss goals are vital for predicting goal achievement. For fitness tracking apps and exercise encouragement apps, Rockmann and Gewald [14], and Hamari et al [60] noted individual differences and found that different motivational affordances, such as self-quantification, gamification, and social networking capabilities, help users achieve their goals.

(2) Usage and Effects of Self-tracking

The second cluster under the end users' perspective focused on the different uses of self-tracking and the outcomes of usage. With respect to the different uses and use cases, prior research has explored the use of self-tracking for physical activities [8,9,35,57,58,73-77], sleep [57,58,74,78], diet [57,58], fertility [55], caloric intake [79], and alcohol-harm reduction [80].

Prior literature also provides classifications of users and self-tracking tools. Makkonen et al [81] identified 4 distinct consumer segments of self-tracking based on their technology adoption patterns: pro-trackers, semitrackers, interested trackers, and nontrackers. In the context of app usage for alcohol consumption reduction, Milward et al [80] identified three types of users: the trackers, cut-downers, and non-committers. Spotswood et al [9] explored the role of self-tracking in supporting healthy behavior and found three mechanisms for teleoaffective shaping: labeling, rewarding, and materializing effort. Lyall and Robards [82] identified three roles of self-tracking devices for the user, namely as a tool, toy, and tutor.

Moreover, previous research has explored the outcomes of self-tracking. Shin and Biocca [63] noted that health consciousness is the most significant factor for staying healthy. Stiglbauer, Weber, and Batinic [67] proved that wearing a fitness tracking device can make users more conscious regarding their physical health but did not have a significant effect on their mental health. Furthermore, according to Ravichandran [78], sleep-tracking devices can help users better understand and improve their overall sleep habits. Rönkkö [74] proposed that certain design features such as graphical feedback, information sharing, and social communities in self-tracking devices may be particularly important in contributing toward positive lifestyle changes, with the most important factor being the presence of personal long-term goals.

(3) Continuance Intention and Long-term Usage

The third cluster of studies from the end users' perspectives focused on sustained usage of self-tracking and the quantified self. For example, Chuah [61] found that the perceived benefits (utilitarian, hedonic, social, and symbolic) indirectly affect users' continuance intentions regarding smartwatch use through inspiration and well-being. Shin and Biocca [63] investigated the relative effects of hedonic and utilitarian motives as determinants of confirmation, satisfaction, and continuance intention regarding wearable devices. In contrast, Suh [62] observed that hedonic motivation has a negative influence on the continuance intention when using self-tracking devices, but utilitarian and eudaemonic motivations have a positive effect. As for web-based fitness community usage, Stragier et al [68] proposed that self-regulatory and social motives directly predict sustained web-based fitness community usage. Rockmann et al

[83] also theorized an emotional "carry-over effect" in activity continuance decisions on activity tracking.

With respect to the long-term use of self-tracking and the quantified self, Hardey [84] employed a mixed-methods approach and found that visualization and a long-term healthy state motivate and support long-term tracking. In addition, Meyer et al [85] built 5 use cases for personal health devices that can be used as a long- and short-term comparison list. The five use cases are supporting health behavior, improved self-understanding, identification of trends and relations, decision-making, and data collection for future use.

(4) Management of Personal Data

The fourth cluster of studies from the end users' perspectives emphasized managing personal data. This comprises two subtopics: coping with personal data and privacy concerns. Feng et al [86] showed that people preferred to use health or fitness apps as their personal health information management tools, and the most popular app was Fitbit. When reflecting on the personal data stored on self-tracking devices, Maltseva and Lutz [28] proposed that individuals who habitually use self-tracking apps and devices are more likely to self-disclose their personal data. When self-tracking devices are to social networks, the proportion of lower-performing friends may positively influence users' physical activities [56]. Moreover, Yli-Kauhaluoma and Pantzar [87] examined the gap between individual experiences and self-tracking data. They found that individuals always feel upset and confused when comparing invisible or inaccurate personal health data with their daily experiences and may eventually refuse to use self-tracking devices. In another words, individuals may experience an emotional reaction, which can then stop changing their behavior [43].

Literature has also documented problems related to tracking, managing, visualizing, and using personal data [88]. Privacy issues are obviously a central concern related to sharing self-tracking data. Chen et al [89] examined the sharing of health data among college students. Their results implied that users are generally willing to share personal health data for research purposes, and the reasons for not sharing are related to privacy concerns. Gui et al [53] pointed out that when fitness devices are connected to social networks, users are encouraged to share fitness data. However, the challenges of balancing awareness and privacy issues were still prevalent. To solve this problem, Zimmer [90] employed the communication privacy management theory to better understand the privacy boundaries related to personal fitness information and found that the advantages outweigh the disadvantages.

(5) Rejection and Discontinuance

The fifth cluster of studies from the end users' perspectives relates to rejection and discontinuance. In their study of experiences during the implementation of self-tracking technology, Kari et al [42] found that the experience during the initial phase of implementation will influence the decision on adoption or rejection. Moreover, previous studies have found that the abandonment of wearable fitness tracking devices can be owing to a loss of motivation, low acceptance levels for such devices, or inaccuracy and uselessness [91]. Harrison et al [92] cited tracking accuracy and device aesthetics as barriers to activity tracking. Esmonde [93] identified the following 4 strategies of resistance used by people to datafication in fitness tracking practices: excessive labeling of some forms of data rather than tracking everything, choosing not to track every day, acknowledging that they cannot be perfect and track without interruption like a machine, and accepting that people's feelings are more important than data.

(6) User Characteristics

The sixth subset of studies under the end users' perspectives focused on user characteristics. McKinney et al [94] showed that in self-tracking of food and activity levels, high health literacy supports health goals, which comprises proper understanding of a record and how to use data to support a health goal and the awareness of privacy and ownership. The digital divide may also influence the adoption of self-tracking and the quantified self [95]. According to Régnier and Chauvel, better-off individuals use self-tracking more than the socially disadvantaged [95]. In addition, Baumgart conducted a series of studies to investigate the interaction of self-tracking with users' cognitions, behaviors, and emotions [64-66].

Stakeholder Group 2: Patients and People With Illnesses

(1) Usage Experience of Patients and People With Illnesses

In the clinical context, patients, and doctors need to track health records for symptoms or contemporary sensor data. The feasibility and acceptability of using tracking technology may differ from that of healthy individuals. Beukenhorst et al [96] focused on the feasibility and acceptability of smartwatches in patients with knee osteoarthritis; people expressed enthusiasm for self-tracking of health data, but there were barriers to full engagement, such as limited battery lives, technical issues, and unfulfilled expectations. Kim et al [54] conducted a food logger applicability test on patients, and their results demonstrated a high adherence rate. However, wrong and unreasonable usage of weight loss apps may contribute to and exacerbate eating disorders [97]. Self-tracking also demonstrated potential in managing chronic diseases and rehabilitation. Goal-directed self-tracking can help people be well prepared in all stages and support chronic condition management [98]. Mishra et al [99] researched how tracking apps and technologies helped patients cope with Parkinson disease. Furthermore, Vogel et al [40] provided evidence for the positive effects of self-tracking technology on a patient's cardiovascular system.

(2) Management of Patient-Generated Data

Patient-generated data for clinical purposes have also attracted research attention. Ancker et al [39] examined personal data tracking for people with multiple chronic conditions, revealing that health-tracking data may come at an emotional cost, such as depression and anxiety, which could ultimately lead to low adoption levels for consumer health information technology. Ivanov et al [27] provided insights into the influencing factors for sharing health-tracking data. They found that health motivation, the severity of the health problem or condition, and age positively influenced patients to share data with professors.

People who had a certain perceived health status preferred to share data with an acquaintance.

(3) Advantages and Disadvantages of Self-tracking in the Clinical Context

Literature contains specialized research on the advantages and disadvantages of self-tracking in the clinical context. With the development of tracking technology, the clinical context now provides more agency, control, and information for patients, which can help them establish a relationship with professionals [100]. Piras and Miele [38] supported this idea, stating that self-tracking mediates the patient–doctor relationship.

As for the disadvantages of self-tracking, collaborations between patients and professionals involving personal data can easily cause misunderstandings regarding the patient-generated data usage, duration, and type of tracked data. There are also patient privacy issues to address [101]. In terms of special disorders, such as Parkinson disease, Riggare et al [102] indicated that self-tracking gives people a deeper understanding of their motor or nonmotor symptoms and contributes to decision-making regarding their self-care. However, the tracking workload is heavy. It is difficult to know what and how to track, and it is also difficult to ignore the risks of obsessive tracking. Therefore, it is necessary to find a proper balance between burdens and benefits [102].

Stakeholder Group 3: Health Care Professionals and Caregivers

(1) Collaboration Among Patients, Health Care Professionals, and Caregivers

Self-tracking technology provides an easier way to collect patients' health information. Health care professionals are also willing to accept self-tracking data to assess the health status of patients [39]. Therefore, patient-provider collaborations with respect to patient-generated data became one of the research topics for self-tracking usage in clinical setting. In a previous study, Chung et al [101] used a stage-based model of personal informatics and the theory of boundary negotiating artifacts to explain misunderstandings and privacy concerns in the collaboration stage. Prior research has also demonstrated that tracking apps and technologies helped patients cope with Parkinson disease. Based on the tracking data, health care partners who are friends and family members can significantly help Parkinson patients adopt positive strategies [99].

(2) Changes in the Roles of Patients and Professionals

Tracking technology has also shifted the roles of patients and doctors. Owing to these technological changes, patients become health managers, and doctors are health organizers [103]. Schroeder et al [98] showed that goal-directed self-tracking can help people in setting goals, preparing knowledge, and contributing to the patient-doctor collaboration. Piras and Miele [38] supported this idea, stating that patients can negotiate a satisfactory relationship with their health care providers when using self-tracking. However, there is still a gap between users, experts, and self-tracking technologies [78]. To gain a professional view of self-tracking in clinical usage, prior research has collected data from doctors, such as in the study by Gabriels and Moerenhout [103]. They conducted an interview study to explore how medical doctors evaluate self-tracking methods and the changes after using those methods. Gabriels and Moerenhout found that regarding self-care, it is important to emphasize the contextual facets of self-tracking and the involvement of the health care professionals [103].

RQ 6: Directions for Future Research Identified From the Literature

Following our analysis of the themes in the extant literature on self-tracking and the quantified self, we move on to describing future research directions presented in the reviewed literature. Based on the analyses and conclusions, we propose 2 perspectives with 5 main future research directions from the reviewed literature: 2 from the perspective of research design, and 3 from the perspective of research topics (see Table 5).

Table 5. Directions for future research.

Perspective	Future research direction	Description
Research design	Employment of longitudinal research designs	Investigate the influence variables corresponding to different usage stages; changes in the evolution of an individual's acceptance level and self-disclosure over time [28,59,91,99].
Research design	User modalities: regular users, intermittent users, nonusers, and former users	Nonuser group: nonusers and their goal orientations and perceptions about the affordances [14,102]
		Former users: comparative studies with short-term and long-term users to identify barriers; people who effectively quit using web-based fitness communities and wearables [53,68]
		Intermittent users, nonusers, and former users [95]
Research topics	Issues related to data sharing and privacy	Data sharing: active sharing and comparing of digital activity data; employees' attitudes about sharing data; browsing others' tracking data and sharing one's own tracking data; means of sharing health-tracking records [8,27,56,104]
		Privacy and security: personal data privacy and security challenges; privacy awareness; perceived risks around data privacy for employees [27,59,105]
Research topics	Psychological and behavioral aspects of self-tracking	Dependent variables: willingness to make in-app purchases; personal health information management; underlying motivations; decisions of consumers to adopt a self‐tracking technology [61,81,86]
		Independent variables: intrinsic and extrinsic motivation, health literacy, duration of self-tracking, number of devices, level of analysis, and demographic characteristics [10,43,55]
		Moderator variables: individual characteristics and personality types [60]
Research topics	Self-tracking in clinical use	Health-tracking data and patient–doctor relationship: whether and how the data are integrated into the patient–doctor relationship; patient–provider interactions with self-tracking data [101,103]
		Patients' and doctors' attitudes toward self-tracking: exploring doctors' communication needs and perspectives and patients' experiences [99,103]

Research Designs Suggested in the Literature

(1) Employment of Longitudinal Research Designs

Changing habits typically takes time, and changes in people's health often occur over time. Moreover, in different stages of usage, people may have different feelings or intentions, and engagement with their health data may change over time [106]. Prior research has extensively underscored the importance of this evolution over time and that of longitudinal research [28,59,91,99]. A longitudinal research design may also provide a better understanding of the causalities by observing and explaining changes over time [28,91]. Therefore, future studies should pay close attention to the evolution of the aforementioned aspects over time. A series of interesting topics were mentioned in prior research. First, the evolution of consumers' acceptance level could be used to discover users' attitudes and identify the barriers for continued usage [59]. Second, in clinical settings, changes in the roles of caregivers (friends and family members who manage the patients' diseases) in helping patients' self-tracking should be identified via longitudinal research [99].

(2) User Modalities: Regular Users, Intermittent Users, Nonusers, and Former Users

As a second research direction, prior research has identified the need to shift focus from the mainstream users of self-tracking toward an increased emphasis on the modalities of use, namely intermittent users, nonusers, or prior users of self-tracking [14,53,68,95,102]. In extant literature, intermittent users comprise those who use self-tracking and self-quantification but not on a regular basis [95,99]. On the other hand, nonusers are those who are reluctant to adopt self-tracking technologies in all social milieus [95]. Former users are those who have used self-tracking or self-quantification in the past but have quit [68,95]. Focusing on the non-mainstream groups to identify and reduce the burden of self-tracking [102] as well as discovering ways to attract new users [14] are topics worthy of future research.

Research Topics Suggested in the Literature

(1) Issues Related to Data Sharing and Privacy

Issues related to sharing data generated by self-tracking, data privacy, and security have frequently been proposed as areas for future research [8,27,56,59,104,105]. Different experiences and concerns may promote different attitudes toward sharing personal data [104]. For example, Zhou et al [56] indicated that sharing self-tracking data may enable social comparisons. Furthermore, multiplatform and cross-platform data sharing has become a common phenomenon, which has consequently created challenges in terms of data privacy and security [27,105]. The tracking process of data collection, analysis, and storage also raises ethical issues [59].

(2) Psychological and Behavioral Aspects of Self-tracking

Psychological and behavioral research encompasses themes such as users' behaviors, attitudes, intentions, acceptance of technologies, and their respective motivations. The literature on motivation for personal health information management can provide new ideas to improve research design [86]. Existing research shows that future research will benefit from exploring the antecedents of the adoption decisions from a microlevel perspective [81]. Moreover, as the freemium business model is extremely popular among various web-based services and software firms [107,108], existing research highlights that future research could examine consumers' willingness to make in-app purchases in the context of self-tracking and the quantified self [61].

In terms of independent variables, the subdivisions of intrinsic and extrinsic motivations should be considered in the research on usage [10]. In addition, the demographic characteristics, such as the educational level, socioeconomic status, and cultural background, could be examined as predictors of self-tracking usage [55]. Hamari et al [60] recommend that individual characteristics and personality types can be regarded as moderating variables of goal setting and the perception of motivational design.

(3) Self-tracking in Clinical Use

With respect to the clinical apps of self-tracking, prior research has highlighted 2 specific areas for future research. The first direction involves health-tracking data and the patient-doctor relationship [38,100,101]. For example, future studies could explore the patient-doctor interaction with self-tracking in different clinical settings [101] and integrate the health-tracking data in the patient-doctor relationship [103]. Second, patients' and doctors' attitudes toward self-tracking also play important roles in clinical research. Patient usage experiments and the feasibility and acceptability of self-tracking should be investigated in the future [103]. As pointed out by Mishra et al [99], self-tracking can help patients communicate with clinicians. Thus, understanding clinicians' communication needs and attitudes toward self-tracking represents another viable future research area.

Discussion

Key Findings

The purpose of this study was to identify the state of the art in self-tracking and the quantified self in health and well-being. To this end, we conducted an SLR of 67 articles, comprising 42 journal articles and 25 conference papers. We have presented the development of the number of publications over time. We also identified the key academic outlets that published research on self-tracking and the quantified self, and the most prevalent research methods and theoretical foundations.

Our results demonstrate that the outputs provided by self-tracking can be used by various stakeholders. By classifying the stakeholders into end users, patients and people with illnesses, and health care professionals and caregivers, and investigating the focal themes of prior research across these 3 groups, our review provides a structured view of the extant body of knowledge. Moreover, we have classified the future research directions provided in the reviewed studies into two categories: suggestions focusing on research designs and suggestions focusing on research topics. Against this backdrop, we highlight three main findings related to the research subjects, research designs, and the role of theory stemming from the reviewed literature.

First, with respect to the stakeholders being considered as the research subjects, our results demonstrate that of the 67 studies that qualified for the review, the majority (54) focused on end users who use self-tracking to obtain feedback on their sport and fitness activities, daily activity levels, and sleep. Although self-tracking technologies are increasingly invading various clinical settings, we claim that adopting a multi-stakeholder perspective to a greater extent could be a beneficial avenue to advance research in this area.

Second, with respect to research design, considering that in many use cases reaping the potential benefits of the self-tracking requires sustained engagement in the process and use of these technologies, it is not surprising that self-tracking continuance is a key focal area of the reviewed literature. The temporal perspective has also been incorporated in the research design as 21 of the 67 reviewed studies feature longitudinal research. This observation also echoes the notion that employment of longitudinal research methods is one of the key future research directions suggested in the reviewed literature.

Third, with respect to the role of theory, 28 out of the 67 reviewed studies did not explicate or build on any theoretical foundation. In light of this observation, there is still scope for more theory-oriented research to reinforce the theoretical and conceptual foundations and enrich the knowledge base on self-tracking and the quantified self.

Limitations

The results and implications of this review should be evaluated in the light of its limitations. There are 2 main limitations to this review that must be acknowledged. First, the scope of the review is limited by the keywords used in the article search. We specifically employed only "self-tracking" and "quantified self" as the search terms. Consequently, superordinate and subordinate words were not included in the literature search. Second, to keep the scope of the review manageable, we focused solely on empirical articles. Therefore, future research could cover nonempirical research and the superordinate and subordinate keywords of self-tracking and the quantified self.

Future Research Directions

By presenting and summarizing the key studies in the existing body of literature in a systematic fashion, this review will assist future research activities and thus accelerate the development of the relevant research areas. To this end, our review has outlined the theories and methods used in the extant literature. Moreover, we constructed a research theme matrix from three stakeholders' perspectives and identified 5 main future research directions provided in the reviewed literature. Beyond these contributions to research and practice, we delineate 4 primary future research directions based on our observations and

reflections of the current review and provide our views on the shortcomings and limitations of the extant literature that future studies could address.

First, surprisingly few studies in our review have focused on investigating how people perceive and interpret the information produced by tracking devices [9,13,56,83]. Although studies have examined the behavioral outcomes of self-tracking, the human information processing aspects seem to have received less attention. There are two groups of factors that largely determine how users perceive information from self-tracking, namely their cognitive processes and the information along with the way the information is being provided to the users. Thus, we suggest that future research should determine how people make sense of the information produced by self-tracking to ensure that this information fits the cognitive style and knowledge structure of the users.

Second, our review revealed a significant paucity of empirical research examining the potentially adverse psychological consequences of self-tracking. This is notable considering that there is a well-established body of literature discussing self-tracking and the quantified self from a critical standpoint [109] and a stream of literature scrutinizing the so-called dark side of information technology (IT) [110,111]. Prior research in this area has classified the dark side of IT into five categories: IT-usage-related stress, work overload, interruptions, addiction, and misuse [112]. It is plausible to assume that these phenomena may also occur in relation to self-tracking and self-quantification devices. Thus, future research could, for example, examine whether the use of self-tracking and quantified-self technologies may induce health-related obsessive-compulsive thoughts and behaviors, particularly over time. Moreover, future research could examine the role of technology addiction in the continued use of self-tracking technologies.

Third, pertaining to the level of analysis, most of the reviewed studies have examined self-tracking and the quantified self essentially as individual-level phenomena. However, as argued by Lupton [113] and Sharon [37], the rise of self-tracking and the emergence of a market for health-related products and services can be related to a societal shift toward neoliberal thinking. Neoliberalism is generally associated with policies of economic liberalization. Lupton claims that individuals who use self-tracking technology "have readily adopted the subject of the responsible, entrepreneurial citizen as it is privileged in neoliberal governmentality in seeking to take action to achieve healthy and fit embodiments and engaging in self-governance" [113]. Therefore, the individual is increasingly emphasized in the societal discourse. Against this backdrop, future studies could investigate the group- and community-level implications of self-tracking and quantified self.

Fourth, related to the stakeholder perspective adopted in this study, self-tracking has the potential to reinforce people's sense of agency in terms of understanding their health. This in turn can alter the roles of various actors and the power balance among these actors in the health care system. Hence, future studies could focus on the systemic effects of self-tracking from a societal perspective.

Acknowledgments

SF was financially supported by Chinese Scholarship Council (grant 201806770140). We extend our thanks to the participants of the Information Systems Research Seminar in Scandinavia held on the internet in August 2020 and the anonymous reviewers for their comments and suggestions.

Authors' Contributions

SF synthesized the literature according to the described methodology and wrote the first draft of the manuscript. SF and MM reviewed the literature and prepared the manuscript. HS and AD commented on the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Article search strategy.
[XLSX File (Microsoft Excel File), 9 KB - jmir_v23i9e25171_app1.xlsx]

Multimedia Appendix 2
Quality assessment.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v23i9e25171_app2.xlsx]

Multimedia Appendix 3
Details of the reviewed literature (quantitative research).
[XLSX File (Microsoft Excel File), 20 KB - jmir_v23i9e25171_app3.xlsx]

Multimedia Appendix 4

Details of the reviewed literature (qualitative research).
[XLSX File (Microsoft Excel File), 20 KB - jmir_v23i9e25171_app4.xlsx]

Multimedia Appendix 5
Details of the reviewed literature (mixed methods research).
[XLSX File (Microsoft Excel File), 19 KB - jmir_v23i9e25171_app5.xlsx]

Multimedia Appendix 6
Study foci and research theme matrices.
[XLSX File (Microsoft Excel File), 16 KB - jmir_v23i9e25171_app6.xlsx]

References

1. Lupton D. The Quantified Self. Cambridge: Polity Press; 2016.
2. Statista. Most popular Apple App Store categories in August 2020, by share of available apps. https://www.statista.com/statistics/270291/popular-categories-in-the-app-store/. 2020. URL: https://www.statista.com/statistics/270291/popular-categories-in-the-app-store/ [accessed 2020-12-26]
3. PocketGamer.biz. Application category distribution. POCKET GAMER.biz. 2020. URL: https://www.pocketgamer.biz/metrics/app-store/categories/ [accessed 2020-12-26]
4. Most popular Google Play app categories as of 3rd quarter 2020. Statista. 2020. URL: https://www.statista.com/statistics/279286/google-play-android-app-categories/ [accessed 2020-12-26]
5. Number of mHealth apps available at Google Play from 1st quarter 2015 to 3rd quarter 2020. Statista. 2020. URL: https://www.statista.com/statistics/779919/health-apps-available-google-play-worldwide/ [accessed 2020-12-26]
6. Number of mHealth apps available in the Apple App Store from 1st quarter 2015 to 3rd quarter 2020. Statista. 2020. URL: https://www.statista.com/statistics/779910/health-apps-available-ios-worldwide/ [accessed 2020-12-26]
7. Smartwatch shipments worldwide from 2018 to 2023. Statista. 2019. URL: https://www.statista.com/statistics/878144/worldwide-smart-wristwear-shipments-forecast/ [accessed 2021-01-30]
8. Maxwell H, O'Shea M, Stronach M, Pearce S. Empowerment through digital health trackers: an exploration of indigenous Australian women and physical activity in leisure settings. Ann Leis Res 2019 Oct;24(1):150-167. [doi: 10.1080/11745398.2019.1674677]
9. Spotswood F, Shankar A, Piwek L. Changing emotional engagement with running through communal self‐tracking: the implications of 'teleoaffective shaping' for public health. Sociol Heal Illn 2020;42(4):772-788. [doi: 10.1111/1467-9566.13057]
10. Attig C, Franke T. I track, therefore I walk – exploring the motivational costs of wearing activity trackers in actual users. Int J Hum Comput Stud 2019;127:211-224. [doi: 10.1016/j.ijhcs.2018.04.007]
11. Gimpel H, Nüske N, Rückel T, Urbach N, von Entreß-Fürsteneck M. Self-tracking and gamification : analyzing the interplay of motivations, usage and motivation fulfillment. 2019 Presented at: 14th International Conference on Wirtschaftsinformatik; February 24-27, 2019; Siegen, Germany p. 1130-1140.
12. Pingo Z, Narayan B. "My smartwatch told me to see a sleep doctor": a study of activity tracker use. Online Inf Rev 2019;44(2):503-519. [doi: 10.1108/oir-04-2018-0115]
13. Gordon ML, Althoff T, Leskovec J. Goal-setting and achievement in activity tracking apps: a case study of MyFitnessPal. In: Proceedings of the International World Wide Web Conference. 2019 Presented at: WWW '19: The World Wide Web Conference; May 13-17, 2019; San Francisco, CA, USA p. 571-582 URL: http://europepmc.org/abstract/MED/32368761 [doi: 10.1145/3308558.3313432]
14. Rockmann R, Gewald H. Individual fitness app use: the role of goal orientations and motivational affordances. 2019 Presented at: 25th AMCIS – The Americas Conference on Information Systems; August 15-17, 2019; Cancún, Mexico.
15. Kitchenham B, Pretorius R, Budgen D, Brereton OP, Turner M, Niazi M, et al. Systematic literature reviews in software engineering – a tertiary study. Inf Softw Technol 2010;52(8):792-805. [doi: 10.1016/j.infsof.2010.03.006]
16. Petersen K, Vakkalanka S, Kuzniarz L. Guidelines for conducting systematic mapping studies in software engineering: an update. Inf Softw Technol 2015;64:1-18. [doi: 10.1016/j.infsof.2015.03.007]
17. Almalki M, Gray K, Martin-Sanchez F. Activity theory as a theoretical framework for health self-quantification: a systematic review of empirical studies. J Med Internet Res 2016 May;18(5):e131 [FREE Full text] [doi: 10.2196/jmir.5000] [Medline: 27234343]
18. Paton C, Hansen M, Fernandez-Luque L, Lau AYS. Self-tracking, social media and personal health records for patient empowered self-care. Yearb Med Inform 2012;21(01):16-24. [doi: 10.1055/s-0038-1639425]
19. West P, Van Kleek M, Giordano R, Weal M, Shadbolt N. Information quality challenges of patient-generated data in clinical practice. Front Public Health 2017;5:284 [FREE Full text] [doi: 10.3389/fpubh.2017.00284] [Medline: 29209601]
20. Moya JD, Pallud J, Scornavacca E. Self-tracking technologies adoption and utilization: a literature analysis. 2019 Presented at: 25th AMCIS – The Americas Conference on Information Systems; August 15-17, 2019; Cancún, Mexico.

21. Moya JD, Pallud J. Quantified self : a literature review based on the funnel paradigm. 2017 Presented at: European Conference on Information Systems; June 5-10, 2017; Guimarães, Portugal p. 1678-1694.
22. Jiang J, Cameron A. IT-enabled self-monitoring for chronic disease self-management: an interdisciplinary review. MISQ 2020 Jan;44(1):451-508. [doi: 10.25300/misq/2020/15108]
23. Behne A, Arlinghaus T, Kotte N, Teuteberg F. Towards functionalities of self-tracking wearables, their effects on humans and their application areas: Where can we improve? 2020 Presented at: Americas Conference on Information Systems; August 10-14, 2020; Virtual.
24. Lentferink AJ, Oldenhuis HK, de Groot M, Polstra L, Velthuijsen H, van Gemert-Pijnen JE. Key components in eHealth interventions combining self-tracking and persuasive eCoaching to promote a healthier lifestyle: a scoping review. J Med Internet Res 2017 Aug;19(8):e277 [FREE Full text] [doi: 10.2196/jmir.7288] [Medline: 28765103]
25. Riphagen M, van Hout M, Krijnen D, Gootjes G. Learning tomorrow: visualising student and staff physical activity and reflect on it. In: EDULEARN13. 2013 Presented at: 5th International Conference on Education and New Learning Technologies; Jul 1-3, 2013; Barcelona, Spain p. 3209-3214.
26. Frigo A. Life-Stowing from a Digital Media Perspective: Past, Present and Future [dissertation]. Huddinge: Södertörn University; 2017.
27. Ivanov A, Sharman R, Rao HR. Exploring factors impacting sharing health-tracking records. Heal Policy Technol 2015;4(3):263-276. [doi: 10.1016/j.hlpt.2015.04.008]
28. Maltseva K, Lutz C. A quantum of self: a study of self-quantification and self-disclosure. Comput Human Behav 2018:102-114. [doi: 10.1016/j.chb.2017.12.006]
29. World Health Organization. Constitution of the World Health Organization. 2006. URL: https://www.who.int/governance/eb/who_constitution_en.pdf [accessed 2020-12-26]
30. Pfeiffer J, Von Entress-fuersteneck M, Urbach N, Buchwald A. Quantify-me: consumer acceptance of wearable self-tracking devices. 2016 Presented at: 24th European Conference on Information Systems; June 12-15, 2016; Istanbul,Turkey p. 99.
31. Swan M. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. Int J Environ Res Public Health 2009;6(2):492-525. [doi: 10.3390/ijerph6020492]
32. Kourogi M, Kurata T. Personal positioning based on walking locomotion analysis with self-contained sensors and a wearable camera. 2003 Presented at: IEEE / ACM International Symposium on Mixed and Augmented Reality; October 7-10, 2003; Tokyo, Japan p. 103-112.
33. Chen Z, Lin M, Chen F, Lane ND, Cardone G, Wang R, et al. Unobtrusive sleep monitoring using smartphones. 2013 Presented at: 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops; May 5-8; Venice, Italy p. 145-152. [doi: 10.4108/icst.pervasivehealth.2013.252148]
34. Kelley C, Lee B, Wilcox L. Self-tracking for mental wellness: understanding expert perspectives and student experiences. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 May Presented at: CHI: Conference on Human Factors in Computing Systems; May 6-11, 2017; Denver, CO, United States p. 629-641 URL: http://europepmc.org/abstract/MED/28920106 [doi: 10.1145/3025453.3025750]
35. Lomborg S, Frandsen K. Self-tracking as communication. Inf Commun Soc 2016;19(7):1015-1027. [doi: 10.1080/1369118x.2015.1067710]
36. Sharon T, Zandbergen D. From data fetishism to quantifying selves: self-tracking practices and the other values of data. New Media Soc 2017;19(11):1695-1709. [doi: 10.1177/1461444816636090]
37. Sharon T. Self-tracking for health and the quantified self: re-articulating autonomy, solidarity, and authenticity in an age of personalized healthcare. Philos Technol 2017;30(1):93-121. [doi: 10.1007/s13347-016-0215-5]
38. Piras EM, Miele F. Clinical self-tracking and monitoring technologies: negotiations in the ICT-mediated patient–provider relationship. Heal Sociol Rev 2017;26(1):38-53. [doi: 10.4324/9781315108285-4]
39. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. "You get reminded you're a sick person": personal data tracking and patients with multiple chronic conditions. J Med Internet Res 2015 Aug;17(8):e202 [FREE Full text] [doi: 10.2196/jmir.4209] [Medline: 26290186]
40. Vogel J, Auinger A, Riedl R, Kindermann H, Helfert M, Ocenasek H. Digitally enhanced recovery: investigating the use of digital self-tracking for monitoring leisure time physical activity of cardiovascular disease (CVD) patients undergoing cardiac rehabilitation. PLoS One 2017;12(10):e0186261 [FREE Full text] [doi: 10.1371/journal.pone.0186261] [Medline: 29020079]
41. Schwennesen N. When self-tracking enters physical rehabilitation: from 'pushed' self-tracking to ongoing affective encounters in arrangements of care. Digit Health 2017;3:2055207617725231 [FREE Full text] [doi: 10.1177/2055207617725231] [Medline: 29942609]
42. Kari T, Koivunen S, Frank L, Makkonen M, Moilanen P. Critical experiences during the implementation of a self-tracking technology. 2016 Presented at: Proceedings of the 20th Pacific Asia Conference on Information Systems; June 27-July 1, 2016; Chiayi, Taiwan.
43. Sjöklint M, Constantiou I, Trier M. The complexities of self-tracking - an inquiry into user reactions and goal attainment. 2015 Presented at: 23rd European Conference on Information Systems; May 26-29, 2015; Münster, Germany. [doi: 10.2139/ssrn.2611193]

44. Gimpel H, Nißen M, Görlitz RA. Quantifying the quantified self: a study on the motivations of patients to track their own health. 2013 Presented at: 34th International Conference on Information Systems; December 15-18, 2013; Milano, Italy.
45. Petticrew M, Roberts H. Systematic Reviews in the Social Sciences: A Practical Guide. Oxford, United Kingdom: Blackwell Publishing; 2006.
46. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions. 2nd Edition. Chichester, United Kindgom: John Wiley & Sons; 2008.
47. Victor L. Systematic reviewing. Soc Res Updat 2008(54):1-4 [FREE Full text]
48. Okoli C, Schabram K. A guide to conducting a systematic literature review of information systems research. Sprouts Work Pap Inf Syst 2010;10(26):1-49. [doi: 10.2139/ssrn.1954824]
49. Idri A, Amazal FA, Abran A. Analogy-based software development effort estimation: a systematic mapping and review. Inf Softw Technol 2015;58:206-230. [doi: 10.1016/j.infsof.2014.07.013]
50. Behera RK, Bala PK, Dhir A. The emerging role of cognitive computing in healthcare: a systematic literature review. Int J Med Inform 2019 Sep;129:154-166. [doi: 10.1016/j.ijmedinf.2019.04.024] [Medline: 31445250]
51. Sollaci LB, Pereira MG. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. J Med Libr Assoc 2004 Jul;92(3):364-367. [Medline: 15243643]
52. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015;39(2):1-19. [doi: 10.1007/s10916-014-0182-2]
53. Gui X, Chen Y, Caldeira C, Xiao D, Chen Y. When fitness meets social networks: investigating fitness tracking and social practices on WeRun. In: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI Conference on Human Factors in Computing Systems; May 6–11, 2017; Denver, CO, United States p. 1647-1659. [doi: 10.1145/3025453.3025654]
54. Kim Y, Ji S, Lee H, Kim J, Yoo S, Lee J. “My doctor is keeping an eye on me!”: exploring the clinical applicability of a mobile food logger. In: CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: 2016 CHI Conference on Human Factors in Computing Systems; May 07-12, 2016; San Jose, CA, United States p. 5620-5631. [doi: 10.1145/2858036.2858145]
55. Gambier-Ross K, McLernon DJ, Morgan HM. A mixed methods exploratory study of women's relationships with and uses of fertility tracking apps. Digit Health 2018;4:2055207618785077 [FREE Full text] [doi: 10.1177/2055207618785077] [Medline: 30225095]
56. Zhou Y, Kankanhalli A, Huang KW. Effects of fitness applications with SNS: how do they influence physical activity. 2016 Presented at: 37th International Conference on Information Systems; December 11-14, 2016; Dublin, Ireland.
57. Kim J. Analysis of health consumers' behavior using self-tracker for activity, sleep, and diet. Telemed e-Health 2014;20(6):552-558. [doi: 10.1089/tmj.2013.0282]
58. Kim J. A qualitative analysis of user experiences with a self-tracker for activity, sleep, and diet. Interact J Med Res 2014 Mar;3(1):e8 [FREE Full text] [doi: 10.2196/ijmr.2878] [Medline: 24594898]
59. Baudier P, Ammi C, Lecouteux A. Employees’ acceptance of the healthcare Internet of Things: a source of innovation in corporate human resource policies. J Innov Econ Manag 2019(3):89-111. [doi: 10.3917/jie.pr1.051]
60. Hamari J, Hassan L, Dias A. Gamification, quantified-self or social networking? matching users’ goals with motivational technology. User Model User-Adap Inter 2018;28(1):35-74. [doi: 10.1007/s11257-018-9200-2]
61. Chuah SHW. You inspire me and make my life better: investigating a multiple sequential mediation model of smartwatch continuance intention. Telemat Inform 2019 Oct;43:101245. [doi: 10.1016/j.tele.2019.101245]
62. Suh A. Sustaining the use of quantified-self technology: a theoretical extension and empirical test. Asia Pacific J Inf Syst 2018;28(2):114-132. [doi: 10.14329/apjis.2018.28.2.114]
63. Shin D, Biocca F. Health experience model of personal informatics: the case of a quantified self. Comput Human Behav 2017;69:62-74. [doi: 10.1016/j.chb.2016.12.019]
64. Baumgart R. Another step towards the understanding of self-tracking: a research model and pilot test. 2017 Presented at: 23rd Americas Conference on Information Systems; August 10-12, 2017; Boston, MA, United States.
65. Baumgart R. How does quantified self run? 2016 Presented at: 22nd Americas Conference on Information Systems; August 11-14, 2016; San Diego, CA, United States.
66. Baumgart R, Holten R. How does self-tracking go? a research model and pre-test. 2018 Presented at: 51st Hawaii International Conference on System Sciences; January 3-6, 2018; Waikoloa Village, Hawaii, United States p. 3189-3196. [doi: 10.24251/hicss.2018.403]
67. Stiglbauer B, Weber S, Batinic B. Does your health really benefit from using a self-tracking device? evidence from a longitudinal randomized control trial. Comput Human Behav 2019;94:131-139. [doi: 10.1016/j.chb.2019.01.018]
68. Stragier J, Vanden Abeele M, Mechant P, De Marez L. Understanding persistence in the use of online fitness communities: comparing novice and experienced users. Comput Human Behav 2016;64:34-42. [doi: 10.1016/j.chb.2016.06.013]
69. Wittkowski K, Klein JF, Falk T, Schepers JJL, Aspara J, Bergner KN. What gets measured gets done: Can self-tracking technologies enhance advice compliance? J Serv Res 2020;23(3):281-298. [doi: 10.1177/1094670520904424]
70. Abril EP. Tracking myself: assessing the contribution of mobile technologies for self-trackers of weight, diet, or exercise. J Health Commun 2016;21(6):638-646. [doi: 10.1080/10810730.2016.1153756]

71. Webster J, Watson RT. Analyzing the past to prepare for the future: writing a literature review. MIS Q 2002;26(2):xiii-xxiii [FREE Full text]
72. Nokes KM, Reyes DM, Hickey DE. Perceptions of low income HIV+ persons about finding accurate web-based health information. Appl Nurs Res 2018;42:70-76. [doi: 10.1016/j.apnr.2018.06.002]
73. Esmonde K. Training, tracking, and traversing: digital materiality and the production of bodies and/in space in runners' fitness tracking practices. Leis Stud 2019;38(6):804-817. [doi: 10.1080/02614367.2019.1661506]
74. Rönkkö K. An activity tracker and its accompanying app as a motivator for increased exercise and better sleeping nabits for youths in need of social care: field study. JMIR Mhealth Uhealth 2018 Dec;6(12):e193 [FREE Full text] [doi: 10.2196/mhealth.9286] [Medline: 30578186]
75. Faust L, Hachen D, Lizardo O, Chawla NV. Quantifying subjective well-being using trends in weekend activity. 2018 Presented at: IEEE International Conference on Healthcare Informatics; June 4-7, 2018; New York, NY, United States p. 123-129. [doi: 10.1109/ichi.2018.00021]
76. Depper A, Howe PD. Are we fit yet? English adolescent girls' experiences of health and fitness apps. Heal Sociol Rev 2017;26(1):98-112. [doi: 10.4324/9781315108285-8]
77. McMahon SK, Lewis B, Oakes M, Guan W, Wyman JF, Rothman AJ. Older adults' experiences using a commercially available monitor to self-track their physical activity. JMIR Mhealth Uhealth 2016 Apr;4(2):e35 [FREE Full text] [doi: 10.2196/mhealth.5120] [Medline: 27076486]
78. Ravichandran R, Sien SW, Patel SN, Kientz JA, Pina LR. Making sense of sleep sensors: how sleep sensing technologies support and undermine sleep health. In: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI Conference on Human Factors in Computing Systems; May 06-11, 2017; Denver, CO, United States p. 6864-6875. [doi: 10.1145/3025453.3025557]
79. Didžiokaitė G, Saukko P, Greiffenhagen C. The mundane experience of everyday calorie trackers: beyond the metaphor of quantified self. New Media Soc 2018;20(4):1470-1487. [doi: 10.1177/1461444817698478]
80. Milward J, Deluca P, Drummond C, Kimergård A. Developing typologies of user engagement with the BRANCH alcohol-harm reduction smartphone app: qualitative study. JMIR Mhealth Uhealth 2018 Dec;6(12):e11692 [FREE Full text] [doi: 10.2196/11692] [Medline: 30545806]
81. Makkonen M, Kari T, Frank L, Moilanen P. Adoption and adoption interests of self-tracking technologies: single and multiple technology perspectives. In: Proceedings of the 29th Bled eConference "Digital Economy". 2016 Presented at: 29th Bled eConference "Digital Economy"; June 19-22, 2016; Bled, Slovenia.
82. Lyall B, Robards B. Tool, toy and tutor: subjective experiences of digital self-tracking. J Sociol 2017;54(1):108-124. [doi: 10.1177/1440783317722854]
83. Rockmann R, Salou T, Gewald H. If you are happy and don't know IT: Continuance? analyzing emotion carry-over effects in activity tracking continuance decisions. 2018 Presented at: Pacific Asia Conference on Information Systems; June 26-30, 2018; Yokohama, Japan p. 238.
84. Hardey MM. On the body of the consumer: performance-seeking with wearables and health and fitness apps. Sociol Health Illn 2019;41(6):991-1004. [doi: 10.1111/1467-9566.12879]
85. Meyer J, Beck E, Wasmann M, Boll S. Making sense in the long run: long-term health monitoring in real lives. 2017 Presented at: IEEE International Conference on Healthcare Informatics; August 23-26, 2017; Park City, UT, United States p. 285-294. [doi: 10.1109/ichi.2017.11]
86. Feng Y, Li K, Agosto DE. Healthy users' personal health information management from activity trackers: the perspective of gym-goers. Proc Assoc Info Sci Tech 2017 Oct;54(1):71-81. [doi: 10.1002/pra2.2017.14505401009]
87. Yli-Kauhaluoma S, Pantzar M. Seeking connectivity to everyday health and wellness experiences: specificities and consequences of connective gaps in self-tracking data. Digit Health 2018;4:2055207618779714 [FREE Full text] [doi: 10.1177/2055207618779714] [Medline: 29942640]
88. Rapp A, Cena F. Personal informatics for everyday life: how users without prior self-tracking experience engage with personal data. Int J Hum Comput Stud 2016;94:1-17. [doi: 10.1016/j.ijhcs.2016.05.006]
89. Chen J, Bauman A, Allman-Farinelli M. A study to determine the most popular lifestyle smartphone applications and willingness of the public to share their personal data for health research. Telemed J E Health 2016 Aug;22(8):655-665. [doi: 10.1089/tmj.2015.0159] [Medline: 26958742]
90. Zimmer M, Kumar P, Vitak J, Liao Y, Chamberlain Kritikos K. 'There's nothing really they can do with this information': unpacking how users manage privacy boundaries for personal fitness information. Inf Commun Soc 2018:1-18. [doi: 10.1080/1369118x.2018.1543442]
91. Attig C, Franke T. Abandonment of personal quantification: a review and empirical study investigating reasons for wearable activity tracking attrition. Comput Human Behav 2020 Jan;102:223-237. [doi: 10.1016/j.chb.2019.08.025]
92. Harrison D, Bianchi-berthouze P, Bird J. Activity tracking: barriers, workarounds and customisation. In: UbiComp '15: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 7–11, 2015; Osaka, Japan p. 617. [doi: 10.1145/2750858.2805832]

93. Esmonde K. 'There's only so much data you can handle in your life': accommodating and resisting self-surveillance in women's running and fitness tracking practices. Qual Res Sport Exerc Health 2019 May;12(1):76-90. [doi: 10.1080/2159676x.2019.1617188]
94. McKinney P, Cox AM, Sbaffi L. Information literacy in food and activity tracking among parkrunners, people with type 2 diabetes, and people with irritable bowel syndrome: exploratory study. J Med Internet Res 2019 Aug;21(8):e13652 [FREE Full text] [doi: 10.2196/13652] [Medline: 31373277]
95. Régnier F, Chauvel L. Digital inequalities in the use of self-tracking diet and fitness apps: interview study on the influence of social, economic, and cultural factors. JMIR Mhealth Uhealth 2018 Apr;6(4):e101 [FREE Full text] [doi: 10.2196/mhealth.9189] [Medline: 29678807]
96. Beukenhorst AL, Howells K, Cook L, McBeth J, O'Neill TW, Parkes MJ, et al. Engagement and participant experiences with consumer smartwatches for health research: longitudinal, observational feasibility study. JMIR Mhealth Uhealth 2020 Jan;8(1):e14368 [FREE Full text] [doi: 10.2196/14368] [Medline: 32012078]
97. Eikey EV, Reddy MC. "It's definitely been a journey": a qualitative study on how women with eating disorders use weight loss apps. In: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: 2017 CHI Conference on Human Factors in Computing Systems; May 6–11, 2017; Denver, CO, United States. [doi: 10.1145/3025453.3025591]
98. Schroeder J, Karkar R, Murinova N, Fogarty J, Munson SA. Examining opportunities for goal-directed self-tracking to support chronic condition management. Proc ACM Interactive, Mobile, Wearable Ubiquitous Technol 2019;3(4):1-26. [doi: 10.1145/3369809]
99. Mishra SR, Predrag K, John MW, Hekler EB, Omberg L, Kellen M, et al. Supporting coping with Parkinson's disease through self-tracking. In: CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: 2019 CHI Conference on Human Factors in Computing Systems; May 4–9, 2019; Glasgow, Scotland, United Kingdom p. 1-16. [doi: 10.1145/3290605.3300337]
100. Trauth E, Browning ER. Technologized talk: wearable technologies, patient agency, and medical communication in healthcare settings. Int J Sociotechnology Knowl Dev 2018;10(3):1-26. [doi: 10.4018/ijskd.2018070101]
101. Chung CF, Dew K, Cole A, Zia J, Fogarty J, Kientz JA, et al. Boundary negotiating artifacts in personal informatics: patient-provider collaboration with patient-generated data. In: CSCW '16: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. 2016 Presented at: 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing; February 27–March 2, 2016; San Francisco, CA, United States p. 770-786. [doi: 10.1145/2818048.2819926]
102. Riggare S, Scott Duncan T, Hvitfeldt H, Hägglund M. "You have to know why you're doing this": a mixed methods study of the benefits and burdens of self-tracking in Parkinson's disease. BMC Med Inform Decis Mak 2019 Aug;19(1):175 [FREE Full text] [doi: 10.1186/s12911-019-0896-7] [Medline: 31470832]
103. Gabriels K, Moerenhout T. Exploring entertainment medicine and professionalization of self-care: interview study among doctors on the potential effects of digital self-tracking. J Med Internet Res 2018 Jan;20(1):e10 [FREE Full text] [doi: 10.2196/jmir.8040] [Medline: 29330140]
104. Chung CF, Gorm N, Shklovski I, Munson SA. Finding the right fit: understanding health tracking in workplace wellness programs. In: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: 2017 CHI CHI Conference on Human Factors in Computing Systems; May 6–11, 2017; Denver, CO, United States p. 4875-4885. [doi: 10.1145/3025453.3025510]
105. Lupton D. 'It's made me a lot more aware': a new materialist analysis of health self-tracking. Media Int Aust 2019;171(1):66-79. [doi: 10.1177/1329878x19844042]
106. Figueiredo MC, Caldeira C, Eikey E, Mazmanian M, Chen Y. Engaging with health data: the interplay between self-tracking activities and emotions in fertility struggles. Proc ACM Hum Comput Interact 2018;2:1-20. [doi: 10.1145/3274309]
107. Mäntymäki M, Islam AN, Benbasat I. What drives subscribing to premium in freemium services? a consumer value‐based view of differences between upgrading to and staying with premium. Info Systems J 2020 Mar;30(2):295-333. [doi: 10.1111/isj.12262]
108. Mäntymäki M, Hyrynsalmi S, Koskenvoima A. How do small and medium-sized game companies use analytics? an attention-based view of game analytics. Inf Syst Front 2019 Apr;22(5):1163-1178. [doi: 10.1007/s10796-019-09913-1]
109. Lupton D. Quantified sex: a critical analysis of sexual and reproductive self-tracking using apps. Cult Health Sex 2015;17(4):440-453. [doi: 10.1080/13691058.2014.920528] [Medline: 24917459]
110. Salo J, Mäntymäki M, Islam A. The dark side of social media–and Fifty Shades of Grey introduction to the special issue: the dark side of social media. Internet Res 2018;28(5):1166-1168. [doi: 10.1108/intr-10-2018-442]
111. Tarafdar M, Gupta A, Turel O. The dark side of information technology use. Inf Syst J 2013 May;23(3):269-275. [doi: 10.1111/isj.12015]
112. D'Arcy J, Gupta A, Tarafdar M, Turel O. Reflecting on the "dark side" of information technology use. Commun Assoc Inf Syst 2014;35(1):109-118. [doi: 10.17705/1cais.03505]
113. Lupton D. Quantifying the body: monitoring and measuring health in the age of mHealth technologies. Crit Public Health 2013 Dec;23(4):393-403. [doi: 10.1080/09581596.2013.794931]

Abbreviations

IT: information technology
RQs: research questions
SLR: systematic literature review
TAM: technology acceptance model

Edited by R Kukafka; submitted 20.10.20; peer-reviewed by H Makkonen, M Mende; comments to author 18.12.20; revised version received 10.02.21; accepted 16.07.21; published 21.09.21.

Please cite as:
Feng S, Mäntymäki M, Dhir A, Salmela H
How Self-tracking and the Quantified Self Promote Health and Well-being: Systematic Review
J Med Internet Res 2021;23(9):e25171
URL: https://www.jmir.org/2021/9/e25171
doi:10.2196/25171
PMID:34546176

Review

Accuracy of Using Generative Adversarial Networks for Glaucoma Detection: Systematic Review and Bibliometric Analysis

Ali Q Saeed[1,2*], MSc; Siti Norul Huda Sheikh Abdullah[1*], PhD; Jemaima Che-Hamzah[3*], PhD; Ahmad Tarmizi Abdul Ghani[1*], PhD

[1]Center for Cyber Security, Faculty of Information Science & Technology, Universiti Kebangsaan Malaysia, Selangor, Malaysia

[2]Computer Center, Northern Technical University, Ninevah, Iraq

[3]Department of Ophthalmology, Faculty of Medicine, Universiti Kebangsaan Malaysia, Cheras, Kuala Lumpur, Malaysia

[*]all authors contributed equally

Corresponding Author:
Ali Q Saeed, MSc
Center for Cyber Security, Faculty of Information Science & Technology
Universiti Kebangsaan Malaysia
Bangi Street, Bangi
Selangor, 43600
Malaysia
Phone: 60 7740870504
Email: ali.qasim@ntu.edu.iq

Abstract

Background: Glaucoma leads to irreversible blindness. Globally, it is the second most common retinal disease that leads to blindness, slightly less common than cataracts. Therefore, there is a great need to avoid the silent growth of this disease using recently developed generative adversarial networks (GANs).

Objective: This paper aims to introduce a GAN technology for the diagnosis of eye disorders, particularly glaucoma. This paper illustrates deep adversarial learning as a potential diagnostic tool and the challenges involved in its implementation. This study describes and analyzes many of the pitfalls and problems that researchers will need to overcome to implement this kind of technology.

Methods: To organize this review comprehensively, articles and reviews were collected using the following keywords: ("Glaucoma," "optic disc," "blood vessels") and ("receptive field," "loss function," "GAN," "Generative Adversarial Network," "Deep learning," "CNN," "convolutional neural network" OR encoder). The records were identified from 5 highly reputed databases: IEEE Xplore, Web of Science, Scopus, ScienceDirect, and PubMed. These libraries broadly cover the technical and medical literature. Publications within the last 5 years, specifically 2015-2020, were included because the target GAN technique was invented only in 2014 and the publishing date of the collected papers was not earlier than 2016. Duplicate records were removed, and irrelevant titles and abstracts were excluded. In addition, we excluded papers that used optical coherence tomography and visual field images, except for those with 2D images. A large-scale systematic analysis was performed, and then a summarized taxonomy was generated. Furthermore, the results of the collected articles were summarized and a visual representation of the results was presented on a T-shaped matrix diagram. This study was conducted between March 2020 and November 2020.

Results: We found 59 articles after conducting a comprehensive survey of the literature. Among the 59 articles, 30 present actual attempts to synthesize images and provide accurate segmentation/classification using single/multiple landmarks or share certain experiences. The other 29 articles discuss the recent advances in GANs, do practical experiments, and contain analytical studies of retinal disease.

Conclusions: Recent deep learning techniques, namely GANs, have shown encouraging performance in retinal disease detection. Although this methodology involves an extensive computing budget and optimization process, it saturates the greedy nature of deep learning techniques by synthesizing images and solves major medical issues. This paper contributes to this research field by offering a thorough analysis of existing works, highlighting current limitations, and suggesting alternatives to support other researchers and participants in further improving and strengthening future work. Finally, new directions for this research have been identified.

(J Med Internet Res 2021;23(9):e27414) doi:10.2196/27414

KEYWORDS

glaucoma; generative adversarial network; deep learning; systematic literature review; retinal disease; blood vessels; optic disc

Introduction

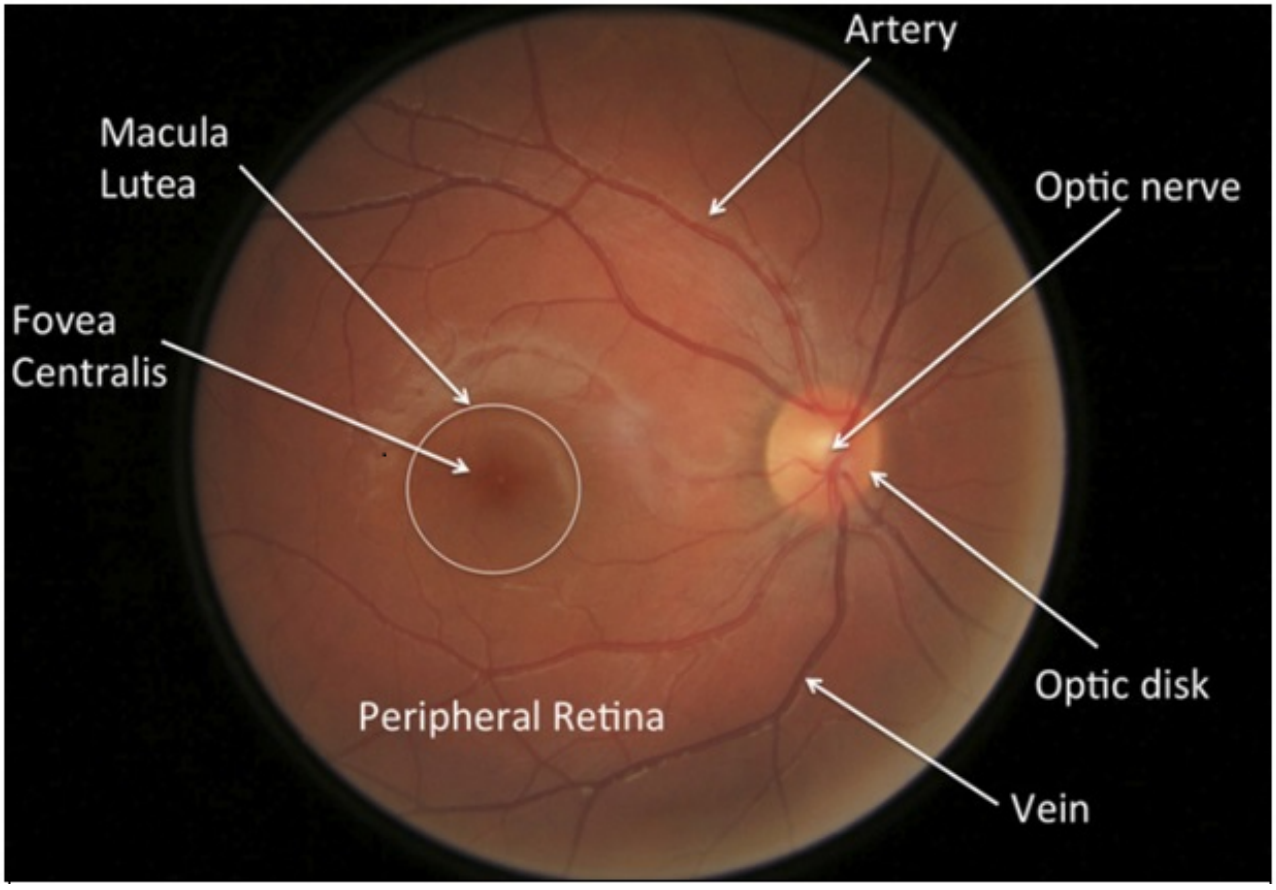
Medical and Statistical Overview

Blindness and visual impairments often result from cataracts, age-related macular degeneration, and glaucoma [1,2]. Glaucoma is a neurodegenerative disease that damages the optic nerve and causes visual field loss [3]. As it is an asymptomatic disease, it is known as the silent thief of sight [4], and patients are unaware of the infection until their vision is irreversibly impaired. Among affected individuals, 50% are ignorant of the disorder [5-7]. Early phases of glaucoma have no symptoms or visual field changes [8]. As the disease progresses, a slow narrowing of the visual field can occur. If left untreated, glaucoma may contribute to total blindness [9]. Loss of vision usually begins on the eye's side and then approaches the middle.

Statistically, glaucoma affects millions of people globally, with more than 64 million cases recorded in 2013, and other studies have estimated that 76 million people will be affected by 2020 and 111.5 million by 2040 [9,10]. Glaucoma is the second leading cause of blindness worldwide, preceded by cataracts [11], and it impacts 4.5 million individuals [9,12], more than 10% of the gross population [10]. Owing to the asymptotic function of glaucoma, approximately 70% of individuals with glaucoma are unaware of the illness's existence [13,14] in the early stage. Thus, we need to provide an early detection and evaluation method [15]. Once glaucoma is detected, a more effective follow-up takes place as a cure can slow down the transmission of the disease [8].

Cataracts may be reversed by surgery, while glaucoma causes lifelong blindness. Elevated intraocular pressure (IOP) is the most common cause of glaucoma. The tonometer measures IOP. However, IOP is not always an accurate and adequate indicator of glaucoma, because glaucoma does not always cause a rise in IOP [16] but rather a deterioration of the optic nerve head (ONH). Visual information flows through the ONH to the brain. The ONH consists of a bright spherical area called the optic disc (OD) and a wider circle-like area called the optic cup (OC). Figure 1 shows these structures in ocular images.

Figure 1. Fundus image structure.



ONH assessment is a widely used glaucoma screening tool that utilizes differential division to distinguish between glaucomatous and normal images [17]. Manual calculations of ONH geometric structures, such as the cup-to-disc ratio (CDR); inferior, superior, nasal, and temporal (ISNT) rule; disc diameter; and rim area, are recommended as diagnostic features for glaucoma screening [18-20]. Among them, the CDR is a reliable therapeutic feature for early glaucoma screening and diagnosis [21,22]. Each of

the derived CDR parameters (diameter or area) is the ratio between the OC and the OD. CDR values rise when the illness progresses and become higher than approximately 0.6-0.7 when the patient has a stronger chance of developing glaucoma [23]. Based on an earlier study [24], a CDR of at least 0.65 is deemed glaucomatous in clinical practice. The CDR score tracks the development of glaucoma over time, effectively screening the condition early [25]. Currently, to check for retinal diseases, specialists tend to manually extract the blood vessel (BV), OD, or OC from retinal images. Accurate segmentation of the retinal structure is very important during the diagnostic process. However, doing this process manually is very labor intensive, time consuming, and risky in terms of human mistakes. Furthermore, the analysis results may lack objectivity, as different experts may produce different results. Therefore, it is important to automate retinal image segmentation/classification while minimizing expert interference.

Research Background

The development of medical imaging technology has helped to accelerate the detection of diseases. Additionally, several studies have been conducted using image processing techniques to automatically process medical images without the intervention of experts [26]. Several studies [22,27,28] have examined vascular tracking and OD and OC segmentation using fundoscopic images. The main segmentation techniques depend on visual features such as color and contrast thresholding, region segmentation, and boundary recognition. Such methods use a learned classifier to classify pixels as foreground pixels (eg, OD, OC, or BV) or as background pixels (regions out of the area of interest) [29,30]. However, most of these methods are based on hand-crafted features (eg, texture, red green blue [RGB] color, gradient and Gabor filter), which are susceptible to low image contrast, pathological regions, and have a lack of deep feature extraction.

In recent years, automatic learning has been significantly improved with the assistance of machine learning (ML) techniques [31]. According to several studies [32,33], ML and deep learning (DL) algorithms have evolved to the point that they can compete with and sometimes even outperform humans on certain tasks, such as object detection [34] and image classification on ImageNet [35]. Currently, deep learning methods (DLMs) are an active research field because they can automatically generate and learn extremely complex features from input data. In particular, DLMs with deeper and complicated perceptron layers [eg, convolutional neural networks (CNNs)] have shown better performance in object detection than other methods [33]. Researchers have attempted to use various types of architectures, such as GoogLeNet [36], AlexNet [33], and DenseNet [37], for glaucoma diagnosis with the introduction of deep neural networks. Such research mainly focuses on 2 aspects: using DL for complex and deep feature extraction and utilizing medical features and spatial domain knowledge in the detection process. However, the use of deep fully connected networks is susceptible to imbalanced learning problems such as high false-negative or false-positive rates, leading to more fake or skinny branches than those of the ground truth [38,39]. In other words, retinal BV segmentation still has issues such as false pathological information segmentation and low microvascular segmentation [40].

For addressing complex learning issues, deep architectures often have advantages over shallow architectures; for example, deep CNNs have demonstrated significant efficiency improvements over conventional vision-based models [41]. A fully connected convolutional network has been used to address insufficient public data. Such methods, however, create very fuzzy vessels with false positives along with tiny and weak roots. This error primarily occurs because the CNNs used in current methods depend solely on pixel-level objective feature to equate the standard image to the image created by the model and are incapable of adapting actively to the fundus image of the natural vascular structure [42]. Empirical studies have proven that deep CNNs can learn invariant representations and attain human-level success if sufficient training data are provided. However, one of the leading shortfalls of DLMs is the lack of available data. Medical data annotation often requires specific domain experts. This shortage leads to the need for CNN training approaches with a limited number of annotated data. However, this can easily lead to underfitting, and as a result, high error rates on both training and testing data are recorded. Lahiri et al [43] demonstrated the effectiveness of using generative adversarial networks (GANs) [44] to perform some discriminative task with only 0.8%-1.6% of the amount of annotation data used by other methods.

GANs belong to the family of unsupervised learning algorithms that have proven their merits in generating synthetic images close to real images and solving image-to-image translation problems in the natural domain [45,46]. GANs have gradually shown their extraordinary ability and have started to shine brilliantly in various application fields [45,47,48]. Inspired by the prevailing learning capability of GANs, Wu et al [49] proposed the generative adversarial network with U-net, referred as (U-GAN), which includes an attention gate model in the generator and a densely connected convolutional network to segment the BVs automatically. Lahiri et al [50] proposed deep convolutional GANs (DCGANs) for retinal segmentation to segment the region of interest (ROI) from a given image. In addition to segmentation tasks, the synthesis of retinal images is a large part of the literature. Haoqi and Ogawara [51] trained a GAN model to learn the mappings of vessels from retinal images to segmented images for training a model to generate a synthesized image close to a given real image.

To date, several review articles summarizing the technology of DL in ophthalmology have been published [20,52-55]. Nevertheless, none of them have particularly focused on the emerging breakthrough GAN techniques using fundus photographs. Moreover, despite the rapid development of telecommunication technology, only a few study groups have examined the possibility of integrating artificial intelligence (AI) technologies with teleophthalmology [56]. To the best of our knowledge, no researchers have adopted telescreening for glaucoma using DL techniques, particularly the GAN.

Shedding light on the importance of telecommunication technology in DL techniques is a current and very urgent need. Alongside the emergence of newer low-cost handheld devices,

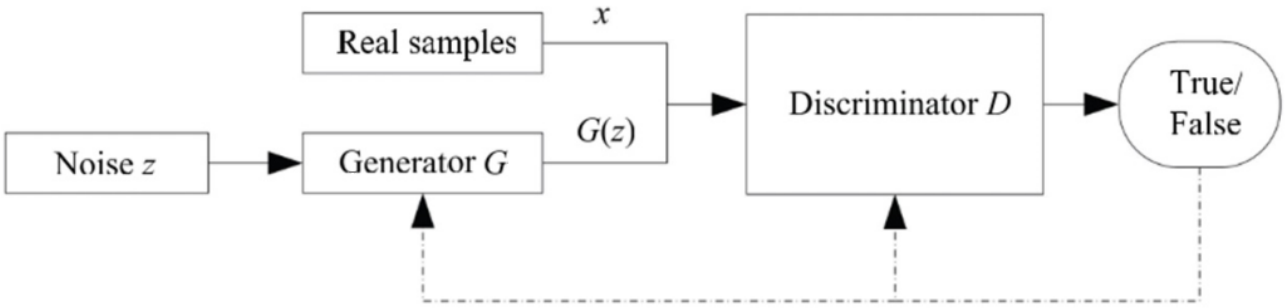
glaucoma screening will become more available, even to distant and poor communities. In addition, maintaining social distance is very important for mitigating the spread of the coronavirus pandemic. This paper summarizes the work in the literature on glaucoma diagnosis and highlights the challenges and gaps of current studies to uncover the possibilities of filling these gaps with the recommended suggestions. We aim to elucidate all research efforts, such as the GAN architectures mentioned earlier, that have been developed in response to the new and disruptive technology, mapping the research landscape from the literature onto a coherent taxonomy of the key features that characterize such an emerging line of research. Finally, the future work of this research will be proposed and described in detail.

Methods

Basic Theory of GANs

We start by reviewing the concept of GANs [44]. GANs consist of 2 separate neural networks, a generation network (G) and a discriminator network (D), plus a noise vector (z) sampled from a known distribution (eg, a Gaussian distribution), which is used to generate data points (fake samples; see Figure 2). A 2-player min-max game inspires the basic idea of this technique. The goal is to train the generator G to learn to capture the potential distribution in the real data sample and generate a new sample close to the real data to deceive the discriminator. The discriminator D is a binary classifier that attempts to discriminate whether the input data are real or fake [44]. To win the game, both G and D need to continuously improve their generation and discrimination capabilities, respectively. The training process lasts until both G and D reach a convergence point (Nash equilibrium), where G generates an output distribution very close to the real data distribution [42,57,58].

Figure 2. GAN architecture. GAN: generative adversarial network.



Mathematically, let G be parameterized by θ, which takes random noise z as input and produces synthetic images G(z) as output. The generated G(z) is mapped from a distribution G(z; θ) ~ p_g. Additionally, the training data set x is sampled from the real data distribution p_{data}, and the objective function of the generator network is used to train G to approximate p_{data} using p_g. By contrast, the discriminator (D) takes either the original image x or G(z) as input and indicates whether the input is from a true data distribution (x) or a synthetic data distribution G(z) by outputting a probability of D(x) or D(G[z]). This can be seen in the followig equation, where p_{data}(x) is the true data distribution and p_z(z) is the noise distribution.

However, the training mechanism of such a model is critical. Unbalanced training between the G and D networks leads to model collapse. This happens when D is trained much better than G. In this case, D is able to easily discriminate between the real and synthetic images generated by G and reject all its outputs; thus, the loss log{1 − D(G[z])} saturates, and G learns nothing from the zero gradient. To avoid the model collapse issue, the loss function of G should be trained to maximize logD(G[z]) instead of minimizing log{1 − D(G[z])}. This can avoid the saturation of the gradient and provides the same gradient direction as that yielded by the old loss function.

Extension Models of GANs

The first GAN [44] was composed of fully connected layers. Later, the DCGAN [59] introduced the use of fully CNNs to increase training stability and improve efficiency. Since then, many GAN models have followed this set up as the main components of GAN architecture. Unlike the DCGAN, the Wasserstein GAN (WGAN) [60] increases the permutation in the fully connected layer. In this model, the Wasserstein distance metric is used instead of the Jensen–Shannon divergence to measure the distance between the generated data distribution and the real data distribution. Therefore, the problems of model collapse and training instability were partially solved in this model. Subsequently, an improved version of the WGAN called the WGAN-GP (gradient penalty) [61] was proposed. The WGAN-GP depends on gradient penalty replacement so that it can solve slow training problems encountered by the WGAN. Moreover, inspired by the WGAN, Mao et al [62] proposed the least-squares GAN (LSGAN) to improve the quality of the generated images. The main idea of the LSGAN is to use a new loss function in the D network for smooth and unsaturated gradients.

The original GAN randomly generated a date distribution that is beyond our control, as the output depends on random noise. Therefore, a conditional GAN (cGAN) was invented to add a vector c as a conditional input to the noise vector z so that the

generator could generate the required data. Hence, the generator output of the cGAN was defined by G(c,z).

Since the cGAN was proposed, many articles have used the cGAN applications, for example, Pix2Pix [45], a cGAN-based technique proposed by PatchGAN to map a set of images to another image using N × N pixels. It classifies each N × N path of the image and averages all the scores of patches to obtain the final score for the image. The main limitation of Pix2Pix is that it requires images x1 and y1 that are paired with each other in the training stage. By contrast, CycleGAN [47], which is also a cGAN-based technique, utilizes an image translation method that does not need paired data, even though Pix2Pix still outperforms CycleGAN's remarkable margin.

Another variation of the GAN combines a variational autoencoder (VAE) and a GAN in a single model named VAE-GAN [63]. The idea behind this technique is to exploit the strength of both the GAN and VAE, as the GAN can generate sharp images but misses some modes while the VAE produces blurry images but with a large variety. Studies have demonstrated that VAE-GAN images are better than those produced by the VAE or GAN alone.

Information Sources

Guided by [64], we conducted a comprehensive search to find all GANs-based articles related to glaucoma by searching the best and most reliable libraries: (1) Scopus, (2) ScienceDirect, (3) IEEE Xplore, (4) Web of Science, and (5) PubMed Central. This collection includes technical and medical literature, perfectly reflecting all research activities in this discipline.

Study Selection Procedure

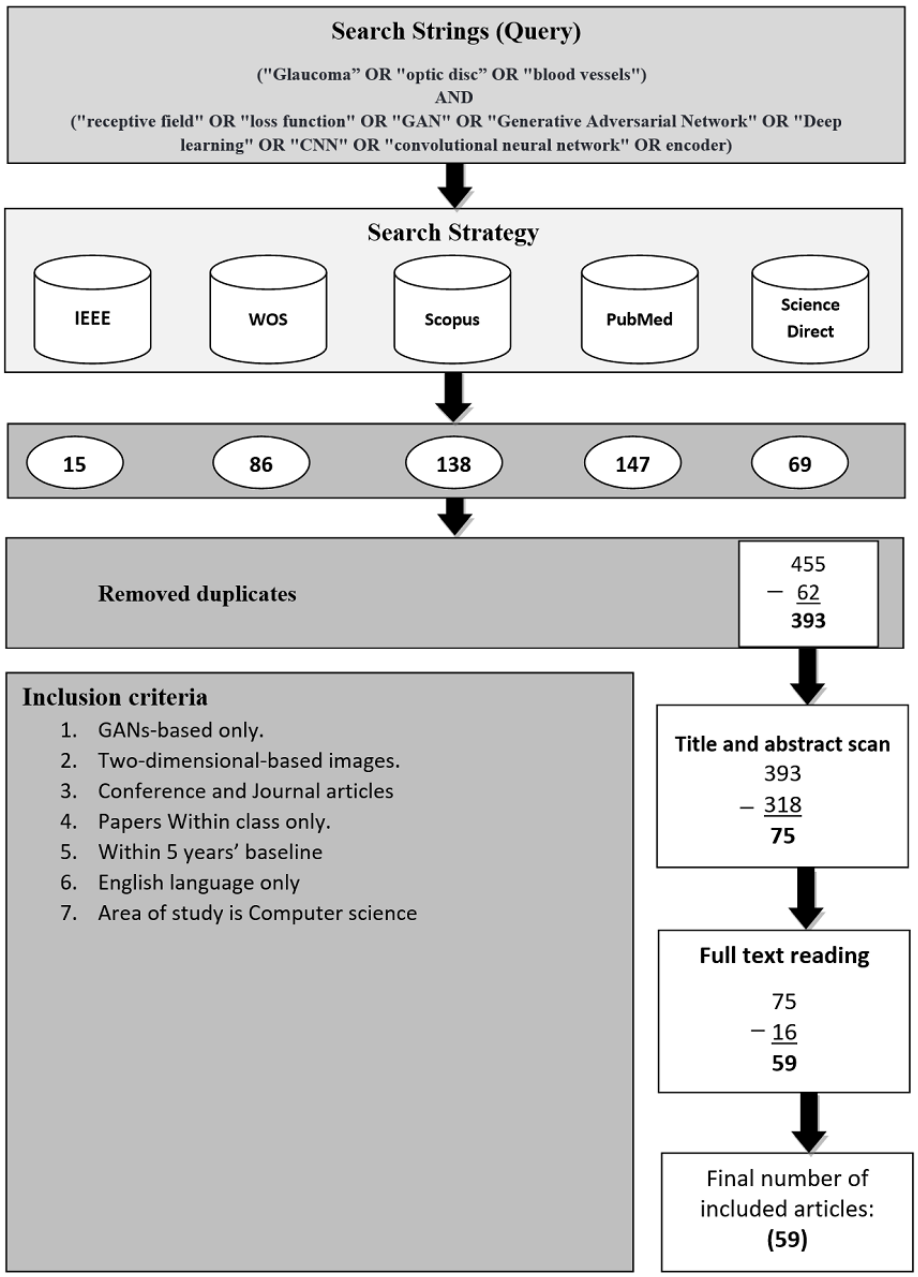
The method for choosing appropriate studies was on the basis of 2 stages: screening and filtering. Successively, both stages met the same criterion for inclusion and exclusion. Both duplicates and unrelated studies by title and abstract skimming were omitted during the first stage. Then, the result in a set of papers was entirely read, analyzed, and summarized in the filtration stage.

Search

This work was carried out between March 2020 and November 2020. Various keyword combinations were used in the search of highly reputable libraries (IEEE Xplore, Science Direct, PubMed, Scopus, and Web of Science). Our search query consist of 2 parts that are connected with each other using the operator "and." The following set of keywords ("glaucoma," "optic disk," "blood vessels") and ("receptive field," "loss function," "GAN," "generative adversarial network," "deep learning," "convolutional neural network," "CNN," Encoder) belong to the first and second parts, respectively. The operator "or" is used to connect keywords within the same part. Based on this, our study scope is formulated.

The quest focused on different journals and conferences and omitted books and all other forms of literature. Therefore, we mainly concentrated on up-to-date and applicable scientific studies related to the use of GANs in retinal disease, especially glaucoma. Figure 3 shows the research query and inclusion criteria used in this work.

Figure 3. Flowchart of the study selection with the research query and inclusion criteria. GAN: generative adversarial network; WOS: Web of Science.



Validity of the Collected Papers (Scope Validation)

The total number of keywords in the collected papers was 115. To validate our research scope, we analyzed these keywords and categorized them according to their co-occurrences. Then, we set a threshold indicating the co-occurrences of each keyword across all papers. Let k≥3, where k is a threshold. As a result, we obtained 15 keywords out of 115 that met the threshold. That is, each of these 15 keywords occurred at least three times in all the collected papers.

Figure 4 illustrates the connections of these 15 keywords to each other. The size of each circle indicates how frequently a single corresponding keyword occurred. The more frequently a keyword occurred, the larger circle size it gets, for example, the keyword "deep learning" has the biggest circle size in the diagram, which means it is the most frequently appeared keyword in the collected papers. The second factor is the color, which indicates how often a single keyword occurred per year. The last factor is the total link strength, which indicates the total connection of a keyword to other keywords. The more frequently 2 keywords appeared in the same article, the thicker is the line drawn between them. For example, the keywords "deep learning" and "glaucoma" were linked by a thicker line than the line between the keywords "generative adversarial network" and "glaucoma," which means that both "deep learning" and "glaucoma" appeared together in the collected articles more than the keywords "generative adversarial network" and "glaucoma" did. This indication reveals that GANs have been used less than other DL techniques in glaucoma detection.

Figure 4. Scope validation diagram.

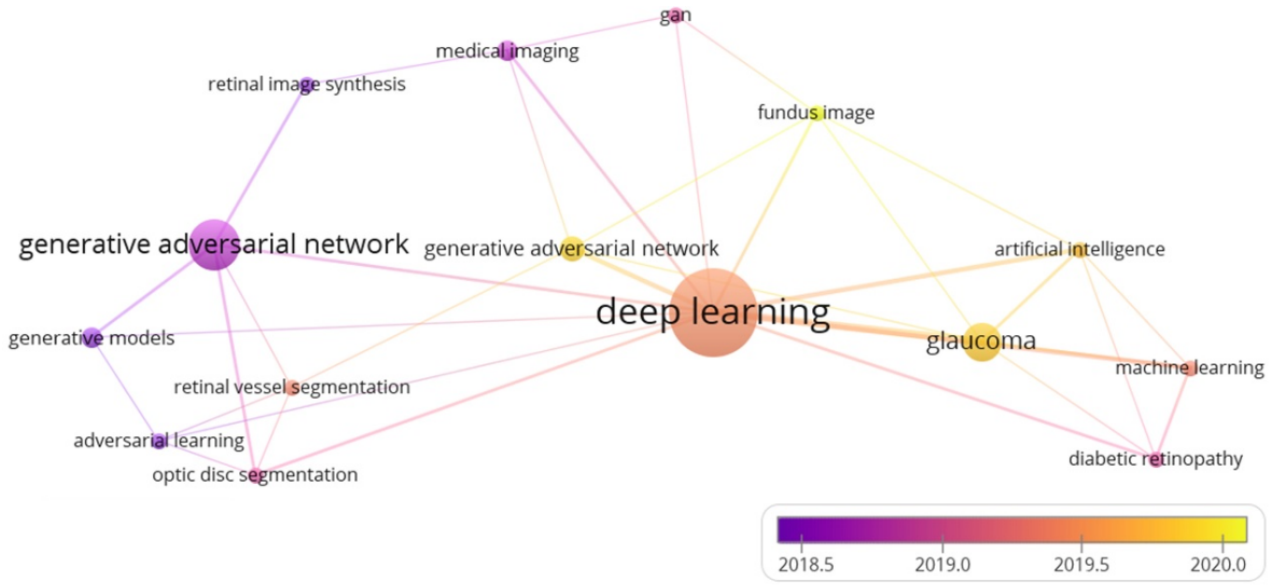


Table 1 shows the occurrences and the total link strength of the 15 keywords that met our threshold (k≥3). Deep learning is the most frequently occurring keyword, which has uniquely appeared 20 times in the collected articles and 27 times with other different keywords. GANs occurred 17 times, with 18 connections to other keywords, while the keyword glaucoma occurred 8 times, with 12 connections to other keywords. In conclusion, these highest scores for the aforementioned keywords empirically demonstrated the validity of our search query that is used to collect literature publications. Furthermore, it proves that our research scope revolves around 3 main keywords, namely, deep learning, generative adversarial network, and glaucoma, as they have the biggest circle sizes with the thickest connection among them.

Table 1. Keywords occurrence.

Keywords	Occurrences	Total link strength
Techniques		
Deep learning	20	27
Generative adversarial network(s)/GAN	17	18
Artificial intelligence	3	8
Machine learning	3	8
Diseases		
Glaucoma	8	12
Diabetic retinopathy	3	6
Imaging		
Fundus image	3	6
Medical imaging	4	5
Papers' contribution		
Adversarial learning	3	4
Optic disc segmentation	3	6
Retinal vessel segmentation	3	4
Generative models	3	3
Retinal image synthesis	3	3

Inclusion and Exclusion Criteria

In this section, papers that met the criteria in Figure 3 were included. We taxonomized the included papers on a general and in-depth diagram consisting of 2 paper groups, namely, the development studies group and the reviews and surveys group. The papers in the first group were classified according to 8 consecutive layers. In the literature, researchers classified GANs into 2-4 categories; these categories were separately used by different researchers, as referenced accordingly in the points

below. However, in our taxonomy, we combined them all. Furthermore, we added 4 more classification criteria as follows: (1) method architecture (direct, hierarchical, iterative) [32]; (2) model structure (2 players, multiple players) [65]; (3) GAN category (optimization function, structure, and conditional) [66-68]; and (4) generator backbone (U-Net based or CNN based) [69].

Further, we added 4 additional categories as follows: (1) type of GAN used in a paper (eg, variational autoencoder with GAN [VAEGAN], DCGAN, cGAN, CycleGAN); (2) discriminator's receptive field (PixelGAN, PatchGAN, ImageGAN); (3) landmarks used during the segmentation/classification process (single, multiple); (4) paper contributions (segmentation, classification, image synthesis, mixed).

The exclusion criteria followed in this paper were as follows: (1) ML approaches, (2) 3D-based imaging methods (optical coherence tomography), (3) between-class papers, and (4) out-of-scope papers.

Data Collection Process

All papers from different sources were summarized and saved in a single spreadsheet file for simplicity and a quick review. Significant remarks and comments were illustrated by full-text reading in our analysis scope and classification stage, which further refined our taxonomy. Finally, our results were summarized on an Excel sheet (Microsoft) and listed in a tabular format. The additional data set includes a list of articles, publishing source, articles' abstracts and contributions, the tools used in papers, audiences, objectives, architecture-based categorization table, and a list of relevant figures.

Results

Overview

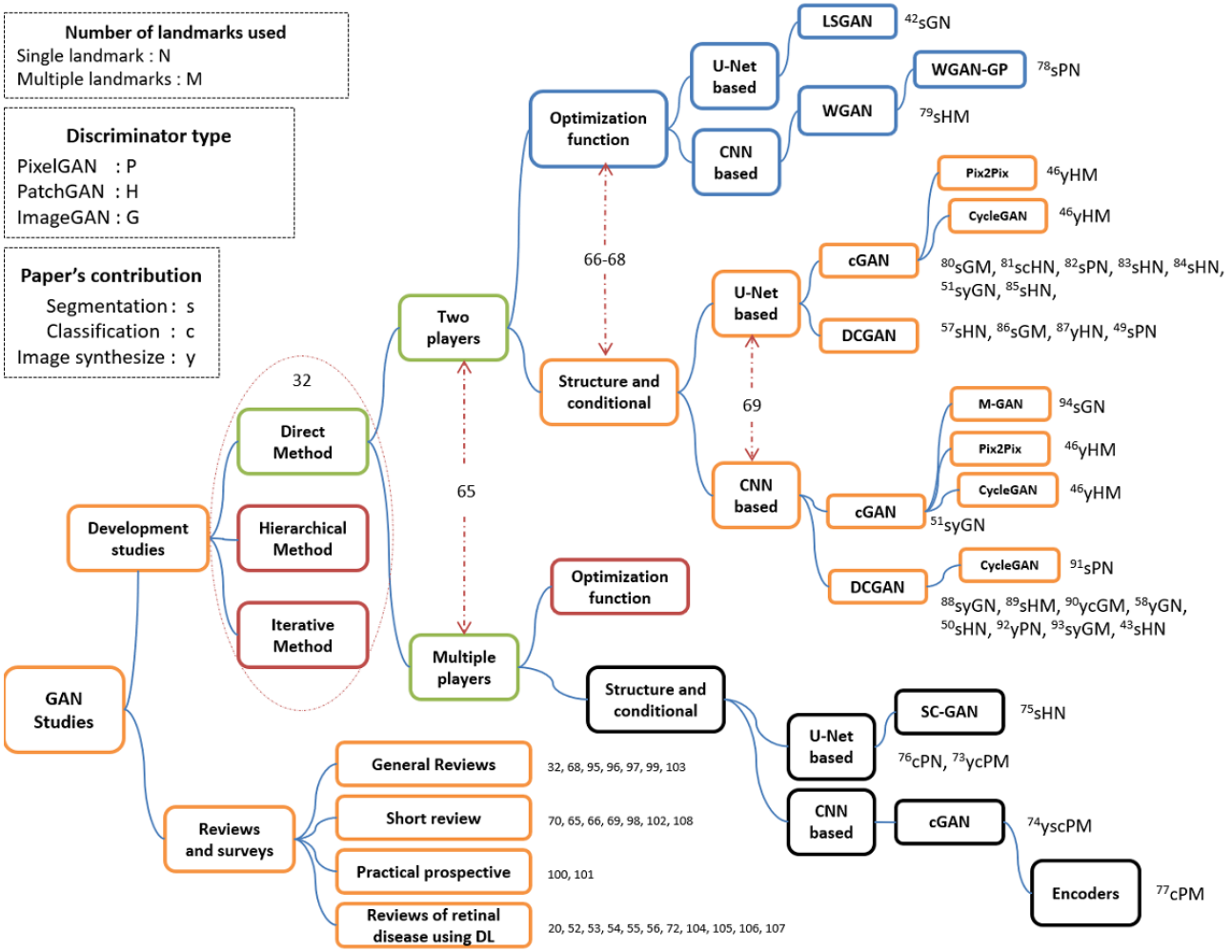
The cumulative number of articles in the original search process was 455. Eighty percent (364/455) of the findings released in 2018-2021 and 20% (91/455) in 2015-2017 were distributed as follows: 15 papers from IEEE Xplore, 86 from Web of Science, 138 from Scopus, 147 from PubMed, and 69 from ScienceDirect. Approximately 62 papers were duplicates across the 5 databases.

Later, 318 papers (not GAN based) were omitted after skimming through the articles' titles and abstracts, leaving only 75 papers. Further screening via full-text reading was carried out on these 75 papers, which resulted in excluding 16 nonrelevant papers. A comprehensive reading was performed on the final 59 papers to create a general map to study this newly emerging methodology.

Of these 59 papers, 51% (n=30) focused on the development and training of various GAN models and real attempts to improve the efficiency of the network architecture to improve segmentation/classification precision, especially at an early stage of the disease with fewer false positives/negatives. Nearly 49% (29/59) of publications included general reviews and surveys relating to GAN technique and its variants; recent GAN applications, limitations, and potential future prospects; reviews of retinal diseases; various DL detection methods; general analytical knowledge such as the most frequently used data sets; and the countries contributing to the current research area. From all these observations, we got a thorough view on the literature, determined the general categories of the study scope, and boosted the taxonomic classification of the literature. Figure 5 presents the groupings of the GAN-based approaches used in the literature according to their structures or optimization functions.

Kumar and Dhawan [70] classified GANs based on their architectures or the loss functions used to train their generators. It is worth noting that the first 4 layers of our taxonomy have been separately used in other papers; therefore, inspired by those studies, we used these categories together as a baseline for our taxonomy. We added other categories to classify brief literature works in depth according to (1) their level of feature discrimination (PixelGAN, PatchGAN, or ImageGAN), (2) the numbers of landmarks used in the segmentation or classification process (a single landmark or multiple landmarks), (3) the backbones of the GANs used in the articles (eg, DCGAN [59], Info-GAN [71], WGAN [60], CGAN [45], Pix2Pix [45], and Cycle-GAN [47]), and (4) the contribution of each paper (eg, segmentation [s], classification [c], or synthesis [y]). In the following sections, we describe each category and provide some accompanying statistics.

Figure 5. Taxonomy of the literature on glaucoma screening based GANs technique. cGAN: conditional GAN; CNN: convolutional neural network; DCGAN: deep convolutional generative adversarial network; DL: deep learning; GAN: generative adversarial network; LSGAN: least-square GAN; WGAN-GP: Wasserstein GAN-gradient penalty.



Development Studies Category

GANs were first developed by Goodfellow et al [44] in 2014. Although researchers have continuously attempted to improve the performance of GANs in various ways, such as weight regularization, new loss functions, weight pruning, and Nash equilibrium, it is still a new research field among deep learning techniques [70,72]. Only recently did this technique start to be adopted by researchers in the field of retinal disease, particularly glaucoma (roughly at the beginning of 2018). Therefore, the total set of papers that described various experiments and tools used for the detection or segmentation of retinal images included 30/59 (51%) articles.

Among these categories, it is notable in Figure 5 that the first 4 layers classified articles based on the method used (direct, hierarchical, or iterative) [32], the model structure [65], the architecture category (optimization function or structure and conditional based) [66-68], and the generator's backbone (CNN based or U-Net based) [69] consecutively.

In the first layer, all the literature work followed the direct methods. This means that all these methods follow the philosophy of using 1 generator and 1 discriminator, and the structures of the G and D are straightforward without any branches. None of the articles used hierarchal or iterative methods; this reveals a new opportunity to apply GANs in the field of retinal disease.

The second layer classified articles based on the number of players. Nearly 25/30 (83%) articles used 2 players, and only 5/30 (17%) articles utilized multiple players. In the latter case, some studies used 3 player-based methods [73-75], with the frameworks of [74] and [75] comprising segmentation, generator, and discriminator networks. In the study by Liu et al [74], the segmentation network and generator enlarged the training data set to improve the segmentation performance, while the discriminator solely focused on identifying fake image–label pairs to ensure compatible utilities. However, in Yu et al [75], the same architecture was used to synthesize images after performing traditional annotation-free methods to obtain coarse segmentations.

A slight difference was observed in Wang et al [73], where a pathology-aware visualization network was used instead of the segmentation network, with both pathology-aware visualization and the generator used to enhance the synthesized glaucoma images in specific pathological areas. The synthesized image was re-enforced to provide a heatmap close to that of the input reference image. The Patho-GAN can thus generate images of glaucoma fundus with clearer pathologies. In Yang et al [76],

the VGG19 network was incorporated with the 3 players to find the topology structure loss, which was combined with the other 3 losses (adversarial loss, weighted cross-entropy loss, and total variation loss) to be used by the generator. However, in [77], the authors used 2 encoders, namely, E_s and E_t, where (s) is the source domain and (t) is the target domain; these encoders were trained to impede the classification performance of the discriminators (D+, D–). In turn, D+ and D– were trained to distinguish between positive/negative source images and positive/negative target images, and finally, a classifier (C) tried to classify source/target images.

Following [66-68], we added a third layer to our taxonomy to classify papers as either structure-based or optimization-based methods. The majority of studies (27/30, 90%) at this level were structure- and conditional-based methods, while only 3/30 (10%) of the studies, namely, those in [42,78,79], were optimization-based methods with 2-player structures; none of these methods have been recorded as multiplayer-based structures.

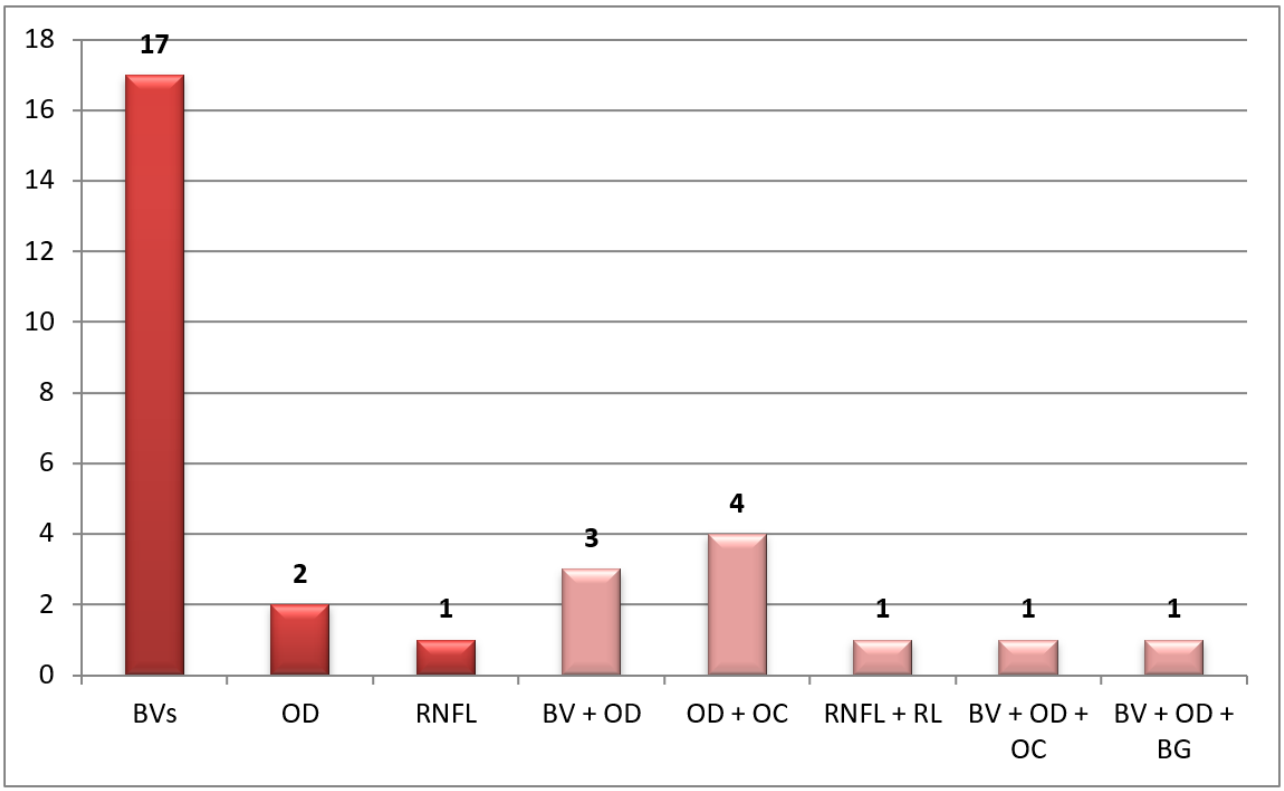
Some researchers tend to use objective function–based methods by updating specific loss functions or using a combination of losses to overcome the model collapse of GANs. This occurs when the generator continuously generates images with the same distribution or generates images with the same texture themes or color as the original image but with marginal differences in human understanding [65]; for example, Ma et al [42] used a least-squares loss function instead of sigmoid cross-entropy. Therefore, their experiment greatly improved the segmentation accuracy of the utilized model on both the digital retinal image for vessels extraction (DRIVE) and structured analysis of the retina (STARE) data sets by forcing the generator to generate images with distributions close to those of the real images. In Tu et al [78], the authors used the WGAN-GP method to overcome the training instability of the traditional GAN and generate accurate probability maps of BVs. The WGAN-GP is an extension of the WGAN; it uses a gradient penalty instead of weight clipping to enforce the Lipschitz constraint. This type of GAN can be trained faster and generates higher-quality samples than those produced by WGANs [61,68,70,78]. Last, Kadambi et al [79] proposed a framework for domain adaptation guided by the Wasserstein distance metric instead of typical adversarial methods for more stable training and better convergence.

The subsequent layer in our taxonomy was to classify methods according to the generator's backbone (eg, U-Net based or CNN based) [69]. Papers [42,46,49,51,57,73,75,76,80-87] represented about 50% of the studies (n=16) and were U-Net-based architectures. However, the other 50% of the papers [43,46,50,51,58,74,77-79,88-94] were CNN-based generators (n=16).

The study by Yu et al [46] was very intensive; the authors proposed multiple-channels-multiple-landmarks as a new preprocessing framework. They used a combination of landmarks (vessel trees, ODs, and OC images) to synthesize colored images with 2 types of GANs (Pix2Pix and CycleGAN). Additionally, they used a Pix2Pix architecture with 2 different generator structures (eg, U-Net-based and CNN-based). They empirically demonstrated that the Pix2Pix network with a ResU-Net generator using high-resolution paired images and multiple-channels-multiple-landmarks outperforms every single landmark-based GAN method regardless of their architectures. Furthermore, they were able to generate significant and realistic images.

The next distinguishing level in our taxonomy addressed the landmarks used in the papers. As Figure 5 shows, references containing "N" letters refer to a single landmark (eg, the BV, OD, OC, retinal nerve fiber layer [RNFL], or rim loss [RL]). These references contributed to 20/30 (67%) of the total papers. Seventeen of them were BV-based methods [42,43,49-51,58,75,76,78,81-85,88,91,92,94]. Only 2 studies [57,81] were OD-based detection approaches, and 1 [82] utilized RNFL-based detection (Figure 6).

Figure 6. Distribution of papers per landmark(s). BG: background; BV, blood vessel; OC: optic cup; OD: optic disc; RL: rim loss; RNFL: retinal nerve fiber layer.



Another set of articles used multiple landmarks and was represented with an "M" letter in Figure 5. These articles contributed to 33% (10/30) of total papers. Some studies [80,90,93] used the BV and OD, while [74,77,79,86,89] used the OD and OC to classify the disease. In addition, Wang et al [73] used RL and RNFL, and Yu et al [46] used BV, OD, and OC.

The rest of the researchers used multiple landmarks, such as [74,77,79,86,89], which involved OD and OC segmentation. Studies [80,93] worked on BV and OD segmentation, and only Wang et al [73] used RNFL and RL. The rest of the papers used triple landmarks in their work, such as [58] and [90], which involved work on BV, OD, and background, and Yu et al [46] used BV, OD, and OC.

In the next layer of our taxonomy, articles were classified according to the discriminator's receptive field. As illustrated in Figure 5, references with P, H, or G letters represent refer to PixelGAN, PatchGAN, or ImageGAN, respectively. ImageGAN papers were [42,51,58,80,86,88,90,93,94], while PixelGAN papers were [49,73,74,76-78,82,91,92]. In addition, PatchGAN papers were [43,46,50,57,75,79,81,83-85,87,89].

Isola et al [45] proposed a Pix2Pix-based conditional adversarial network (cGAN) as a general-purpose solution to image-to-image translation problems, and demonstrated that a 70 × 70 PatchGAN alleviates artifacts and achieves the best scores. Scaling beyond 70 × 70 to a full 286 × 286 ImageGAN did not appear to improve the quality of the results and, in fact, the latter model obtained a considerably lower fully connected network (FCN) score. This scaling mechanism may have been effective because there are more parameters in ImageGAN than PatchGAN and greater depth, which made it harder to train. By contrast, 3 studies [57,81,89] proved that the 64 × 64 Patch-SAN is the best, while one [84] concluded that a 120 × 120 patch is better than a 64 × 64 patch size. Studies [80,88] concluded that ImageGAN is better than PatchGAN. Last, pixel-level annotation [50] is much more tedious than image-level annotation.

Each reference in Figure 5 is denoted with a letter indicating the contribution of the relevant paper. Nearly 57% (17/30) of papers worked on the segmentation task and were denoted by (s), 17% (5/30) worked on image synthesis and were denoted by (y), and only 2 papers worked on the classification task and were denoted by (c). The remaining 6/30 (20%) papers worked on multiple tasks (eg, sc, sy, ysc). Multimedia Appendices 1-7 summarize the literature results reported in the papers.

Reviews and Surveys Category

In this category, 2 sets of reviews were identified. In the first set, detailed discussion is presented about recent breakthrough techniques of GANs, their development, variations, and medical field applications. The second set shows the impact of deep learning on ophthalmology. In total, this category includes 29/59 (49%) papers.

For the first set, studies [32,65,66,68-70,95-98] provided detailed reviews about GANs including their basic background, theory, and implementations. Also, they present current research hotspots and proposed GANs in different applications. They provided the reader with a clear insight into GANs' advantages and disadvantages, its different evaluation metrics, and proposed a bright prospect of this technique. Studies [32,95] focused on the importance of GANs, especially in medical field

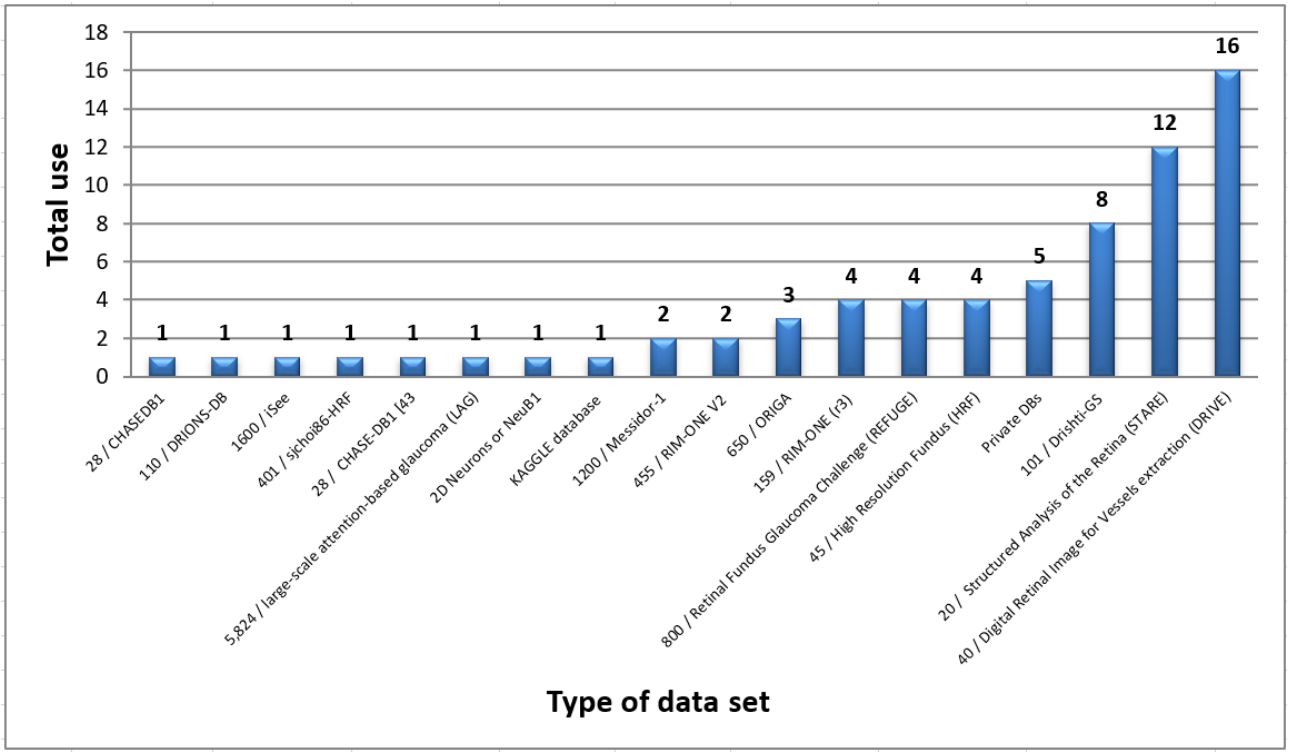
applications, and their capability to generate data through image synthesis technique without explicitly modeling the probability of density function. Wang et al [96] provided a further investigation of GAN in parallel intelligence. Another study, [99], discussed incorporating GANs in the signal processing community, showing different training methods, constructing GANs, and highlighting current challenges to their theories and applications. References [100,101] are practical prospective studies, and in [100], the authors tried to assess GAN algorithms and find the best architecture among all. However, they concluded that most of the models could achieve similar scores with enough hyperparameter optimization and random restarts. Additionally, they tried to overcome the limitation of evaluation metrics by computing precision and recall on several proposed data sets. Also, in [101], the authors reproduced the current state-of-the-art GANs, aiming to explore their landscape, discussing their pitfalls, and reproducibility issues. Turhan and Bilge [102] presented a comprehensive study about generative models such as GANs and autoencoders (AEs) and identified the relationship among them for better understanding and emphasizing on the importance of generative models. Oussidi and Elhassouny [103] proposed a starting point survey for those who have interests in deep generative models such as deep belief networks (DBNs), deep Boltzmann machine (DBM), restricted Boltzmann machines (RBMs), VAE, and GAN. They explained their building blocks, learning procedures, and limitations.

In the second set of articles, [52,54,72,104] presented an overview of DL applications in ophthalmic disorder using digital fundus images. They summarized the publicly available data sets used for different retinal diseases such as cataracts, retinopathy, glaucoma, and age-related macular degeneration. They also provided a detailed summary of the pros and cons of this emerging technique for both computer scientists and ophthalmologists and specified the clinical and technical aspects to address deep learning challenges and future directions. Some studies [56,105,106] discussed the importance of clinical considerations and potential challenges for clinical adoption and telemedicine integration to reduce cost, increase accuracy, and facilitate health care accessibility. Ting et al [53] described the importance of deploying deep learning algorithms within clinical settings. Hogarty et al [55] clarified the misunderstanding between ML and deep learning terms and presented an overview of AI and its development in the ophthalmology field. Mayro et al [107] also provided an overview of AI and deep learning DL applications in glaucoma detection using fundus images, optical coherence tomography, and visual field interpretation.

Other studies, [20,108], followed the systematic framework in their reviews: [20] discussed the main algorithms used for glaucoma detection using ML, indicating the importance of this technology from a medical aspect, especially retinal image processing, whereas [108] performed a systematic review on investigating and evaluating DL methods' performance for automatically detecting glaucoma using fundus images.

Figure 7 illustrates the publicly available data sets, their sizes, and how often researchers used them. Each data set is collected using a particular camera with different standards and used for a specific disease type. Thus, generalization is the key problem of DL approaches as described in the "Challenges" section.

Figure 7. Total use of various datasets in glaucoma screening.

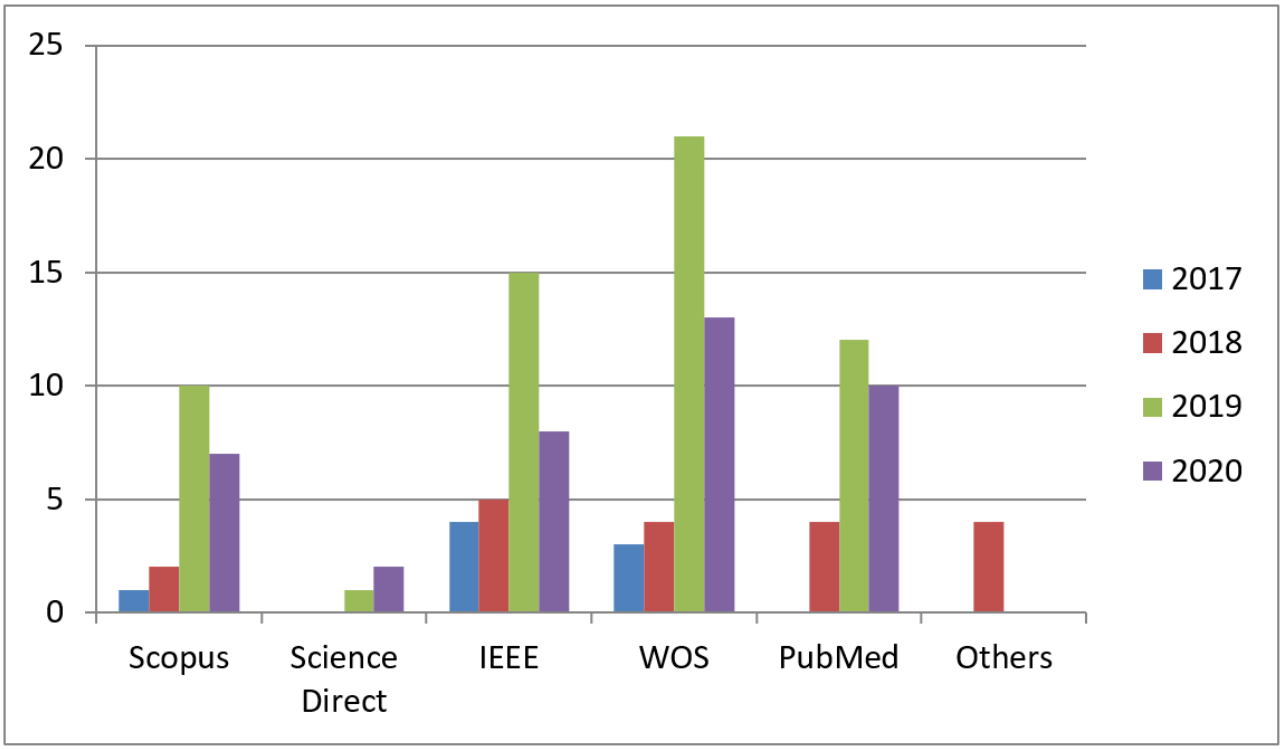


As Figure 7 illustrates, DRIVE and STARE are the most frequently used data sets. In other words, researchers often rely on BV segmentation in the diagnosing process [72]. However, few researchers have used Messidor-1, high-resolution fundus, 2D Neurons(NeuB1), and CHASEDB. For OD and OC landmarks segmentation, DRIONS-DB, retinal fundus glaucoma

challenge (REFUGE), ORIGA, RIM-ONE (r3/v2), and Drishti-GS were the most used, while seldom used is the large-scale attention-based glaucoma (LAG) data set, which is for RNFL and RL landmarks segmentation.

Figure 8 shows the distribution of the collected papers per year regardless of their duplications. The statistics in Figure 8 indicates the recent interest of researchers to adopt GANs techniques. Furthermore, it reveals the need to explore this newly emerging technique in ophthalmology. Therefore, extensive further work is needed to cover this area of research.

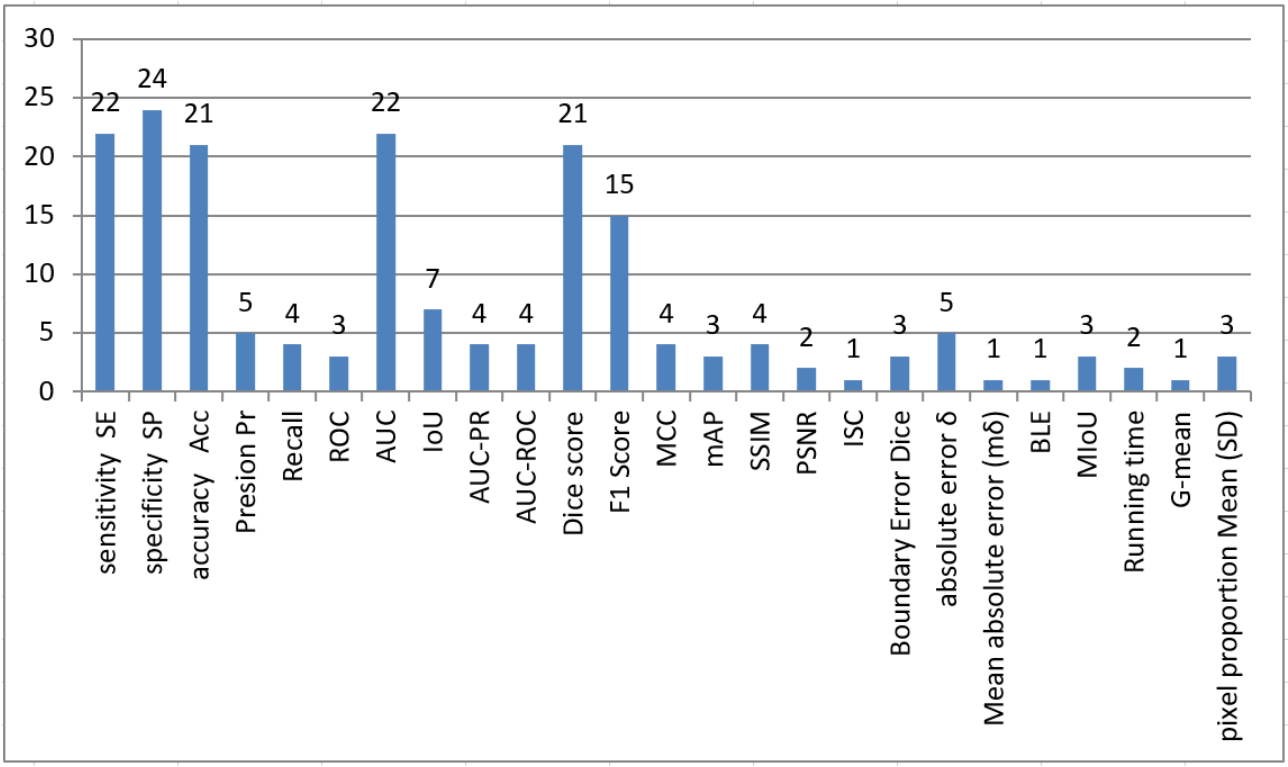
Figure 8. Distribution of papers per libraries. WOS: Web of Science.



This work has targeted 5 search engines: Scopus, ScienceDirect, Web of Science, IEEE, and PubMed, which are highly reputed and reliable resources for research. They include studies on implementation of deep learning techniques for different retinal disorder fields to help ophthalmologists and patients. Journal articles comprised 36 papers and only 23 were published in conferences.

According to Multimedia Appendices 1-7, each paper has used a different set of evaluation metrics; thus, we concur with Yu et al [46] in concluding that there are no uniform evaluation indexes in the literature to evaluate synthetic and real images. To further clarify this issue, Figure 9 shows the distribution of evaluation metrics used in the collected papers. To present Multimedia Appendices 1-7 visually, a T-shaped matrix diagram in Multimedia Appendix 8 illustrates in the upper part, named "Metrics used," the total use of each metric in all articles according to the used data set. Similarly, in the lower part, named "Task," the diagram shows the total use of each data set in all articles according to a specific task (classification, segmentation, or synthesizing).

Figure 9. Distribution of frequently used evaluation metrics in glaucoma screening. AUC: area under the curve; BLE: Boundary Distance Localization Error; IoU: Intersection over Union; ISC: Image Structure Clustering; MCC: Matthews correlation coefficient; mAP: Mean Average Precision; MIoU: Mean Intersection over Union; PSNR: peak signal-to-noise ratio; ROC: receiver operating characteristic curve; SSIM: structural index similarity.



Based on the observations of the upper part of the diagram, the top 5 metrics (sensitivity, specificity, accuracy, area under the curve [AUC], F1 score, and Dice Co.) were used the most with various data sets. Furthermore, 87% (13/15) of metrics were mainly performed on STARE and DRIVE data sets, unlike other data sets, such as Rim-ONEv3 and Drishti-GS, that use another set of metrics (eg, F1 score, Dice Co, peak signal-to-noise ratio, structural index similarity, and δ) to evaluate the performance. This indicates the need to consider standard effect metrics in future research irrespective of the type of data set used.

By contrast, in the lower part, the segmentation task was reported as the most applied task in the collected articles, followed by images synthesizing, with the classification task being the seldom applied. Nevertheless, the best results reported were in a classification study by Bisneto et al [81], which utilized a combination of Dristh-GS and RIM-ONE data sets. They achieved 100% in sensitivity, specificity, accuracy, and AUC in OD/OC classification. Their method was based on cGANs with taxonomic diversity and distinction indexes. Although most of the studies are on segmentation tasks and professionally segmented BVs and ODs [42,78,80,91], they still lack segmenting fine and small vessels and suffer from false positives. By contrast, images synthesizing attracted increased interest of researchers, as it assists in overcoming the shortage of medical images. Some researchers, such as [58,87], have used GAN with adversarial AE to enhance the generated image and some others tend to rely on using different loss functions to better train G and D networks. However, generated images are blurry, noisy, and of low quality with lack of details. Other studies, such as [76,88,89,94], adopted preprocessing (eg, data augmentation, localization of ROI, automatic color equalization) and postprocessing (eg, Lanczos resampling method, morphological operation, contrast enhancement) to enhance the performance of their methods, and they experienced a further improvement in their segmentation result.

Discussion

Principal Findings

This study aimed to provide a detailed summary of the literature on retinal disease detection or segmentation, particularly glaucoma, using GANs and highlight the recent trends exhibited by researchers on this topic. We mainly focused on articles that worked on enhancing the segmentation or detection of the disease rather than improving GAN techniques. Furthermore, we provide a taxonomy of papers related to this area to further assist future research.

Several benefits may arise from our taxonomy. First, organizing tens of papers in a single diagram provides better understanding of literature work, as people with less experience may be confused if many papers remain unorganized. Second, the taxonomy helps sort literature works and activities into meaningful, easy to manage, and coherent frameworks. Third, it provides researchers with better insights into a given theme, thus finding current literature gaps and discovering new research directions. Last and most importantly, it helps highlighting articles' strengths and weaknesses of a particular research scope.

From the developed taxonomy, we can quickly see that all the published papers followed the direct method of the GAN architecture; hence, there is an urgent need to discover the impact of the hierarchical or iterative method on glaucoma screening. Moreover, almost all of the researchers worked on

BV segmentation, and very few used OD and OC segmentation, which are the most reliable indications of glaucoma according to ophthalmologists. Future GAN research should focus on disease classification rather than on the segmentation of retinal anatomy. Most of the literature studies faced difficulties in terms of the early detection of glaucoma and low segmentation of fine vessels; therefore, alternatives should be developed, for example, using the RNFL to indicate the early presence of the disease or exploiting the prior knowledge of vascular connectivity to improve upon the segmentation performance of the current methods. Although the RNFL is a good sign for early glaucoma screening and has been incorporated as one of the gold standards of glaucoma evaluation [109], very few studies utilized the RNFL with GANs. OD/OC segmentation may lead to interference with pathological aspects such as large genetic OD sizes. Based on the reviewed papers, we noticed that only one article [1] has used RNFL for glaucoma screening. Although that study achieved impressive results, the authors used a private data set.

Most of the previous studies concentrated on the segmentation task. As much as 17/30 papers worked on retinal landmark segmentation [1-17], while only 2 papers worked on disease classification [18,19], and 5 papers worked on image synthesis to address the lack of medical images [20-24]. However, the rest of the papers (6/30) performed multiple tasks (eg, segmentation and classification, synthesis and segmentation) [25-30]. In conclusion, more than 50% (17/30) of the literature worked on segmentation task and few researchers have worked on classification and synthesizing retinal images. Therefore, future studies should take these statistics into considerations.

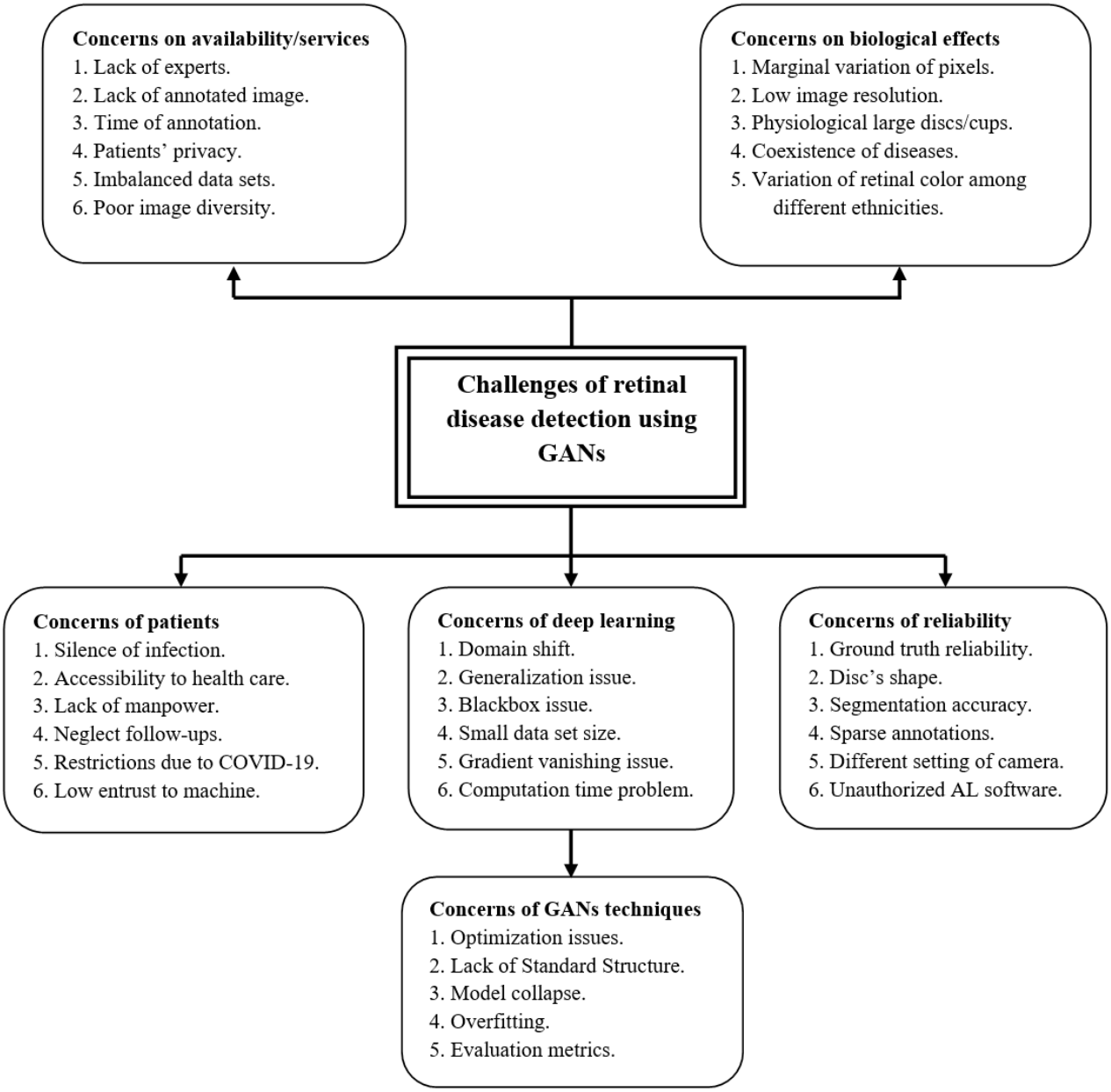
In the following sections, the included papers will be discussed in detail. We present comprehensive diagrams showing the factors that motivate researchers to carry out their work in this area, highlighting their encountered challenges, and summarizing significant recommendations for addressing their faults in future work.

Challenges

Overview

Glaucoma is a serious disease. Therefore, researchers and developers attempt to exploit the magic of DL technique to help doctors and patients diagnose the disease at its early stage. However, various challenges hinder their expectations; some of those challenges implicitly exist in the nature of DLMs, or are somehow incorporated within DLMs (eg, data richness, diversity of data, and powerful hardware), besides the challenges of GANs architectures (eg, model collapse, optimization, Nash equilibrium, and evaluation metrics). All these challenges have been summarized and discussed in this section along with their relevant references to provide the readers with direct access to the original papers for further discussion. Figure 10 categorizes literature challenges into 6 groups to further assist discussion. Each group is indicated with a separate shape.

Figure 10. Challenges of glaucoma screening using GANs technique. AL: artificial learning; GAN: generative adversarial network.



Challenges Related to Patients

The silent progress of glaucoma disease constitutes a crucial challenge worldwide. Half of the infected people do not experience any symptoms at early stages [5-7]. According to various studies, more than 60 million cases were diagnosed globally in 2013, and it is expected to exceed 75 million and 111 million cases by 2020 and 2040, respectively [9,10]. Especially among rural populations, China and India are considered to be the home to approximately 40% of glaucoma cases globally [110]. These populations, mostly in developing countries, suffer from difficulties in accessing medical centers, unavailability of experts, high costs of health care, and sustainability of health care services [111], in contrast to Western countries, where health care is cost-effective and different socioeconomic situations of patients are supported, and thus treatment for glaucoma remains affordable [106].

In addition, the recent pandemic, COVID-19, has enforced social distancing during communication. Therefore, there is a great need to promote ocular screening in conjunction with telemedicine as a remote monitoring tool [112], alongside the presence of handy cheap smartphones, whereby patients can collect their own IOP data themselves with accurate tonometers and free anesthesia [113]. Although DLMs positively affect both doctors and patients' style in terms of decision making, cost affordability, and health care accessibility, there remain some serious challenges, such as technical and clinical challenges, interpretation of the results, and patient trust in machines [112]. Zapata et al [114] predict that very soon AI will start assisting specialists in achieving high levels of consistency and accuracy beyond human abilities.

Challenges Related to Reliability

Reliability is a key to adopting computer technology in the medical field. Deep learning techniques may misclassify

segmenting some pixels due to low image contrast or heavy overlap between foreground and background pixels, leading to false-positive/false-negative result [57,81]. In some cases, doctors are dissatisfied with deep learning segmentation performance, as it is not as real as their expectations. Taking RNFL segmentation as an example, the segmentation results do not have specific geometrical shape of RNFLD as the gold standards and large segmentation errors of fundus images [83]. Furthermore, the variability of shape and extremely inhomogeneous OD structure appearance result in inaccurate CDR measurement compared with ideal ones [115-117]. In some cases, deep learning approaches neglect domain knowledge that doctors care about, such as CDR [118].

Existing methods often suffer from poor segmentation of the fine vessels [78,80] due to weak ability of antinoise interference or insufficient segmentation of vessels [49]; therefore, prior knowledge of BVs connectivity may improve the segmentation performance. Meanwhile, the low reliability of manual detection and the small size of public data sets increase the complexity of morphological assessment of nonglaucomatous optic neuropathy [119,120]. Robust ground truth labeling must be generated after a comprehensive evaluation, including structural imaging, clinical examination, and perimetry [121]. Doctors mostly decide the disease status. Although all clinical symptoms occur, it can lead to differences within annotators, and thus exaggerated annotations [52,106,122,123]. The reliability of glaucoma algorithms is restricted due to the lack of reference ground reality for glaucoma [115,124]. DLMs have a remarkable ability to address glaucoma. However, it is critical to have gold-standard algorithms for assessing and detecting glaucoma [54], as well as for editing or synthesizing images using the GAN techniques [97].

Sometimes, researchers tend to exclude low-quality or sparsely annotated images during the training phase; this kind of regime weakens the algorithm and leads to less reliability in real-life cases [111]. Furthermore, incorporating nonspecialists for image grading limits the reliability of identification [125]. Finally, although most of the reviewed papers have shown outstanding diagnostic performance, at times researchers do not mention some hyperparameter values used in the training stage, particularly when they use their own private dataset [112]. Excessive screening can result in overdiagnosis. DLMs could also be harmful if the diagnostic software is issued directly to patients, as future opportunities and risk of AI could be magnified [55].

Challenges Related to Biological Effects

Pathological change and image quality play a major role in the accuracy of glaucoma diagnosis [57,73,123]. Early and moderate glaucoma stages are considered one of the biggest challenges faced by ophthalmological practice due to the marginal variation size of CDR compared with normal eye [126]. Serener and Serte [127] have used ResNet-50 and GoogLeNet with transfer learning for early and advanced glaucoma detection, and found that GoogLeNet outperforms ResNet-50 with a trade-off performance between sensitivity and specificity. Besides, Bisneto et al [81] proposed GAN-based OD segmentation allied with an index of taxonomic diversity for extracting texture attributes aiming to detect early stages of glaucoma. They achieved outstanding results reaching up to 100% for accuracy and 1 for the receiver-operating characteristic curve. The misclassification of glaucoma and nonglaucoma is usually due to heavy overlap and extremely bad contrast between ocular structure and the background, leading to unsatisfied segmentation performance due to OC's undistinguishable boundaries [116]. Low-quality images (blurring and contrast) can result in unreliable model predictions. Furthermore, the lack of a clear OC border increases the misclassification rate [128].

There is a trade-off between image's quality and computational parameters of the network [129]. Therefore, the need for DLMs to downsample images into lower resolution (ie, 224 × 224) to reduce the computation time leads to reducing image contrast, and hence deteriorating key diagnostic parts of ocular images and weakening the capability to recover contextual information [86]. By contrast, performance of DLMs varied among ethnicities, for example, the Saudi population's performance is not the same as on Western populations. The differences among populations is due to the richness of melanocytes in the retinal pigmented epithelium of darkly skinned people compared with Whites [52]. Therefore, data sets used in glaucoma detection must follow specific standards to ensure heterogeneity and diversity of images.

Multiple eye disorders such as high myopia or pathologic are another major challenge leading to false-negative and false-positive results [54]. The main reason for the incorrect segmentation of glaucoma in myopia cases is the alteration of the macula and optic nerve appearance. In addition, the use of RNFL imaging for glaucoma diagnosis in patients with diabetes should be made carefully [130]. Myopia affects macular and RNFL thickness measurements due to the thinning and stretching of these layers caused by the increased axial length and optical projection artifact of the scanning region [131]. Myopia mostly causes misclassification of glaucoma due to its irregular ONH appearance [132]. In severe myopic cases, the color contrast between the foreground (OC) and the neuroretinal rim decreases due to an increased pallor in the rim. Furthermore, the increased pixels' values brighten the underlying peripapillary tissue and lead to difficult evaluation of the RNFL in the peripapillary area. In addition, torsion or tilting of the OD can occur, and the OD's rotation can result in an oblique view of the ONH [128].

In other cases, it is hard to distinguish between physiologic large cups and glaucomatous cases because both cases share a common feature (eg, large CDR) [117]. Diseases such as OD edema, OD hemorrhage, and glaucoma frequently make segmentation of OD rather difficult [133]. By contrast, retinal BV segmentation also has inherent challenges such as incorrect segmentation of pathological details and low microvascular segmentation [40].

Challenges Related to Availability/Services

Time, efforts, and lack of experts are the main challenges of medical care centers [88,134]. Therefore, computers have been increasingly used for automatic retinal segmentation to serve as a second opinion to the doctors, improve the diagnostic accuracy, and reduce the tedious work of annotating images

[43,46,135]. Particularly, GANs showed impressive performance in medical image synthesis and it is usually employed to tackle the shortage of annotated data or lack of experts [74,79,95]. Generally, medical images are usually rare, expensive, and full of patient privacy issues [51,88] and the publicly available data sets are often imbalanced in size and annotation [46,57,84]. In general, segmentation tasks suffer from an immense problem of class imbalance. Thus, the accuracy metric is not sufficient alone until concluding a system's efficiency on both sensitivity and specificity. They should, however, be considered as an essential evaluation metric [72].

Diaz-Pinto et al [90] proposed a GAN method with semisupervised learning to develop a good image synthesizer to tackle the shortage of retinal image availability and support generalization ability. Additionally, Liu et al [136] created a large-scale glaucoma diagnostic fundus images (FIGD) database. They proposed the glaucoma diagnosis with a complicated neural networks method for automatic detection of glaucomatous optic neuropathy. Importantly, the method has the potential to be generalized throughout populations.

Various GAN-based methods have been proposed to mitigate image labeling [43,50,51,75,87,92]. However, this challenge remained open as the current literature results are still inaccurate (eg, fail to generate very thin vessels). Lahiri et al [43] concluded that the diversity of annotated images is more important than the actual number of annotations. Finally, rural areas experience difficulties in locating ophthalmologists. This also necessitates more future work to use telemedicine in ophthalmology [55].

Challenges Related to the Nature of Deep Learning

With the recent advancements in DLM methodologies, promising results in the field of ophthalmology have been obtained. Many GANs and CNNs models are proposed in computer vision. However, DL approaches face several difficulties, such as domain shift.

Domain shift is the disparity in appearance distribution between various data sets due to different camera settings, illumination variation, different screening angles, or out-of-focus ROI. As a result, domain shift hinders the generalization capability of deep networks [89]. In most literature, training and test data sets come from the same image distribution. However, this is not always the case in real life. Therefore, it may significantly damage the real-life applications if not handled beforehand [72]. Kadambi et al [79] proposed an unsupervised domain adaptation framework by allowing the model to learn domain-invariant features to enhance segmentation performance and generalization capability. Wang et al [77] tried to align the distributions of the source and target domains so that the labeled source images can be used to enhance the classification efficiency of the target domain.

Deep learning addressed many issues in the traditional methods of ML. However, it also brought new difficulties. The most crucial issue is the ambiguity of the diagnosing result; in other words, the blackbox problem [53,56]. DLMs are blackbox in nature and do not have diagnostic explanations to confirm their effectiveness in a real clinical setting. Wang et al [73] proposed a pathology-aware visualization approach for feature visualization using DNNs to explain better how decisions are taken by computer, and therefore find pathological evidence through computer-aided diagnosis. Furthermore, for this purpose, Zhao et al [115] proposed a weakly supervised model due to its ability to simultaneously learn the clinical evidence identification and perform the segmentation task from large-scale weak-label data that further improves glaucoma diagnosis.

The lack of publicly available data sets for training the model is another significant challenge concerning deep learning approaches. Therefore, Orlando et al [132] proposed a data set named REFUGE, which contains 1200 fundus photographs with standard gold segmentations and clinical glaucoma marks. Moreover, Li et al [137] created the LAG database containing 11,760 fundus photographs classified as either positive glaucoma (4,878) or negative glaucoma (6,882), which is the largest among the currently existing databases. According to Asiri et al [52], the key problem of constructing a robust deep CNN method is not the availability of broad data sets but instead the diversity of annotation of those images [43]. A major difficulty of each algorithm is its validity in multiple patient cohorts with diverse conditions. Therefore, for a DLM to be sturdy, it must be effective across various data sets [105].

Recent studies demonstrated that more complicated and informative image features might be discovered when growing the depth of the network [138,139]. However, as the network depth rises, deeper CNN has poor diagnostic efficiency due to the gradient disappearance issue or the gradient explosion problem [88,140,141]. Researchers mostly use shortcut links (skip connections) that skip one or more layers while training deep networks, as was the case with [88,126,128,129]. Alternatively, in GANs techniques, using WGAN or LSGAN gives a smoother gradient that contributes to stable training [42,79]. Another concern that should be considered before building up deep models is the *computation time*. As there is a trade-off between model's depth and the efficiency, the deeper the architecture the greater the number of parameters it gets, which eventually increases computation time [140].

Challenges Related to GAN Technique

Despite all the ongoing developments and studies, GANs suffer from several challenges and weaknesses besides the challenges related to deep learning nature (eg, blackbox, generalization capability, computation time, and annotation cost). The most critical concern with GANs is the instability of the training process (Nash equilibrium point) [98,142]. Zhao et al [82] used the residual module that allowed easy optimization of competitive networks, while Tu et al [78] used WGAN-GP to alleviate training instability of the traditional GAN. Biswas et al [92] carefully adjusted hyperparameters to balance between the 2 networks (G and D). Park et al [94] improved learning performance and mitigated imbalanced learning by introducing new loss functions for the generator and re-designing the discriminator's network. However, it remains challenging to determine which algorithm works better than others or what modifications are critical to enhancing the results. Lucic et al [100] found that most models could achieve comparable scores

with appropriate hyperparameter optimization and random restarts. According to Kurach et al [101], the nonsaturating loss over data sets, architectures, and hyperparameters is sufficiently stable.

Besides, in GANs, the possibility of mode failure/collapse persists while training the model. Model collapse occurs when data generated from GANs mostly concentrate on very narrower modes (partial collapse) or 1 single mode (complete collapse) [68,99]. By contrast, if the discriminator becomes very strong during training, the generator gradient gradually decreases and eventually disappears. As a result, the generator learns nothing. The imbalance between generator and discriminator networks contributes to overfitting. Many approaches have been proposed to tackle these challenges; for example, Radford et al [59] aimed to address instability training issues, and Kadambi et al [79] created a new adversarial domain adaptation architecture, led by Wasserstein for better stability and convergence.

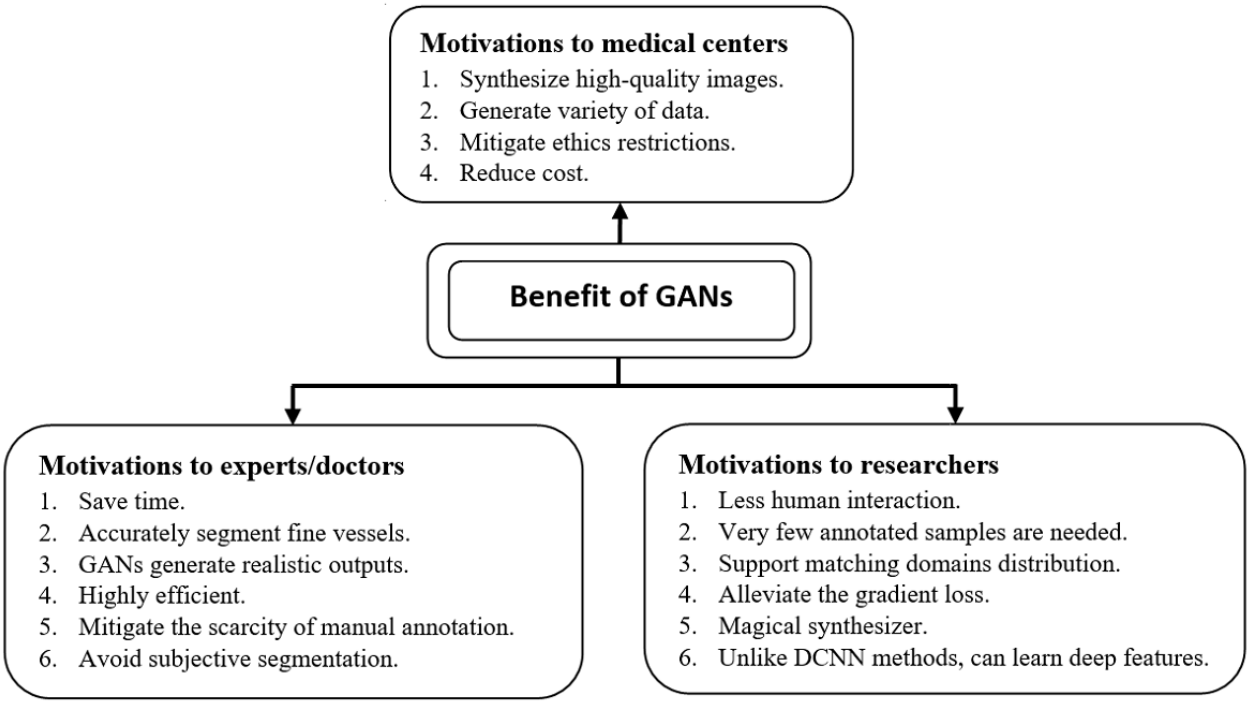
The lack of standard evaluation metrics is another big issue in GANs compared with other generative models. Inception score (IS), average log likelihood, Fréchet inception distance (FID), Wasserstein metric, etc. are quantitative measurements of GANs. There is no majority vote on which assessing measurement is the best. Different scores rely on various aspects of image generation. However, some measurements seem more plausible than others (eg, FID is more durable to noise). FID can compare the similarity between real and generated images [143], which is considered more effective than IS [70].

In conclusion, the main causes of GAN problems can be summarized as follows: (1) The distance calculation of the corresponding optimization (such as Kullback–Leibler divergence and Jensen–Shannon divergence) is unreasonable. (2) It is difficult to overlap the generated distribution with real distribution. Although the GAN technique is a new, interesting, and attractive field of study in many applications, further studies are needed to resolve the uniqueness of generated samples, poor convergence, and complete model collapse challenges.

Motivations

Adopting deep GAN in ophthalmology is a promising and significant field of study. This section reports some of the literature's characteristics, which we classified on the basis of references to support further discussion (Figure 11).

Figure 11. Benefits of GANs-based methods for glaucoma screening. DCNN: deep convolutional neural network; GAN: generative adversarial network.



Motivations Related to Experts/Doctors

Detection of any retinal defects must be through analysis of ocular images. Analysis of retinal images, however, must involve trained physicians to analyze and assess digital color fundus images. Such a process requires a great deal of time and human work; therefore, GANs support doctors in mitigating this extensive bottleneck [50,51,91]. Furthermore, deep GANs techniques are unlike CNNs, where the same GAN approach could be applied to a wide variety of cases and still produce reasonable results [45]. GANs can detect the OD in fundus photos with pathological changes or irregular highlights [57,86]. In the case of vessel segmentation with CNN-based methods, outputs are usually blurry around small and weak branches or suffer from a problem of nonconnectivity of segmented vessels; however, GANs better segment capillary/thin vessels of fundus images [76,80,84], and thus serve as a second opinion to ophthalmologists [72]. GANs are the framework that allows to create and use practical outputs as a gold standard [44]. Therefore, these frameworks were adopted by Lu et al [83] due to their ability to generate the required specific geometry of RNFLD, which is close to ground truth with high precisions, accuracy, and fewer segmentation errors, despite the existence of multiple pieces of RNFL or low-contrast images. Thus, its

segmentation results are much more trusted by doctors than CNN's.

Adversarial learning avoids scarcity of manual annotation and subjective segmentation made by non-expert clinicians as this methodology is mainly data driven [72,85]. In glaucoma classification, enforcing GANs to synthesize images with similar visualization results as the reference image will help mitigate the drawbacks of binary labels (negative or positive) that limit the visualization methods to recognize pathological facts underlying diagnosis by DNNs [73].

Motivations Related to Researchers

Deep learning in retina images is very effective and useful [72]. However, they are often affected by domain shifts across data sets. As a result, a generalization of DLMs was severely hindered. Therefore, researchers tend to exploit generative adversarial learning for domain adaptation by encouraging the target domain predictions to be close to the source ones [79,89]. Domain adaptation is often used to overcome the lack of large pixel annotation using off-the-shelf annotated images from other relevant domains. Alternatively, researchers exploit the existence of a large amount of unlabeled data to train a classifier using the power of DCGAN in a semisupervised learning scenario [90]. Semisupervised learning is in the middle way between unsupervised and supervised learning; therefore, less human intervention is required when combined with GANs for better semantic segmentation [74]. Using GANs techniques, Lahiri et al [43] performed image segmentation with very few annotated samples (0.8%-1.6%), nearly 500-1000 annotations. Further, Zhao et al [93] proposed an image synthesizer using GANs with style transfer and then integrated the outputs into the training stage to boost segmentation efficiency using just 10 samples.

With deep adversarial learning, researchers aim to reduce domain discrepancy [144,145] by improving the quality of the generated outputs to be as close as possible as the inputs. Wang et al [77] exploited label information for matching domain distribution. Ma et al [42] applied the least-squares loss function instead of sigmoid cross-entropy to generate images with distribution close to the real ones and also alleviate gradient vanishing problems. Furthermore, Liu et al [57] added a patch-level adversarial network to enhance image consistency between ground truth and the generated samples, which further boosts segmentation performance.

GANs are capable of learning the mapping from the input image to the output image as well as learning a loss function to train this mapping [45], unlike existing DLMs, which use a unified loss function for retinal vessels segmentation, thereby producing blurry outputs with false positives around faint and tiny vessels [84], which is in contrast to GAN variations (eg, WGAN-GP and M-GAN) that provide accurate segmentation results around small and weak branches [78], reduce low microvascular segmentation [94], and preserve the connectivity of arteriovenous vessels [76]. Moreover, AEs and GANs in a single system facilitate generating vessel maps without the previous existence of retinal vessel tree [87]. Besides, unconditional GANs can synthesize retinal images without using prior vessel images [92].

Although researchers recommend using DCNN for efficient segmentation tasks [146], the existing limitations of DCNNs are insufficiency of feature extraction, weak generalization capability, and poor capability to recover low-context information, unlike GANs, which are used to alleviate these problems as in Jiang et al [86], who proposed GAN with transfer learning, data augmentation, and skip connection concepts to overcome these challenges. Bisneto et al [81] impressively improved glaucoma segmentation and classification results using GANs allied with texture attributes identified by taxonomic diversity indexes. They achieved promising results (sensitivity, specificity, and accuracy of up to 100%).

For optimizing network complexity, Wu et al [49] applied the attention Gates technique in a standard GAN to encourage the propagation of features, promote reuse of features, and greatly reduce network parameters when paired with DenseNet instead of conversion layer. Alternatively, using dilated convolutions in the generative networks effectively expands the generator's receptive field without the number of calculations [82]. Adversarial training has been shown to improve the long-range spatial label interaction without expanding the segmentation network's complexity [147].

Motivations Related to Medical Centers

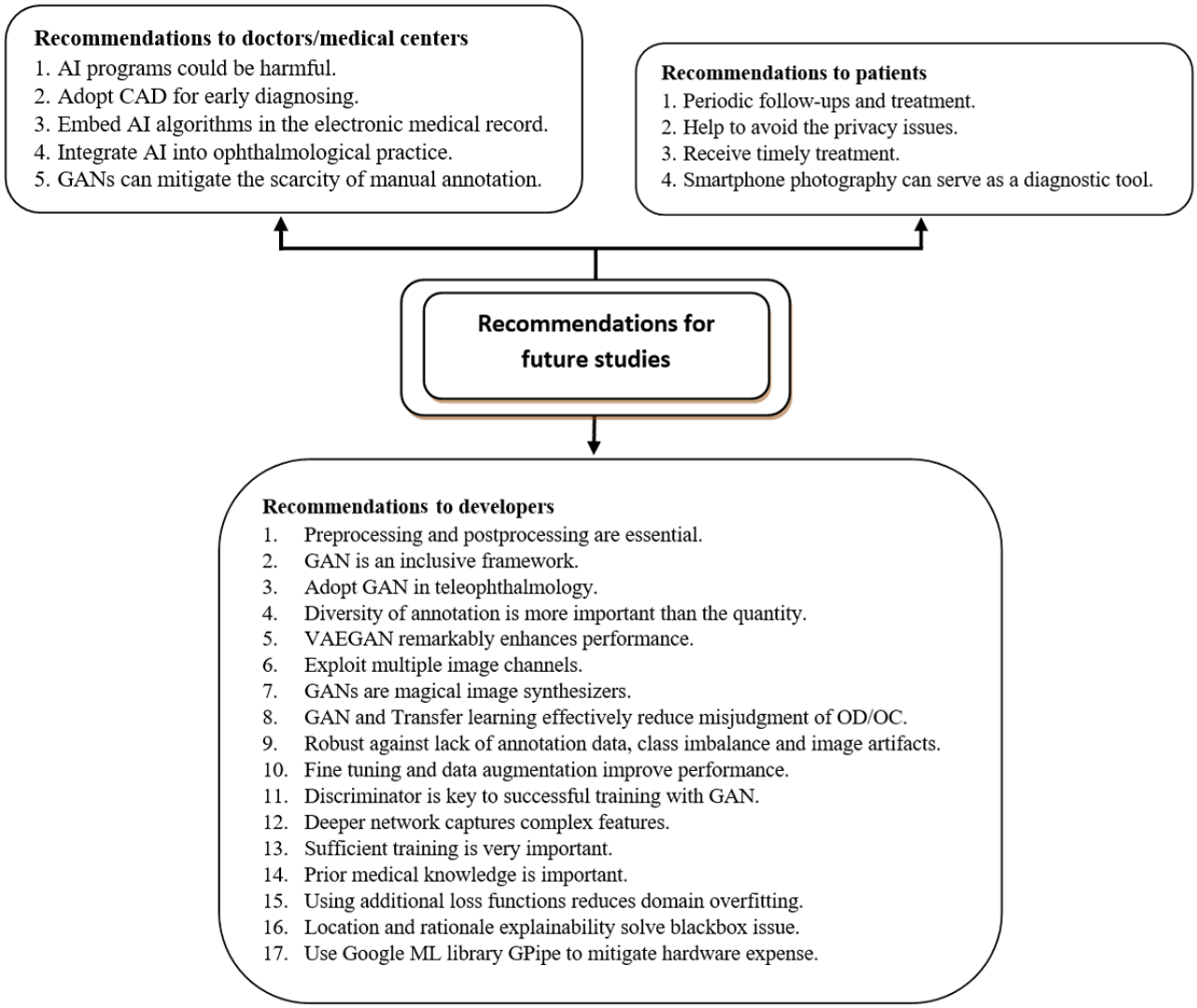
We think the best medical treatment is achieved when the doctor–patient relationship is built on honesty and concern. DL cannot substitute real relationships, but can complement them [104]. GAN architectures are versatile. For various training samples, the objective feature can be re-designed and more free model designs can be used [98]. The extraordinary feature of GANs in the medical field is synthesizing high-quality images with global consistency(eg, color consistency and both BV and OD occupy the same proportional area as the real images) [58,92]. Bisneto et al [81] proposed a method that learns the mapping function between retinal landmarks (BV, OD, and OC) and synthesizes images using the 3 channels (RGB). Furthermore, the method exploits the merit of a large receptive field of GANs to generate good segmentation results [82].

Incorporating GAN techniques in the medical field helps enrich health care centers with various data and effectively solves data imbalance problem [87,134]. As a result, this feature facilitates solving ethical issues surrounding patients' privacy [72], saves memory and time needed to collecting images [79], reduces costs [88], and saturates the nature of data-hungry DLMs [51].

Recommendation

In this section, we briefly include guidelines from the literature to alleviate existing challenges faced by researchers, doctors, medical centers, and patients, as well as present ways to achieve a correct diagnosis of retinal defects (Figure 12).

Figure 12. Recommendations of using GANs-based methods in glaucoma screening. AI: artificial intelligence; CAD: computed-aided design; GAN: generative adversarial network; OC: optic cup; OD: optic disc; VAEGAN: variational autoencoder with GAN.



Recommendations to Doctors and Medical Centers

Higher-image resolutions significantly improve performance of GANs [87,148]. The key factor in obtaining GAN's high-quality synthetic outputs is the high-resolution paired images and the architecture of the generator [46]. Moreover, annotation variety is more important than the actual number of annotations [43]. Therefore, doctors must develop a public data set with high-resolution images that meet the quality assessment system [105]. Furthermore, it must be accessible and include multiethnicities to ensure generalization capability [108]. Besides, experts must validate deep learning models on the sizable heterogeneous population under different conditions [52], as direct release of DL application without prior checking could be harmful [55].

To improve public health, reduce health care costs, and enhance patients' perception, doctors shall adopt DL techniques in the medical field to tackle these challenges [53]. Adopting deep learning applications in magnetic resonance imaging and X-ray image processing is an interesting area of research [93]. All glaucoma studies emphasized the importance of CAD programs for early disease detection and for improvement of screening reliability [20].

In the future, GANs may be utilized to speed up AI development and application, allowing AI to comprehend and explore the environment [66]. Innovative and radical solutions for the health care system must be improved alongside glaucoma screening [106]. Significant improvements in instrumentation and interpretation can lower the cost of glaucoma screening in the future. Embedding glaucoma AI algorithms in the electronic medical record will improve outpatient management [107]. However, it is up to the physicians to lead the way in deciding how to incorporate AI in a new era of glaucoma management.

Automated retinal imaging technologies can reduce barriers to access and monitoring of the health system. Thus, AI integration into ophthalmology can improve patient care [56], help clinicians focus on patient relationships, and enhance health services [104], all of which can decrease irreversible blindness [54]. GANs can reduce the scarcity of manual data annotation and also be used as a clinical support tool [72].

Recommendations to Developers

A CNN in a generative learner is used for image segmentation tasks and obtaining successful outcomes [149]. GAN is an inclusive system that can be combined with various deep learning models to address problems that conventional ML algorithms cannot solve, such as poor quality of outputs, insufficient training samples, and deep feature extraction [68]. Furthermore, it outperforms conventional methods in editing and synthesizing image [69]. GAN allied with transfer learning can effectively reduce misjudgment of OD/OC in glaucoma cases and improve accuracy and generalization capability; however, better backbone network and different upsampling methods are required to improve performance [86] and exploring other downstream tasks may enhance the model's performance [93]. Although there is a vast increase of GAN applications, further studies are required to improve its efficiency and performance [70]. Incorporating spatial information, attention-based information, feature-maps information, and image channels (RGB) to improve network performance is a current research trend [140].

GANs can generate samples with distribution close to real data. Thus, they can be used in a systematic study of parallel systems [96]. GANs or its variants remain the future trends for mitigating imbalanced learning through generating samples close to real data, or enhancing model performance when combined with VAEs [58,65,72]. Thus, it is used as a sophisticated data augmentation technique to generate heterogeneous samples and ensure prognostic characteristics of images [52].

To date, only a few studies have experienced AI technologies in teleophthalmology [56]; photography using smartphones can be used as a diagnostic tool for ocular diseases [105]. Nowadays, there is a great need for remote disease monitoring and screening [107], especially during the COVID-19 pandemic and the vast infection transmission [150]. Thus, future study should emphasize on deep learning and telemedicine/teleretinal as potential gamechangers in the eye-care field [106].

Wang et al [89] proposed a very lightweight network architecture for joint OD and OC segmentation based on the MobileNetV2 backbone, which has few parameters and half testing time compared with the XCeption backbone, which promotes the network as a mobile app for glaucoma detection. Bisneto et al [81] presented GAN and texture features for automatic detection of glaucoma, and achieved impressive results that reached up to 100% for sensitivity, specificity, and accuracy. The authors indicated a proposal to transfer their method into a mobile app in a future study.

Future research should emphasize GANs and semisupervised learning for image synthesizing, aiming to improve the classification accuracy and the quality of the generated images simultaneously [43,75,90]. Adopting GANs in the medical field remains in its infancy, with no breakthrough application yet clinically implemented for GAN-based approaches [95]. For better feature extraction, researchers must exploit full feature information on RBG channels, spatial structure, and geometry of landmarks [83]. Semantic segmentation may reduce manual labeling effort [50,74] and enhance model performance when incorporated with WGAN domain adaptation [79]. In ophthalmology diagnosis, adversarial domain adaptation can be an important and effective direction for future research [72,88,151]. In addition, exploring the relationship between the quality of the generated image and the performance of the CAD system is needed [46].

With the envision to improve deep learning performance, preprocessing and postprocessing are essential for accurate segmentation [52,80,94]. Barros et al [20] concluded that data set size has a huge impact on the results. However, Lahiri et al [43] amazingly demonstrated that annotation diversity is more important than annotation count. GAN can make use of large amounts of unlabeled data [66,87].

Regarding GAN evaluation metrics, future studies should focus on more objective and systematic evaluation methods. However, further FID examination is required [100]. Developing quantitative assessment metrics thus remains a crucial research direction [152,153]. Researchers should evaluate their segmentation performance on public data sets [74] with heterogeneous and multimodal designs using less data-hungry algorithms [105]. In addition, the performance other classifiers (eg, XGBoost) and other cGAN architectures should be examined for faster and more accurate learning [81].

For glaucoma diagnosis, CDR and ISNT metrics present substantial information to be assessed [20]. More studies are needed to assess the validity of ophthalmology applications to detect AMD, diabetic retinopathy, and glaucoma in terms of accuracy, sensitivity, and specificity [54]. AUC, sensitivity, and specificity should be included in AI studies as the bare minimum [53].

Moreover, future research may utilize fine-tuning and data augmentation techniques to effectively improve model performance [52,81,86] and increase data set size for better training, and thus, synthesizing better classifiers [77]. GANs strength lies in its discriminator [32,80]. Duplicating the generator's structure improves robustness [94]. Adding more network layers help capture more in-depth features [82]. Training and optimizing the model remain critical [84,87], with regard to careful balancing between G and D [92]. Patch-based images should be used as input for both G and D [51,84]. U-GAN instead of U-Net should be used to improve the model's performance [49]. Additionally, exploiting previous knowledge of vessel structure [78,80,92] is critical for accurate segmentation [91]. Objective function supported with various loss functions may enhance model performance [84]; for example, WGAN-GP can avoid gradient disappearing and enhance training [92], Dice coefficient loss function for segmenting hard images [57], and least-squares loss function with dilated convolution can enhance small vessel segmentation [42]. On top of that, topological structure loss can enhance the connectivity of A/V classification [76], whereas binary cross-entropy loss function with false-negative loss function can improve training efficiency and increase segmentation robustness [94]. Furthermore, an adversarial loss can reduce the domain overfitting [154], and Wasserstein distance is preferable for domain adaptation, as it decreases the probability of mode collapse and avoids the gradient vanishing [79]. Weight normalization along with average pooling is the best design

setting when structured prediction is used with U-Net [43]. Exploring a combination of different styles instead of training dedicated models for a particular style is necessary [93]. MISH is a modern activation function that presented better results than ReLU on most current benchmark data sets [155].

To date, explainable DLMs for glaucoma screening utilizing retinal fundus images have not been proposed [156]. Researchers should focus on relational and locational explanation using saliency maps, heatmaps, or other invented methods to provide plausible explanations of DL decisions.

Lastly, future research should incorporate the distributed ML library GPipe proposed by Google [157] to mitigate hardware limitations. This may help train large-sized models and enhance performance without tuning hyperparameters [140].

Recommendations to Patients

Increasing the amount of data using a successful GANs synthesizer significantly saves the privacy of patients [72]. Good DLM offers timely treatment by providing wealthy information regarding patients' eye conditions [49]. In the near future, AI can support telemedicine platforms by facilitating the self-monitoring by patients through home-based diagnosis [56]. The availability of cheap, handy smartphones may also assist as a remote diagnostic tool [105]. This eventually could improve patient's perception and satisfactions [53], as well as encourage continuous follow-up and treatment [106].

New Direction of DL

Recently, DLMs have achieved positive retinal disease identification and segmentation outcomes. These technologies can revolutionize our way of life, and, probably in the next few decades, the field of medicine will change rapidly [53]. However, these techniques involve expensive hardware (eg, GPU requirements) and are greedy for images by nature. Thus, more advanced data augmentation techniques must be introduced to create heterogeneous samples while preserving the prognostic features of fundus images. A possible approach in this regard is to explore GANs [52,158]. Building systematic deep learning models trained on heterogeneous and multimodal data with fewer data-hungry algorithms can boost the effectiveness of AI in clinical settings [105]. Additionally, AI algorithms should be incorporated into electronic medical records to promote outpatient management, which is another fascinating subject [107].

From the viewpoints of accessibility, cost-effectiveness, and health care protection, there is a tremendous need to promote remote glaucoma monitoring in developed countries and rural communities, allow patients with glaucoma to obtain their own IOP data with anesthesia-free and reliable tonometers [113], and enable home-based evaluation and disease control (eg, rendering home tonometry accessible at a lower cost). Most importantly, within the current situation of the COVID-19 pandemic, new directions for DLMs can be implemented via teleretinal screening apps in ophthalmic settings to maintain maximum protection for both physicians and patients at a lower cost.

Improving the quality of diagnosis in terms of class imbalance, refining the training phases of GANs, and enhancing the computation time to better diagnose glaucoma variants remain obstacles [20,52,100]. Furthermore, it is necessary to note that GANs have not been used to diagnose difficult retinal disease to date, and GAN evaluation metrics are yet another challenging path of study [68].

Finally, combining GANs with other approaches is another prospective research approach; for example, the fusion of GANs with reinforcement learning, function learning, or conventional learning to create new AI applications and facilitate the advancement of these methods is also worth investigating [66,98].

Limitations of the Study

This most important limitation of our analysis is the number and identification of the source databases; however, the selected works form a reasonable and broadly representative selection of the chosen sources. Furthermore, the exclusion of other retinal diseases besides glaucoma, due to its severity worldwide, is considered another limitation. In addition, a quick view of the research activities on this critical retinal disease and GANs does not necessarily reflect the research community's response.

Conclusion

Providing adequate health services to people with retinal disorders has been a global issue. Studies are still ongoing to diagnose retinal disorders using deep learning; however, papers adopting GANs for glaucoma detection are not as abundant as those utilizing DL or ML methods. Consequently, insights into this emerging area are needed. Six papers [18,19,26-29] have worked on glaucoma classification–based GANs, and the majority tended to use GANs for segmentation or synthesizing retinal images.

The contribution of this study lies in analyzing and taxonomizing literature works in the field of glaucoma detection using GAN-based methods. To the best of our knowledge, all the previous studies generally discussed AL or DL effects on retinal diseases, and none particularly surveyed GANs for glaucoma detection. This makes our work first to address this emerging technique.

According to our taxonomy, the majority of the collected papers paid more attention to single landmark segmentation (eg, BVs) than to the segmentation of multiple landmarks. Some techniques were of tremendous or little interest (eg, the DCGAN and cGANs). Researchers worked in this field, identified their difficulties, and suggested recommendations to overcome the current and expected challenges. Other studies focused on improving GAN architectures rather than adopting them for diagnosis. To date, there has been no specific work adopting a GAN as a smartphone app or in telemedicine. Therefore, filling this gap is important for both patients and physicians to ensure fewer physical meetings during the global COVID-19 pandemic. Furthermore, new directions in this field have been explained.

Acknowledgments

A long-term research grant scheme (LRGS/1/2019/UKM-UKM/2/7) supports this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Segmentation papers.
[XLSX File (Microsoft Excel File), 17 KB - jmir_v23i9e27414_app1.xlsx]

Multimedia Appendix 2
Segmentation and classification papers.
[DOCX File , 20 KB - jmir_v23i9e27414_app2.docx]

Multimedia Appendix 3
Classification papers.
[DOCX File , 21 KB - jmir_v23i9e27414_app3.docx]

Multimedia Appendix 4
Synthesizing images papers.
[DOCX File , 23 KB - jmir_v23i9e27414_app4.docx]

Multimedia Appendix 5
Synthesizing, segmentation, and classification paper.
[DOCX File , 20 KB - jmir_v23i9e27414_app5.docx]

Multimedia Appendix 6
Synthesizing and classification of papers.
[DOCX File , 21 KB - jmir_v23i9e27414_app6.docx]

Multimedia Appendix 7
Segmentation and synthesizing images papers.
[DOCX File , 20 KB - jmir_v23i9e27414_app7.docx]

Multimedia Appendix 8
T-shaped matrix diagram.
[DOCX File , 135 KB - jmir_v23i9e27414_app8.docx]

References

1. Flaxman SR, Bourne RRA, Resnikoff S, Ackland P, Braithwaite T, Cicinelli MV, Vision Loss Expert Group of the Global Burden of Disease Study. Global causes of blindness and distance vision impairment 1990-2020: a systematic review and meta-analysis. Lancet Glob Health 2017 Dec;5(12):e1221-e1234 [FREE Full text] [doi: 10.1016/S2214-109X(17)30393-5] [Medline: 29032195]
2. Pascolini D, Mariotti SP. Global estimates of visual impairment: 2010. Br J Ophthalmol 2012 May;96(5):614-618 [FREE Full text] [doi: 10.1136/bjophthalmol-2011-300539] [Medline: 22133988]
3. de Carvalho Junior ASV, Carvalho ED, de Carvalho Filho AO, de Sousa AD, Corrêa Silva A, Gattass M. Automatic methods for diagnosis of glaucoma using texture descriptors based on phylogenetic diversity. Computers & Electrical Engineering 2018 Oct;71:102-114 [FREE Full text] [doi: 10.1016/j.compeleceng.2018.07.028]
4. Wood K. Glaucoma: the silent thief of sight. The Lamp 1995 Sep;52(8):15 [FREE Full text] [Medline: 7500740]
5. Villain MA. [The epidemiology of glaucoma]. J Fr Ophtalmol 2005 Jun;28 Spec No 2:2S9-2S12 [FREE Full text] [Medline: 16208234]
6. Mitchell P, Cumming RG, Attebo K, Panchapakesan J. Prevalence of cataract in Australia: the Blue Mountains eye study. Ophthalmology 1997 Apr;104(4):581-588 [FREE Full text] [doi: 10.1016/s0161-6420(97)30266-8] [Medline: 9111249]
7. Michelson G, Wärntges S, Hornegger J, Lausen B. The papilla as screening parameter for early diagnosis of glaucoma. Dtsch Arztebl Int 2008 Aug;105(34-35):583-589 [FREE Full text] [doi: 10.3238/arztebl.2008.0583] [Medline: 19471619]

8. Costagliola C, dell'Omo R, Romano MR, Rinaldi M, Zeppa L, Parmeggiani F. Pharmacotherapy of intraocular pressure - part II. Carbonic anhydrase inhibitors, prostaglandin analogues and prostamides. Expert Opin Pharmacother 2009 Dec;10(17):2859-2870 [FREE Full text] [doi: 10.1517/14656560903300129] [Medline: 19929706]
9. Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. Br J Ophthalmol 2006 Mar;90(3):262-267 [FREE Full text] [doi: 10.1136/bjo.2005.081224] [Medline: 16488940]
10. Tham Y, Li X, Wong TY, Quigley HA, Aung T, Cheng C. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology 2014 Nov;121(11):2081-2090 [FREE Full text] [doi: 10.1016/j.ophtha.2014.05.013] [Medline: 24974815]
11. Singh SM, Yahya AN, Mohamad CM, C.A M. Risk Factors for Cataract: A Case Study at National University of Malaysia Hospital. Jurnal Sains Kesihatan Malaysia (Malaysian Journal of Health Sciences) 2012 Oct;4(1):85-98.
12. Kingman S. Glaucoma is second leading cause of blindness globally. Bull World Health Organ 2004 Nov;82(11):887-888 [FREE Full text] [Medline: 15640929]
13. Shen SY, Wong TY, Foster PJ, Loo J, Rosman M, Loon S, et al. The prevalence and types of glaucoma in malay people: the Singapore Malay eye study. Invest Ophthalmol Vis Sci 2008 Sep;49(9):3846-3851 [FREE Full text] [doi: 10.1167/iovs.08-1759] [Medline: 18441307]
14. Vijaya L, George R, Paul PG, Baskaran M, Arvind H, Raju P, et al. Prevalence of open-angle glaucoma in a rural south Indian population. Invest Ophthalmol Vis Sci 2005 Dec;46(12):4461-4467 [FREE Full text] [doi: 10.1167/iovs.04-1529] [Medline: 16303934]
15. Nazri MZA, Kurniawan R, Sheikh-Abdullah SNH, Othman ZA, Abdullah S. Bayesian network and Dempster-Shafer theory for early diagnosis of eye diseases. COMPUSOFT 2020 Apr 30;9(4):A.
16. Centre for Eye Research Australia. Tunnel Vision-The Economic Impact of Primary Open Angle Glaucoma. 2011 Jul 19. URL: http://www.icoph.org/resources/249/Tunnel-Vision-The-Economic-Impact-of-Primary-Open-Angle-Glaucoma.html [accessed 2020-06-25]
17. Garway-Heath DF, Hitchings RA. Quantitative evaluation of the optic nerve head in early glaucoma. Br J Ophthalmol 1998 Apr;82(4):352-361 [FREE Full text] [doi: 10.1136/bjo.82.4.352] [Medline: 9640180]
18. Thakur N, Juneja M. Survey of classification approaches for glaucoma diagnosis from retinal images. In: Advanced Computing and Communication Technologies. Advances in Intelligent Systems and Computing, vol 562. Singapore: Springer; Oct 25, 2017:91-99.
19. Guo J, Azzopardi G, Shi C, Jansonius NM, Petkov N. Automatic Determination of Vertical Cup-to-Disc Ratio in Retinal Fundus Images for Glaucoma Screening. IEEE Access 2019;7:8527-8541. [doi: 10.1109/ACCESS.2018.2890544]
20. Barros DMS, Moura JCC, Freire CR, Taleb AC, Valentim RAM, Morais PSG. Machine learning applied to retinal image processing for glaucoma detection: review and perspective. Biomed Eng Online 2020 Apr 15;19(1):20 [FREE Full text] [doi: 10.1186/s12938-020-00767-2] [Medline: 32293466]
21. Jiang Y, Duan L, Cheng J, Gu Z, Xia H, Fu H, et al. JointRCNN: A Region-Based Convolutional Neural Network for Optic Disc and Cup Segmentation. IEEE Trans Biomed Eng 2020 Feb;67(2):335-343. [doi: 10.1109/TBME.2019.2913211] [Medline: 31021760]
22. Cheng J, Liu J, Xu Y, Yin F, Wong DWK, Tan N, et al. Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. IEEE Trans Med Imaging 2013 Jun;32(6):1019-1032. [doi: 10.1109/TMI.2013.2247770] [Medline: 23434609]
23. Abràmoff MD, Garvin MK, Sonka M. Retinal imaging and image analysis. IEEE Rev Biomed Eng 2010;3:169-208 [FREE Full text] [doi: 10.1109/RBME.2010.2084567] [Medline: 22275207]
24. Akram MU, Tariq A, Khalid S, Javed MY, Abbas S, Yasin UU. Glaucoma detection using novel optic disc localization, hybrid feature set and classification techniques. Australas Phys Eng Sci Med 2015 Dec;38(4):643-655. [doi: 10.1007/s13246-015-0377-y] [Medline: 26399880]
25. Wong DK, Liu J, Lim JH, Jia X, Yin F, Li H, et al. Level-set based automatic cup-to-disc ratio determination using retinal fundus images in ARGALI. Annu Int Conf IEEE Eng Med Biol Soc 2008;2008:2266-2269. [doi: 10.1109/IEMBS.2008.4649648] [Medline: 19163151]
26. Khatami A, Khosravi A, Nguyen T, Lim CP, Nahavandi S. Medical image analysis using wavelet transform and deep belief networks. Expert Systems with Applications 2017 Nov 15;86:190-198 [FREE Full text] [doi: 10.1016/j.eswa.2017.05.073]
27. Joshi GD, Sivaswamy J, Krishnadas SR. Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment. IEEE Trans Med Imaging 2011 Jun;30(6):1192-1205. [doi: 10.1109/TMI.2011.2106509] [Medline: 21536531]
28. Zheng Y, Stambolian D, O'Brien J, Gee JC. Optic disc and cup segmentation from color fundus photograph using graph cut with priors. Med Image Comput Comput Assist Interv 2013;16(Pt 2):75-82 [FREE Full text] [doi: 10.1007/978-3-642-40763-5_10] [Medline: 24579126]
29. Roychowdhury S, Koozekanani DD, Parhi KK. Iterative Vessel Segmentation of Fundus Images. IEEE Trans Biomed Eng 2015 Jul;62(7):1738-1749. [doi: 10.1109/TBME.2015.2403295] [Medline: 25700436]
30. Câmara Neto L, Ramalho GL, Rocha Neto JF, Veras RM, Medeiros FN. An unsupervised coarse-to-fine algorithm for blood vessel segmentation in fundus images. Expert Systems with Applications 2017 Jul 15;78:182-192 [FREE Full text] [doi: 10.1016/j.eswa.2017.02.015]

31. Kurniawan R, Nazri MZA, Sheikh-Abdullah SNH, Hamzah JC, Yendra R, Oktaviana W. Automatic Rule Generator via FP-Growth for Eye Diseases Diagnosis. International Journal on Advanced Science, Engineering and Information Technology 2019;9(3):960-966 doi = {10.18517/ijaseit.9.3.7025} [FREE Full text]
32. Huang H, Yu PS, Wang C. An Introduction to Image Synthesis with Generative Adversarial Nets. Computer Vision and Pattern Recognition 2018 Nov 17:1-17 [FREE Full text]
33. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun. ACM 2017 May 24;60(6):84-90 [FREE Full text] [doi: 10.1145/3065386]
34. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans Pattern Anal Mach Intell 2017 Jun;39(6):1137-1149. [doi: 10.1109/TPAMI.2016.2577031] [Medline: 27295650]
35. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis 2015 Apr 11;115(3):211-252 [FREE Full text] [doi: 10.1007/s11263-015-0816-y]
36. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. : IEEE; 2015 Oct 15 Presented at: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015; Boston, MA, USA p. 1-9 URL: https://ieeexplore.ieee.org/document/7298594 [doi: 10.1109/CVPR.2015.7298594]
37. Huang G, Liu Z, Maaten LVD, Weinberger KQ. Densely Connected Convolutional Networks. New York, NY: IEEE; 2017 Nov 9 Presented at: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017; Honolulu, HI p. 2261-2269 URL: https://ieeexplore.ieee.org/document/8099726 [doi: 10.1109/CVPR.2017.243]
38. Dasgupta A, Singh S. A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation. New York, NY: IEEE; 2017 Jun 19 Presented at: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017); 18-21 April 2017; Melbourne, VIC, Australia p. 248-251. [doi: 10.1109/isbi.2017.7950512]
39. Maninis KK, Pont-Tuset J, Arbeláez P, Van Gool L. Deep retinal image understanding. In: Ourselin S, Joskowicz L, Sabuncu M, Unal G, Wells W, editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. MICCAI 2016. Lecture Notes in Computer Science, vol 9901. Cham, Switzerland: Springer; Oct 02, 2016:140-148.
40. Xiuqin P, Zhang Q, Zhang H, Li S. A Fundus Retinal Vessels Segmentation Scheme Based on the Improved Deep Learning U-Net Model. IEEE Access 2019 Aug 13;7:122634-122643 [FREE Full text] [doi: 10.1109/ACCESS.2019.2935138]
41. Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Machine Intell 2002 Jul;24(7):971-987 [FREE Full text] [doi: 10.1109/TPAMI.2002.1017623]
42. Ma J, Wei M, Ma Z, Shi L, Zhu K. Retinal vessel segmentation based on Generative Adversarial network and Dilated convolution. New York, NY: IEEE; 2019 Sep 23 Presented at: 2019 14th International Conference on Computer Science Education (ICCSE); 19-21 August 2019; Toronto, ON, Canada p. 282-287. [doi: 10.1109/iccse.2019.8845491]
43. Lahiri A, Jain V, Mondal A, Biswas PK. Retinal vessel segmentation under extreme low annotation: a gan-based semi-supervised approach. New York, NY: IEEE; 2020 Sep 30 Presented at: 2020 IEEE International Conference on Image Processing (ICIP); 25-28 Oct. 2020; Abu Dhabi, UAE p. 418-422. [doi: 10.1109/icip40778.2020.9190882]
44. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. Commun. ACM 2020 Oct 22;63(11):139-144 [FREE Full text] [doi: 10.1145/3422622]
45. Isola P, Zhu J, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. New York, NY: IEEE; 2017 Nov 09 Presented at: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 21-26 July 2017; Honolulu, HI p. 5967-5976 URL: https://ieeexplore.ieee.org/document/8100115 [doi: 10.1109/CVPR.2017.632]
46. Yu Z, Xiang Q, Meng J, Kou C, Ren Q, Lu Y. Retinal image synthesis from multiple-landmarks input with generative adversarial networks. Biomed Eng Online 2019 May 21;18(1):62 [FREE Full text] [doi: 10.1186/s12938-019-0682-x] [Medline: 31113438]
47. Zhu J, Park T, Isola P, Efros AA. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. New York, NY: IEEE; 2017 Dec 25 Presented at: 2017 IEEE International Conference on Computer Vision (ICCV); 22-29 October 2017; Venice, Italy p. 2242-2251. [doi: 10.1109/ICCV.2017.244]
48. Shrivastava A, Pfister T, Tuzel O, Susskind J, Wang W, Webb R. Learning from Simulated and Unsupervised Images through Adversarial Training. New York, NY: IEEE; 2017 Nov 09 Presented at: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 21-26 July 2017; Honolulu, HI p. 2242-2251. [doi: 10.1109/CVPR.2017.241]
49. Wu C, Zou Y, Yang Z. U-GAN: generative adversarial networks with U-Net for retinal vessel segmentation. New York, NY: IEEE; 2019 Sep 23 Presented at: 2019 14th International Conference on Computer Science & Education (ICCSE); 19-21 Aug. 2019; Toronto, ON, Canada p. 642-646. [doi: 10.1109/ICCSE.2019.8845397]
50. Lahiri A, Ayush K, Biswas PK, Mitra P. Generative adversarial learning for reducing manual annotation in semantic segmentation on large scale microscopy images: automated vessel segmentation in retinal fundus image as test case. New York, NY: IEEE; 2017 Aug 24 Presented at: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); 21-26 July 2017; Honolulu, HI p. 794-800. [doi: 10.1109/CVPRW.2017.110]
51. Haoqi G, Ogawara K. CGAN-based synthetic medical image augmentation between retinal fundus images and vessel segmented images. New York, NY: IEEE; 2020 May 19 Presented at: 2020 5th International Conference on Control and Robotics Engineering (ICCRE); 24-26 April 2020; Osaka, Japan p. 218-223. [doi: 10.1109/ICCRE49379.2020.9096438]

52. Asiri N, Hussain M, Al Adel F, Alzaidi N. Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey. Artif Intell Med 2019 Aug;99:101701. [doi: 10.1016/j.artmed.2019.07.009] [Medline: 31606116]
53. Ting DSW, Peng L, Varadarajan AV, Keane PA, Burlina PM, Chiang MF, et al. Deep learning in ophthalmology: The technical and clinical considerations. Prog Retin Eye Res 2019 Sep;72:100759. [doi: 10.1016/j.preteyeres.2019.04.003] [Medline: 31048019]
54. Balyen L, Peto T. Promising Artificial Intelligence-Machine Learning-Deep Learning Algorithms in Ophthalmology. Asia Pac J Ophthalmol (Phila) 2019;8(3):264-272. [doi: 10.22608/APO.2018479] [Medline: 31149787]
55. Hogarty DT, Mackey DA, Hewitt AW. Current state and future prospects of artificial intelligence in ophthalmology: a review. Clin Exp Ophthalmol 2019 Jan;47(1):128-139. [doi: 10.1111/ceo.13381] [Medline: 30155978]
56. Ting DSJ, Foo VH, Yang LWY, Sia JT, Ang M, Lin H, et al. Artificial intelligence for anterior segment diseases: Emerging applications in ophthalmology. Br J Ophthalmol 2021 Feb;105(2):158-168. [doi: 10.1136/bjophthalmol-2019-315651] [Medline: 32532762]
57. Liu Y, Fu D, Huang Z, Tong H. Optic disc segmentation in fundus images using adversarial training. IET Image Processing 2019 Feb 1;13(2):375-381 [FREE Full text] [doi: 10.1049/iet-ipr.2018.5922]
58. Diaz-Pinto A, Colomer A, Naranjo V, Morales S, Xu Y, Frangi AF. Retinal Image Synthesis for Glaucoma Assessment Using DCGAN and VAE Models. Yin H., Camacho D., Novais P., Tallón-Ballesteros A. (eds) Intelligent Data Engineering and Automated Learning – IDEAL 2018. IDEAL 2018. Lecture Notes in Computer Science 2018 Nov 09;11314:224-232 [FREE Full text] [doi: 10.1007/978-3-030-03493-1_24]
59. Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. 4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings 2016 Jan 07:1-16 [FREE Full text]
60. Arjovsky M, Chintala S, Bottou L. Wasserstein Generative Adversarial Networks. 2017 Presented at: Proceedings of the 34th International Conference on Machine Learning; 6-11 August 2017; Sydney, Australia p. 214-223.
61. Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A. Improved training of wasserstein GANs. Adv Neural Inf Process Syst Internet Neural information processing systems foundation 2017 Mar 31:5768-5778 [FREE Full text]
62. Mao X, Li Q, Xie H, Lau RY, Wang Z, Smolley SP. Least Squares Generative Adversarial Networks. New York, NY: IEEE; 2017 Dec 25 Presented at: 2017 IEEE International Conference on Computer Vision (ICCV); 22-29 October 2017; Venice, Italy p. 2813-2821. [doi: 10.1109/ICCV.2017.304]
63. Larsen ABL, Sønderby SK, Larochelle H, Winther O. Autoencoding beyond pixels using a learned similarity metric. 2016 Presented at: Proceedings of The 33rd International Conference on Machine Learning; 20-22 June 2016; New York, NY p. 1558-1566.
64. Yas QM, Zaidan AA, Zaidan BB, Hashim M, Lim CK. A Systematic Review on Smartphone Skin Cancer Apps: Coherent Taxonomy, Motivations, Open Challenges and Recommendations, and New Research Direction. J CIRCUIT SYST COMP 2018 Feb 06;27(05):1830003-1830040 [FREE Full text] [doi: 10.1142/s0218126618300039]
65. Zhang S, Zhai J, Luo D, Zhan Y, Chen J. Recent Advance On Generative Adversarial Networks. New York, NY: IEEE; 2018 Nov 12 Presented at: 2018 International Conference on Machine Learning and Cybernetics (ICMLC); 15-18 July 2018; Chengdu, China p. 69-74. [doi: 10.1109/ICMLC.2018.8526990]
66. Gonog L, Zhou Y. A review: generative adversarial networks. New York, NY: IEEE; 2019 Sep 16 Presented at: 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA); 19-21 June 2019; Xi'an, China p. 505-510. [doi: 10.1109/ICIEA.2019.8833686]
67. Liu J, Ke Y, Zhang Z, Lei Y, Li J, Zhang M, et al. Recent Advances of Image Steganography With Generative Adversarial Networks. IEEE Access 2020;8:60575-60597 [FREE Full text] [doi: 10.1109/ACCESS.2020.2983175]
68. Cao Y, Jia L, Chen Y, Lin N, Yang C, Zhang B, et al. Recent Advances of Generative Adversarial Networks in Computer Vision. IEEE Access 2019;7:14985-15006 [FREE Full text] [doi: 10.1109/ACCESS.2018.2886814]
69. Zhu K, Liu X, Yang H. A Survey of Generative Adversarial Networks. New York, NY: IEEE; 2019 Jan 24 Presented at: 2018 Chinese Automation Congress (CAC); 30 November to 2 December 2018; Xi'an, China p. 2768-2773. [doi: 10.1109/cac.2018.8623645]
70. Kumar S, Dhawan S. A Detailed Study on Generative Adversarial Networks. New York, NY: IEEE; 2020 Jul 10 Presented at: 2020 5th International Conference on Communication and Electronics Systems (ICCES); 10-12 June 2020; Coimbatore, Tamil Nadu, India p. 641-645. [doi: 10.1109/ICCES48766.2020.9137883]
71. Chen X, Duan Y, Houthooft R, Schulman J, Sutskever I, Abbeel P. InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets. Advances in Neural Information Processing Systems 2016 Dec 5:2180-2188 [FREE Full text]
72. Sengupta S, Singh A, Leopold HA, Gulati T, Lakshminarayanan V. Ophthalmic diagnosis using deep learning with fundus images - A critical review. Artif Intell Med 2020 Jan;102:101758. [doi: 10.1016/j.artmed.2019.101758] [Medline: 31980096]
73. Wang X, Xu M, Li L, Wang Z, Guan Z. Pathology-Aware Deep Network Visualization and Its Application in Glaucoma Image Synthesis. Shen D. et al. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. MICCAI 2019. Lecture Notes in Computer Science 2019 Oct 10;11764:423-431 [FREE Full text] [doi: 10.1007/978-3-030-32239-7_47]

74. Liu S, Hong J, Lu X, Jia X, Lin Z, Zhou Y, et al. Joint optic disc and cup segmentation using semi-supervised conditional GANs. Comput Biol Med 2019 Dec;115:103485. [doi: 10.1016/j.compbiomed.2019.103485] [Medline: 31630029]
75. Yu F, Dong H, Zhang M, Zhao J, Dong B, Li Q, et al. AF-SEG: An Annotation-Free Approach for Image Segmentation by Self-Supervision and Generative Adversarial Network. New York, NY: IEEE; 2020 May 22 Presented at: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI); 3-7 April 2020; Iowa City, IA p. 1503-1507. [doi: 10.1109/isbi45749.2020.9098535]
76. Yang J, Dong X, Hu Y, Peng Q, Tao G, Ou Y, et al. Fully Automatic Arteriovenous Segmentation in Retinal Images via Topology-Aware Generative Adversarial Networks. Interdiscip Sci 2020 Sep;12(3):323-334. [doi: 10.1007/s12539-020-00385-5] [Medline: 32725575]
77. Wang J, Yan Y, Xu Y, Zhao W, Min H, Tan M, et al. Conditional Adversarial Transfer for Glaucoma Diagnosis. Annu Int Conf IEEE Eng Med Biol Soc 2019 Jul;2019:2032-2035. [doi: 10.1109/EMBC.2019.8857308] [Medline: 31946300]
78. Tu W, Hu W, Liu X, He J. DRPAN: A novel Adversarial Network Approach for Retinal Vessel Segmentation. New York, NY: IEEE; 2019 Sep 16 Presented at: 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA); 19-21 June 2019; Xi'an, China p. 228-232. [doi: 10.1109/ICIEA.2019.8833908]
79. Kadambi S, Wang Z, Xing E. WGAN domain adaptation for the joint optic disc-and-cup segmentation in fundus images. Int J Comput Assist Radiol Surg 2020 Jul;15(7):1205-1213. [doi: 10.1007/s11548-020-02144-9] [Medline: 32445127]
80. Son J, Park SJ, Jung K. Towards Accurate Segmentation of Retinal Vessels and the Optic Disc in Fundoscopic Images with Generative Adversarial Networks. J Digit Imaging 2019 Jun;32(3):499-512 [FREE Full text] [doi: 10.1007/s10278-018-0126-3] [Medline: 30291477]
81. Bisneto TRV, de Carvalho Filho AO, Magalhães DMV. Generative adversarial network and texture features applied to automatic glaucoma detection. Applied Soft Computing 2020 May;90:106165 [FREE Full text] [doi: 10.1016/j.asoc.2020.106165]
82. Zhao H, Qiu X, Lu W, Huang H, Jin X. High-quality retinal vessel segmentation using generative adversarial network with a large receptive field. Int J Imaging Syst Technol 2020 Apr 09;30(3):828-842 [FREE Full text] [doi: 10.1002/ima.22428]
83. Lu S, Hu M, Li R, Xu Y. A Novel Adaptive Weighted Loss Design in Adversarial Learning for Retinal Nerve Fiber Layer Defect Segmentation. IEEE Access 2020;8:132348-132359 [FREE Full text] [doi: 10.1109/ACCESS.2020.3009442]
84. Rammy SA, Abbas W, Hassan N, Raza A, Zhang W. CPGAN: Conditional patch ‐ based generative adversarial network for retinal vesselsegmentation. IET image process 2020 Apr 10;14(6):1081-1090 [FREE Full text] [doi: 10.1049/iet-ipr.2019.1007]
85. Dong Y, Ren W, Zhang K. Deep supervision adversarial learning network for retinal vessel segmentation. 2020 Jan 23 Presented at: 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI); 19-21 October 2019; Suzhou, China p. 1-6. [doi: 10.1109/CISP-BMEI48845.2019.8965924]
86. Jiang Y, Tan N, Peng T. Optic Disc and Cup Segmentation Based on Deep Convolutional Generative Adversarial Networks. IEEE Access 2019 May 17;7:64483-64493 [FREE Full text] [doi: 10.1109/ACCESS.2019.2917508]
87. Costa P, Galdran A, Meyer MI, Niemeijer M, Abramoff M, Mendonca AM, et al. End-to-End Adversarial Retinal Image Synthesis. IEEE Trans Med Imaging 2018 Mar;37(3):781-791 [FREE Full text] [doi: 10.1109/TMI.2017.2759102] [Medline: 28981409]
88. Iqbal T, Ali H. Generative Adversarial Network for Medical Images (MI-GAN). J Med Syst 2018 Oct 12;42(11):231 [FREE Full text] [doi: 10.1007/s10916-018-1072-9] [Medline: 30315368]
89. Wang S, Yu L, Yang X, Fu C, Heng P. Patch-Based Output Space Adversarial Learning for Joint Optic Disc and Cup Segmentation. IEEE Trans Med Imaging 2019 Nov;38(11):2485-2495 [FREE Full text] [doi: 10.1109/TMI.2019.2899910] [Medline: 30794170]
90. Diaz-Pinto A, Colomer A, Naranjo V, Morales S, Xu Y, Frangi AF. Retinal Image Synthesis and Semi-Supervised Learning for Glaucoma Assessment. IEEE Trans Med Imaging 2019 Sep;38(9):2211-2218 [FREE Full text] [doi: 10.1109/TMI.2019.2903434] [Medline: 30843823]
91. He J, Jiang D. Fundus image segmentation based on improved generative adversarial network for retinal vessel analysis. New York, NY: IEEE; 2020 Jul 09 Presented at: 2020 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD); 28-31 May 2020; Chengdu, China p. 231-236. [doi: 10.1109/icaibd49809.2020.9137459]
92. Biswas S, Rohdin J, Drahansky M. Synthetic Retinal Images from Unconditional GANs. Annu Int Conf IEEE Eng Med Biol Soc 2019 Jul;2019:2736-2739 [FREE Full text] [doi: 10.1109/EMBC.2019.8857857] [Medline: 31946460]
93. Zhao H, Li H, Maurer-Stroh S, Cheng L. Synthesizing retinal and neuronal images with generative adversarial nets. Med Image Anal 2018 Oct;49:14-26 [FREE Full text] [doi: 10.1016/j.media.2018.07.001] [Medline: 30007254]
94. Park K, Choi SH, Lee JY. M-GAN: Retinal Blood Vessel Segmentation by Balancing Losses Through Stacked Deep Fully Convolutional Networks. IEEE Access 2020;8:146308-146322 [FREE Full text] [doi: 10.1109/ACCESS.2020.3015108]
95. Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: A review. Med Image Anal 2019 Dec;58:101552 [FREE Full text] [doi: 10.1016/j.media.2019.101552] [Medline: 31521965]
96. Wang K, Gou C, Duan Y, Lin Y, Zheng X, Wang F. Generative adversarial networks: introduction and outlook. IEEE/CAA Journal of Automatica Sinica 2017;4(4):588-598 [FREE Full text] [doi: 10.1109/JAS.2017.7510583]

97. Wu X, Xu K, Hall P. A survey of image synthesis and editing with generative adversarial networks. Tsinghua Science and Technology 2017 Dec;22(6):660-674 [FREE Full text] [doi: 10.23919/TST.2017.8195348]
98. Fan Z, Hu J. Review and Prospect of Research on Generative Adversarial Networks. New York, NY: IEEE; 2019 Nov 21 Presented at: 2019 IEEE 11th International Conference on Communication Software and Networks (ICCSN); 12-15 June 2019; Chongqing, China p. 726-730. [doi: 10.1109/iccsn.2019.8905263]
99. Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative Adversarial Networks: An Overview. IEEE Signal Processing Magazine 2018 Jan;35(1):53-65 [FREE Full text] [doi: 10.1109/MSP.2017.2765202]
100. Lucic M, Kurach K, Michalski M, Gelly S, Bousquet O. Are GANs Created Equal? A Large-Scale Study. New York, NY: ACM; 2018 Oct 29 Presented at: NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems; December 3-8, 2018; Montréal, QC, Canada p. 698-707.
101. Kurach K, Lucic M, Zhai X, Michalski M, Gelly S. The GAN Landscape: Losses, Architectures, Regularization, and Normalization. 2018 Sep 28 Presented at: ICLR 2019 Conference; May 6-9, 2019; New Orleans, LA p. 1-17 URL: https://openreview.net/forum?id=rkGG6s0qKQ
102. Turhan CG, Bilge HS. Recent trends in deep generative models: a review. New York, NY: IEEE; 2018 Dec 10 Presented at: 2018 3rd International Conference on Computer Science and Engineering (UBMK); 20-23 September 2018; Sarajevo, Bosnia and Herzegovina p. 574-579. [doi: 10.1109/UBMK.2018.8566353]
103. Oussidi A, Elhassouny A. Deep generative models: survey. New York, NY: IEEE; 2018 May 07 Presented at: 2018 International Conference on Intelligent Systems and Computer Vision (ISCV); 2-4 April 2018; Fez, Morocco p. 1-8. [doi: 10.1109/ISACV.2018.8354080]
104. Grewal PS, Oloumi F, Rubin U, Tennant MTS. Deep learning in ophthalmology: a review. Can J Ophthalmol 2018 Aug;53(4):309-313 [FREE Full text] [doi: 10.1016/j.jcjo.2018.04.019] [Medline: 30119782]
105. Wu X, Liu L, Zhao L, Guo C, Li R, Wang T, et al. Application of artificial intelligence in anterior segment ophthalmic diseases: diversity and standardization. Ann Transl Med 2020 Jun;8(11):714 [FREE Full text] [doi: 10.21037/atm-20-976] [Medline: 32617334]
106. Tan NYQ, Friedman DS, Stalmans I, Ahmed IIK, Sng CCA. Glaucoma screening: where are we and where do we need to go? Curr Opin Ophthalmol 2020 Mar;31(2):91-100. [doi: 10.1097/ICU.0000000000000649] [Medline: 31904596]
107. Mayro EL, Wang M, Elze T, Pasquale LR. The impact of artificial intelligence in the diagnosis and management of glaucoma. Eye (Lond) 2020 Jan;34(1):1-11 [FREE Full text] [doi: 10.1038/s41433-019-0577-x] [Medline: 31541215]
108. Islam M, Poly TN, Yang HC, Atique S, Li YJ. Deep Learning for Accurate Diagnosis of Glaucomatous Optic Neuropathy Using Digital Fundus Image: A Meta-Analysis. Stud Health Technol Inform 2020 Jun 16;270:153-157 [FREE Full text] [doi: 10.3233/SHTI200141] [Medline: 32570365]
109. Mustapha M. Retinal Nerve Fiber Layer Thickness Post-Laser Treatment in Diabetic Retinopathy: Argon versus Pattern Scanning Laser. Medicine & Health 2016;11(2):199-208 [FREE Full text] [doi: 10.17576/mh.2016.1102.09]
110. Varma R, Lee PP, Goldberg I, Kotak S. An assessment of the health and economic burdens of glaucoma. Am J Ophthalmol 2011 Oct;152(4):515-522 [FREE Full text] [doi: 10.1016/j.ajo.2011.06.004] [Medline: 21961848]
111. Shibata N, Tanito M, Mitsuhashi K, Fujino Y, Matsuura M, Murata H, et al. Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. Sci Rep 2018 Oct 02;8(1):14665 [FREE Full text] [doi: 10.1038/s41598-018-33013-w] [Medline: 30279554]
112. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, et al. Artificial intelligence and deep learning in ophthalmology. Br J Ophthalmol 2019 Feb;103(2):167-175 [FREE Full text] [doi: 10.1136/bjophthalmol-2018-313173] [Medline: 30361278]
113. Cvenkel B, Velkovska MA. Self-monitoring of intraocular pressure using Icare HOME tonometry in clinical practice. Clin Ophthalmol 2019 May 10;13:841-847 [FREE Full text] [doi: 10.2147/OPTH.S198846] [Medline: 31190727]
114. Zapata MA, Royo-Fibla D, Font O, Vela JI, Marcantonio I, Moya-Sánchez EU, et al. Artificial Intelligence to Identify Retinal Fundus Images, Quality Validation, Laterality Evaluation, Macular Degeneration, and Suspected Glaucoma. Clin Ophthalmol 2020 Feb 13;14:419-429 [FREE Full text] [doi: 10.2147/OPTH.S235751] [Medline: 32103888]
115. Zhao R, Liao W, Zou B, Chen Z, Li S. Weakly-Supervised Simultaneous Evidence Identification and Segmentation for Automated Glaucoma Diagnosis. AAAI 2019 Jul 17;33(1):809-816 [FREE Full text] [doi: 10.1609/aaai.v33i01.3301809]
116. Zhao R, Chen X, Liu X, Chen Z, Guo F, Li S. Direct Cup-to-Disc Ratio Estimation for Glaucoma Screening via Semi-Supervised Learning. IEEE J Biomed Health Inform 2020 Apr;24(4):1104-1113. [doi: 10.1109/JBHI.2019.2934477] [Medline: 31403451]
117. Chai Y, Liu H, Xu J. A new convolutional neural network model for peripapillary atrophy area segmentation from retinal fundus images. Applied Soft Computing 2020 Jan;86:105890 [FREE Full text] [doi: 10.1016/j.asoc.2019.105890]
118. Chai Y, Liu H, Xu J. Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models. Knowledge-Based Systems 2018 Dec 1;161:147-156 [FREE Full text] [doi: 10.1016/j.knosys.2018.07.043]
119. Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a Deep Learning System for Detecting Glaucomatous Optic Neuropathy Based on Color Fundus Photographs. Ophthalmology 2018 Aug;125(8):1199-1206. [doi: 10.1016/j.ophtha.2018.01.023] [Medline: 29506863]

120. Yang HK, Kim YJ, Sung JY, Kim DH, Kim KG, Hwang J. Efficacy for Differentiating Nonglaucomatous Versus Glaucomatous Optic Neuropathy Using Deep Learning Systems. Am J Ophthalmol 2020 Aug;216:140-146. [doi: 10.1016/j.ajo.2020.03.035] [Medline: 32247778]
121. Phene S, Dunn RC, Hammel N, Liu Y, Krause J, Kitade N, et al. Deep Learning and Glaucoma Specialists: The Relative Importance of Optic Disc Features to Predict Glaucoma Referral in Fundus Photographs. Ophthalmology 2019 Dec;126(12):1627-1639 [FREE Full text] [doi: 10.1016/j.ophtha.2019.07.024] [Medline: 31561879]
122. Blumberg DM, De Moraes CG, Liebmann JM, Garg R, Chen C, Theventhiran A, et al. Technology and the Glaucoma Suspect. Invest Ophthalmol Vis Sci 2016 Jul 01;57(9):OCT80-OCT85 [FREE Full text] [doi: 10.1167/iovs.15-18931] [Medline: 27409509]
123. Bhatkalkar BJ, Reddy DR, Prabhu S, Bhandary SV. Improving the Performance of Convolutional Neural Network for the Segmentation of Optic Disc in Fundus Images Using Attention Gates and Conditional Random Fields. IEEE Access 2020 Feb 07;8:29299-29310 [FREE Full text] [doi: 10.1109/ACCESS.2020.2972318]
124. Liu S, Graham SL, Schulz A, Kalloniatis M, Zangerl B, Cai W, et al. A Deep Learning-Based Algorithm Identifies Glaucomatous Discs Using Monoscopic Fundus Photographs. Ophthalmol Glaucoma 2018;1(1):15-22. [doi: 10.1016/j.ogla.2018.04.002] [Medline: 32672627]
125. Christopher M, Belghith A, Bowd C, Proudfoot JA, Goldbaum MH, Weinreb RN, et al. Performance of Deep Learning Architectures and Transfer Learning for Detecting Glaucomatous Optic Neuropathy in Fundus Photographs. Sci Rep 2018 Nov 12;8(1):16685 [FREE Full text] [doi: 10.1038/s41598-018-35044-9] [Medline: 30420630]
126. Gu Z, Cheng J, Fu H, Zhou K, Hao H, Zhao Y, et al. CE-Net: Context Encoder Network for 2D Medical Image Segmentation. IEEE Trans Med Imaging 2019 Oct;38(10):2281-2292. [doi: 10.1109/TMI.2019.2903562] [Medline: 30843824]
127. Serener A, Serte S. Transfer Learning for Early and Advanced Glaucoma Detection with Convolutional Neural Networks. New York, NY: IEEE; 2019 Nov 11 Presented at: 2019 Medical Technologies Congress (TIPTEKNO); 3-5 October 2019; Izmir, Turkey p. 1-4. [doi: 10.1109/TIPTEKNO.2019.8894965]
128. Li F, Yan L, Wang Y, Shi J, Chen H, Zhang X, et al. Deep learning-based automated detection of glaucomatous optic neuropathy on color fundus photographs. Graefes Arch Clin Exp Ophthalmol 2020 Apr;258(4):851-867. [doi: 10.1007/s00417-020-04609-8] [Medline: 31989285]
129. Fu H, Cheng J, Xu Y, Zhang C, Wong DWK, Liu J, et al. Disc-Aware Ensemble Network for Glaucoma Screening From Fundus Image. IEEE Trans Med Imaging 2018 Nov;37(11):2493-2501. [doi: 10.1109/TMI.2018.2837012] [Medline: 29994764]
130. Mustapha M. Retinal Nerve Fibre Layer Thickness Changes after Pan-Retinal Photocoagulation in Diabetic Retinopathy. JSA 2016 Jun 3;6(1):4-9 [FREE Full text] [doi: 10.17845/jsa.2016.0601.02]
131. Russakoff DB, Mannil SS, Oakley JD, Ran AR, Cheung CY, Dasari S, et al. A 3D Deep Learning System for Detecting Referable Glaucoma Using Full OCT Macular Cube Scans. Transl Vis Sci Technol 2020 Feb 18;9(2):12 [FREE Full text] [doi: 10.1167/tvst.9.2.12] [Medline: 32704418]
132. Orlando JI, Fu H, Barbosa Breda J, van Keer K, Bathula DR, Diaz-Pinto A, et al. REFUGE Challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Med Image Anal 2020 Jan;59:101570. [doi: 10.1016/j.media.2019.101570] [Medline: 31630011]
133. Xiong L, Li H. An approach to locate optic disc in retinal images with pathological changes. Comput Med Imaging Graph 2016 Jan;47:40-50. [doi: 10.1016/j.compmedimag.2015.10.003] [Medline: 26650403]
134. Wang Z, Wang Z, Qu G, Li F, Yuan Y, Lam DSC, et al. Intelligent Glaucoma Diagnosis Via Active Learning And Adversarial Data Augmentation. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). New York, NY: IEEE; 2019 Jul 11 Presented at: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019); 8-11 April 2019; Venice, Italy p. 1234-1237. [doi: 10.1109/ISBI.2019.8759178]
135. Tian C, Fang T, Fan Y, Wu W. Multi-path convolutional neural network in fundus segmentation of blood vessels. Biocybernetics and Biomedical Engineering 2020 Apr;40(2):583-595 [FREE Full text] [doi: 10.1016/j.bbe.2020.01.011]
136. Liu H, Li L, Wormstone IM, Qiao C, Zhang C, Liu P, et al. Development and Validation of a Deep Learning System to Detect Glaucomatous Optic Neuropathy Using Fundus Photographs. JAMA Ophthalmol 2019 Dec 01;137(12):1353-1360 [FREE Full text] [doi: 10.1001/jamaophthalmol.2019.3501] [Medline: 31513266]
137. Li L, Xu M, Liu H, Li Y, Wang X, Jiang L, et al. A Large-Scale Database and a CNN Model for Attention-Based Glaucoma Detection. IEEE Trans Med Imaging 2020 Feb;39(2):413-424. [doi: 10.1109/TMI.2019.2927226] [Medline: 31283476]
138. Dengpan Y, Shunzhi J, Shiyu L, ChangRui L. Faster and transferable deep learning steganalysis on GPU. J Real-Time Image Proc 2019 Apr 1;16(3):623-633 [FREE Full text] [doi: 10.1007/s11554-019-00870-1]
139. He K, Zhang X, Ren S, Sun J, editors. Deep Residual Learning for Image Recognition. New York, NY: IEEE; 2016 Dec 12 Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 27-30 June 2016; Las Vegas, NV p. 770-778. [doi: 10.1109/cvpr.2016.90]
140. Khan A, Sohail A, Zahoora U, Qureshi AS. A survey of the recent architectures of deep convolutional neural networks. Artif Intell Rev 2020 Apr 21;53(8):5455-5516 [FREE Full text] [doi: 10.1007/s10462-020-09825-6]
141. Dong C, Loy CC, He K, Tang X. Image Super-Resolution Using Deep Convolutional Networks. IEEE Trans Pattern Anal Mach Intell 2016 Feb;38(2):295-307. [doi: 10.1109/TPAMI.2015.2439281] [Medline: 26761735]

142. Ratliff LJ, Burden SA, Sastry SS. On the Characterization of Local Nash Equilibria in Continuous Games. IEEE Trans. Automat. Contr 2016 Aug;61(8):2301-2307 [FREE Full text] [doi: 10.1109/TAC.2016.2583518]
143. Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. GANs trained by a two time-scale update rule converge to a local nash equilibrium. Red Hook, NY: Curran Associates Inc; 2017 Dec 04 Presented at: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems; 4 December 2017; Long Beach, CA p. 6629-6640.
144. Ganin Y, Lempitsky V. Unsupervised Domain Adaptation by Backpropagation. 2015 Presented at: Proceedings of the 32nd International Conference on Machine Learning; 07-09 Jul 2015; Lille, France p. 1180-1189.
145. Tzeng E, Hoffman J, Saenko K, Darrell T. Adversarial Discriminative Domain Adaptation. New York, NY: IEEE; 2017 Nov 09 Presented at: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 21-26 July 2017; Honolulu, HI p. 2962-2971. [doi: 10.1109/CVPR.2017.316]
146. Mahayuddin ZR, Saifuddin Saif AFM. A Comprehensive Review Towards Segmentation and Detection of Cancer Cell and Tumor for Dynamic 3D Reconstruction. APJITM 2020 Jun 15;09(01):28-39 [FREE Full text] [doi: 10.17576/apjitm-2020-0901-03]
147. Luc P, Couprie C, Chintala S, Verbeek J. Semantic segmentation using adversarial networks. 2016 Presented at: NIPS 2016: The Thirtieth Annual Conference on Neural Information Processing Systems; December 4-9, 2016; Barcelona, Spain p. 1-9 URL: https://research.fb.com/wp-content/uploads/2016/11/luc16wat.pdf
148. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. JAMA 2017 Dec 12;318(22):2211-2223 [FREE Full text] [doi: 10.1001/jama.2017.18152] [Medline: 29234807]
149. Kahng M, Thorat N, Chau DHP, Viegas FB, Wattenberg M. GAN Lab: Understanding Complex Deep Generative Models using Interactive Visual Experimentation. IEEE Trans Vis Comput Graph 2018 Aug 20;25(1):310-320 [FREE Full text] [doi: 10.1109/TVCG.2018.2864500] [Medline: 30130198]
150. Taha BA, Al Mashhadany Y, Hafiz Mokhtar MH, Dzulkefly Bin Zan MS, Arsad N. An Analysis Review of Detection Coronavirus Disease 2019 (COVID-19) Based on Biosensor Application. Sensors (Basel) 2020 Nov 26;20(23):6764 [FREE Full text] [doi: 10.3390/s20236764] [Medline: 33256085]
151. Javanmardi M, Tasdizen T. Domain adaptation for biomedical image segmentation using adversarial training. New York, NY: IEEE; 2018 May 24 Presented at: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018); 4-7 April 2018; Washington, DC p. 554-558. [doi: 10.1109/ISBI.2018.8363637]
152. Sajjadi MSM, Bachem O, Lucic M, Bousquet O, Gelly S. Assessing generative models via precision and recall. In: Advances in Neural Information Processing Systems. Red Hook, NY: Curran Associates, Inc; 2018 May 31 Presented at: 32nd Conference on Neural Information Processing Systems (NeurIPS 2018); 2018; Montréal, QC, Canada.
153. Arora S, Risteski A, Zhang Y. Do GANs learn the distribution? Some Theory and Empirics. In: ICLR 2018 Conference Track 6th International Conference on Learning Representations. 2018 Feb 16 Presented at: ICLR 2018 Conference Track 6th International Conference on Learning Representations; April 30-May 3, 2018; Vancouver Convention Center, Vancouver, BC, Canada URL: https://openreview.net/forum?id=BJehNfW0-
154. Madani A, Moradi M, Karargyris A, Syeda-mahmood T. Semi-supervised learning with generative adversarial networks for chest X-ray classification with ability of data domain adaptation. New York, NY: IEEE; 2018 May 24 Presented at: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018); 4-7 April 2018; Washington, DC p. 1038-1042. [doi: 10.1109/ISBI.2018.8363749]
155. Misra D. Mish: A Self Regularized Non-Monotonic Neural Activation Function. 2020 Presented at: BMVC 2020 : British Machine Vision Conference; September 7-11, 2020; Manchester, UK p. 1-14.
156. Chang J, Lee J, Ha A, Han YS, Bak E, Choi S, et al. Explaining the Rationale of Deep Learning Glaucoma Decisions with Adversarial Examples. Ophthalmology 2021 Jan;128(1):78-88. [doi: 10.1016/j.ophtha.2020.06.036] [Medline: 32598951]
157. Huang Y, Cheng Y, Bapna A, Firat O, Chen MX, Chen D, et al. GPipe: Efficient Training of Giant Neural Networks using Pipeline Parallelism. 2019 Presented at: Advances in Neural Information Processing Systems 32 (NeurIPS 2019); December 8-14, 2019; Vancouver, BC, Canada p. 103-112. [doi: 10.1109/emc2-nips53020.2019.00008]
158. Neff T, Payer C, Stern D, Urschler M. Generative Adversarial Network based Synthesis for Supervised Medical Image Segmentation. Graz, Austria: Verlag der Technischen Universität Graz; 2017 May Presented at: Proceedings of the OAGM&ARW Joint Workshop 2017: Vision, Automation and Robotics; May 10-12, 2017; OAGM/AAPR ARW 2017, Wien, Austria p. 140-145. [doi: 10.3217/978-3-85125-524-9-30]

Abbreviations

BV: blood vessel
CDR: cup-to-disc ratio rule
cGANs: conditional generative adversarial networks
CNN: convolutional neural network
DCGAN: deep convolutional generative adversarial network
DLM: deep learning method

DRIVE data set: digital retinal image for vessels extraction
GANs: generative adversarial networks
ISNT: inferior, superior, nasal, and temporal rule
LAG data set: large-scale-attention-based glaucoma
LSGAN: least-square GAN
OC: optic cup
OD: optic disc
ONH: optic nerve head
REFUGE data set: retinal fundus glaucoma challenge
RGB: red green blue
RL: rim loss
RNFL: retinal nerve fiber layer
ROI: region of interest
STARE data set: structured analysis of the retina
VAEGAN: variational autoencoder with GAN
WGAN: Wasserstein GAN
WGAN-GP: Wasserstein GAN-gradient penalty

Edited by R Kukafka; submitted 26.01.21; peer-reviewed by M Ahmed Kamal, RS Mahmoud, G Lim, A Naser, I Gabashvili; comments to author 06.04.21; revised version received 11.05.21; accepted 05.07.21; published 21.09.21.

Please cite as:
Saeed AQ, Sheikh Abdullah SNH, Che-Hamzah J, Abdul Ghani AT
Accuracy of Using Generative Adversarial Networks for Glaucoma Detection: Systematic Review and Bibliometric Analysis
J Med Internet Res 2021;23(9):e27414
URL: https://www.jmir.org/2021/9/e27414
doi:10.2196/27414
PMID:34236992

Review

Factors Affecting Engagement in Web-Based Health Care Patient Information: Narrative Review of the Literature

Liam Alperen Oktay[1], BSc; Eyad Abuelgasim[1], BSc; Aida Abdelwahed[1], BSc; Nour Houbby[1], BSc; Smaragda Lampridou[2], MSc, RN; Pasha Normahani[2], BSc, MBBS, MSc, MRSC; Nicholas Peters[1], MBBS, MD, FRCP, FHRS; Usman Jaffer[2], BSc, MBBS, MSc, PhD, PGCE, FRCS

[1]Imperial College London, London, United Kingdom

[2]Imperial College NHS Trust, London, United Kingdom

Corresponding Author:
Usman Jaffer, BSc, MBBS, MSc, PhD, PGCE, FRCS
Imperial College NHS Trust
Praed Street
London, W2 0NE
United Kingdom
Phone: 44 7968872992
Email: usman.jaffer@nhs.net

Abstract

Background: Web-based content is rapidly becoming the primary source of health care information. There is a pressing need for web-based health care content to not only be accurate but also be engaging. Improved engagement of people with web-based health care content has the potential to inform as well as influence behavioral change to enable people to make better health care choices. The factors associated with better engagement with web-based health care content have previously not been considered.

Objective: The aims of this study are to identify the factors that affect engagement with web-based health care content and develop a framework to be considered when creating such content.

Methods: A comprehensive search of the PubMed and MEDLINE database was performed from January 1, 1946, to January 5, 2020. The reference lists of all included studies were also searched. The Medical Subject Headings database was used to derive the following keywords: “patient information,” “online,” “internet,” “web,” and “content.” All studies in English pertaining to the factors affecting engagement in web-based health care patient information were included. No restrictions were set on the study type. Analysis of the themes arising from the results was performed using inductive content analysis.

Results: The search yielded 814 articles, of which 56 (6.9%) met our inclusion criteria. The studies ranged from observational and noncontrolled studies to quasi-experimental studies. Overall, there was significant heterogeneity in the types of interventions and outcome assessments, which made quantitative assessment difficult. Consensus among all authors of this study resulted in six categories that formed the basis of a framework to assess the factors affecting engagement in web-based health care content: easy to understand, support, adaptability, accessibility, visuals and content, and credibility and completeness.

Conclusions: There is a paucity of high-quality data relating to the factors that improve the quality of engagement with web-based health care content. Our framework summarizes the reported studies, which may be useful to health care content creators. An evaluation of the utility of web-based content to engage users is of significant importance and may be accessible through tools such as the Net Promoter score. Web 3.0 technology and development of the field of psychographics for health care offer further potential for development. Future work may also involve improvement of the framework through a co-design process.

(J Med Internet Res 2021;23(9):e19896) doi:10.2196/19896

KEYWORDS

patient education; web-based health information; internet; patient engagement; mobile phone

Introduction

Background

In the United Kingdom, up to two-third use the internet to obtain health-related information at some point in their journey [1,2]. The internet has become an important source of education for patients, who are increasingly expected to, and are motivated to, play an active role in making decisions related to their health [3]. Patient education is defined as "the process by which health professionals and others impart information to patients that will alter their health behaviours or improve their health status" [4]. This may include information that is factual or related to patient experience, depending on the issue being addressed [5,6]. Reports suggest that 70% of the patients would like their physicians to recommend a source of web-based information relating to their medical condition, but only 4% of the patients receive such a recommendation [7].

Web-based patient health care information has several potential benefits, including convenient 24-hour access potentially wherever you are, ability to enhance knowledge acquisition [8,9], reduce anxiety [9], and improve the quality of conversations during health-related encounters. However, the effect of web-based content on patient empowerment, self-efficacy, and health attitudes has been found to be variable [9]. This may be due to the absence of an evidence-based framework outlining the factors that should be considered for improving engagement with web-based health care information.

Objectives

In this narrative review, we aim to identify and evaluate the factors that should be considered when producing engaging and high-quality web-based health care patient information. We also aim to incorporate these findings into a framework that may be useful as a guide to developing web-based health care information.

Methods

Overview

A comprehensive search of the PubMed and MEDLINE database was performed from January 1, 1946, to January 5, 2020. The Medical Subject Headings database was used to derive keywords and search term combinations, which included "patient information," "online," "internet," "web," and "content." All studies pertaining to the factors affecting engagement in web-based health care patient information were included. No restrictions were set on the study type. Only studies in English were included. Analysis of themes arising from the results was performed using inductive content analysis. All retrieved abstracts and titles were reviewed by 2 independent investigators (EA and LAO) for relevance pertaining to engagement with web-based health care content. Disagreements between the reviewers were solved by consensus. Manual cross-checking of the reference lists of the identified papers was carried out to identify any other potentially relevant studies.

Analysis of themes arising from the results was performed using inductive content analysis [10]. This involved the reviewing of titles and abstracts by an author (UJ), with free generation of the categories relating to *factors associated with high-quality and engaging web-based content*. The categories were named using content characteristic words, and these were expanded into subcategories. The number of times a category was reported in the articles was totaled, and a list of categories was created in order of the frequency of mentions. All authors reviewed this list to decide which categories would be included in the final list.

Eligibility Criteria

As this is the first narrative review of its kind, no limitation was placed on study type or on surrogate measures of the outcome described.

Outcomes

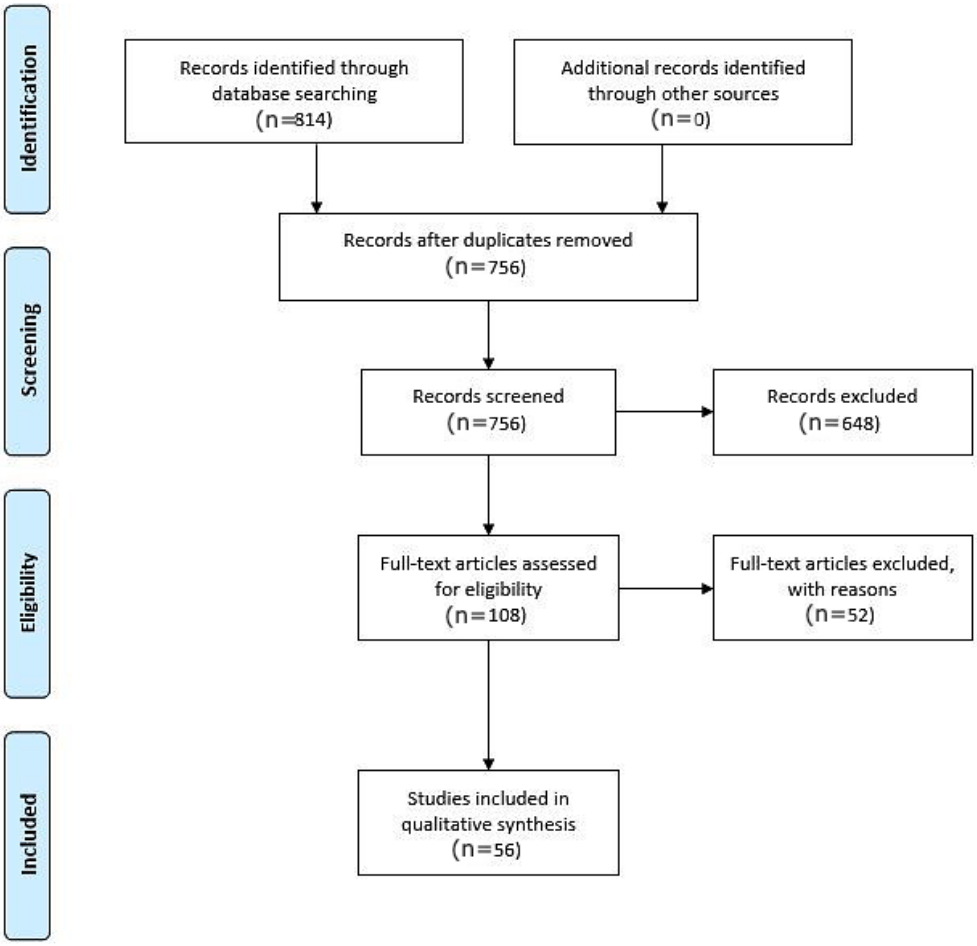
All studies relevant to quality of engagement in web-based health care content were included.

Results

Overview

A total of 814 articles were identified, and of these, 108 (13.3%) were selected for full-text review based on their title and abstract. Full-text screening of the 108 articles resulted in the final selection of 56 (51.9%) articles, from which seven categories were derived. Figure 1 shows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart which depicts the stages of article selection.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart showing article selection.



Description of Studies

Development of Categories

The studies were mostly observational and qualitative. A total of 3 randomized controlled trials and 5 systematic reviews were also identified. Categories relating to *factors associated with high-quality and engaging web-based content* were developed according to themes that were found to arise in the studies identified. The categories settled upon were as follows (with the number of papers reporting included in parentheses): (1) textual information (16 papers); (2) discussion boards or web-based groups (3 papers); (3) video content (11 papers); (4) visuals or pictographs (1 paper); (5) device accessibility (12 papers); (6) stage of patient journey (8 papers); and (7) credibility and completeness of information (4 papers; Table 1).

Table 1. Studies pertaining to engagement with web-based content. Studies are according to mode of engagement (N=56).

Study	Study type	Content	Outcomes pertaining to quality of web-based health care patient content	Main results
Textual information				
Chedid et al (2018) [11]	Observational study	Government-hosted website, textual, and visual aids for prenatal health promotion	• Comprehensiveness, evidence-based information, accessibility, and inclusivity. • Minimum of three referenced prenatal health topics necessary to be classed as evidence-based.	• Assessment of prenatal health promotion material revealed that 66.4% of the federal, 84.6% of the provincial or territorial, and 80% of the public health regional unit–hosted websites, and 87.5% of the e-classes were evidence based. Only 25% of the municipal websites met this standard. • No *P* values stated.
Ernst et al (2019) [12]	Observational study	Disorders of sex development, affiliated health care system's web-based information	• The SMOGa Readability Formula determined reading level, the PEMATb evaluated content for understandability and actionability, and the DISCERN Tool assessed treatment decision-making information.	• Reading level of webpages as determined with the SMOG Readability Formula met or exceeded high school grade level. The mean PEMAT understandability score for team pages and team links was 68% (SD 6%). On average, the pages met less than 70% of the understandability criteria. The mean PEMAT actionability score was 23% (SD 20%). The DISCERN Tool found that the quality of information relating to hormone treatment and to surgery was poor. • No *P* values stated.
Hjelmager et al (2019) [13]	Qualitative study	Web-based information material for patients with low back pain in general practice discussed in the context of app development	• Barriers to and facilitators for future use of the health information technology app for patients with low back pain.	• Eight 1-hour interviews with general practitioners revealed the following: content for lower back pain should be validated by general practitioners; from a trustworthy source; support ongoing treatment plan. • No *P* values stated.
Rofaiel et al (2018) [14]	Observational study	Websites that describe the biologic agents used as treatment options for inflammatory bowel disease	• The DISCERN model was used to evaluate the quality of the information content.	• The mean DISCERN score across all websites was 3.21 out of a 5-point scale. No significant difference was found between patient-searched and physician-recommended websites, with a mean score of 3.21 versus 3.63, respectively (*P*=.16).
Alfonso et al (2019) [15]	Observational study	American Cleft Palate–Craniofacial Association–approved teams' websites	• Content and readability of team websites.	• The mean reading level 10.7 (SD 1.9) exceeded the American Medical Association-recommended sixth grade reading level. • Children's Hospital–affiliated teams (n=86) yielded significantly higher content scores (14.8 vs 13.5; *P*=.03). • Children's Hospital teams also had better readability as indicated by lower reading grade level (10.5 vs 11.4; *P*=.04).
Ayyaswami et al (2019) [16]	Observational study	Web-based cardiovascular disease–related health education articles accessed through Google	• Readability according to 10 readability measures (Flesch Reading Ease, Coleman-Liau Index, Flesch-Kincaid Grade Level, Gunning Fog Index, FORCAST Readability Formula, New Dale-Chall formula, New Fog Count, SMOG Index, Fry Readability Formula, and Raygor Readability Estimate).	• All measures that assessed mean reading grade level found that 196 articles were written at a mean 10.9 (SD 1.8) grade reading level. • 99.5% of the articles were written beyond the fifth to sixth grade reading level. • No *P* values stated.

Study	Study type	Content	Outcomes pertaining to quality of web-based health care patient content	Main results
Fajardo et al (2019) [17]	Systematic review	Web-based deprescribing patient education materials	• PEMAT and International Patient Decision Aids Standards Inventory. • Readability using Gunning Fog Index and Flesch-Kincaid Grade Level.	• Patient education materials addressing deprescribing of medications for symptom control (81%) were most common. • 37% of deprescribing patient education materials present potential benefits and harms of deprescribing. • Most patient education materials are pitched above average reading levels (average minimum reading level of grade 12). • No *P* values stated.
Vivekanantham et al (2017) [18]	Observational study	Web-based health information for patients with polymyalgia rheumatica	• Readability using the Flesch Reading Ease and SMOG Readability Formula tools. • 8-item Credibility Indicator (incorporating authorship, affiliation, editorial team, date of creation, date of update, backing, accreditation, and financing).	• Of the websites identified (n=52), the mean Flesch Reading Ease and SMOG Readability Formula scores were 48 (SD 15) and 10 (SD 2), respectively. • The mean Credibility Indicator was 2 (SD 1). • Of 52 websites, 50 (96%) of the websites were accurate. • No *P* values stated.
Harris et al (2018) [19]	Cross-sectional descriptive	Evaluation of leading web-based content on tympanostomy tube placement	• PEMAT understandability and actionability scores	• The PEMAT scores found that all sites (n=10) were understandable (mean 81.9%). • Seven of the sites had a low actionability score (mean 44.6%). • No *P* values stated.
Maciolek et al (2017) [20]	Observational study	Prostate biopsy web-based patient education materials	• Readability was assessed using Flesch-Kincaid Grade Level. Understandability and actionability were measured using the PEMAT.	• Of the 148 sites, 31 (20.9%) met the recommended below eighth grade reading level. • The PEMAT understandability score for academic institution–sourced patient education materials was higher than that for patient education materials sourced from the private sector (*P*=.02) and from institutions unaffiliated with urologists (*P*=.01).
Siddhanamatha et al (2017) [21]	Observational study	Websites providing educational content for patients with rheumatoid arthritis	• Accuracy, completeness, technical elements, design and aesthetics, readability, usability, and accessibility of the websites	• Of 46 websites in total, 45 (98%) provided accurate information. The mean reading level was grade 12.1 (SD 2.3). • In total, 78% (36/46) were easy to navigate, but only 33% (15/46) were user friendly for people with visual and or hearing impairments. • No *P* values stated.
Nielsen-Bohlman et al (2004) [22]	Observational study	An evaluation of health literacy in the United States; formulate solution to overcome associated obstacles	• Current level of readability of web-based content	• More than 300 studies indicate that health-related materials exceed the average reading grade level of US adults. • No *P* values stated.
Chin et al (2018) [23]	Observational study	Improve patient understanding of web-based content pertaining to adults with hypertension.	• Information retention and comprehension	• The revised passages yielded improved retention and comprehension, with less reading time required per unit uptake of information also noted. The methods included simplifying language and signaling clear organization. • On average, the participants were found to significantly better remember the revised passages (mean 0.74, SD 0.14) compared with the typical passages (mean 0.70, SD 0.11; *P*<.01).

Study	Study type	Content	Outcomes pertaining to quality of web-based health care patient content	Main results
Boudewyns et al (2015) [24]	Randomized controlled trial	Web-based information handouts	• Information comprehension and application	• Individuals who received the revised and improved web-based formats had higher comprehension scores than those who received the MedGuide (P<.001).
Morrow et al (2005) [25]	Observational study	Patient-centered medication instructions to empower patients to plan a medication-taking regimen	• Information comprehension and retention, health-related literacy, and verbal working memory	• Patient-centered instructions were more accurately understood for unfamiliar medications (P<.05). • The standard instructions were more accurate for familiar medications (P<.05).
Pander Maat et al (2010) [26]	Observational study	Revision of patient information leaflets	• Usability, speed of information retrieval, and comprehension.	• Once located, comprehension of the information was approximately 90%. • Revisions led to better performance. Information was found more quickly. Comprehension scores were also improved. • No P values stated.
Discussion boards or web-based groups				
Cedars et al (2019) [27]	Qualitative and thematic analysis	Web-based discussion boards for urethral stricture disease and urethroplasty	• To describe the patient experience and chief concerns with urethroplasty to improve physician understanding and patient education • To understand how men use web-based discussion boards and what information is available about urethroplasty	• Problems in navigating the health care system with urethral stricture disease (n=141) and weak urine stream (n=70) were the most frequent preurethroplasty complaints. • The patients participated in web-based discussions to share experiences with urethral stricture disease and urethroplasty, share emotional support, and search for answers. • No P values stated.
Teaford et al (2019) [28]	Descriptive statistics and qualitative content analysis	New mothers' experiences with web-based postpartum forums	• To explore women's experiences with a web-based forum during the postpartum period	• Five themes were produced through data analysis: (1) social support, (2) anonymity, (3) in-groups, (4) drama, and (5) entertainment or pastime. • The participants found that the discussion board could facilitate sharing of information, act as an entertainment source, and provide community. • No P values stated.
Castaneda et al (2019) [29]	Qualitative study	eHealth peripheral artery disease community forums	Original posts and related responses were analyzed for thematic content.	• The themes identified included medical advice (41%), personal experiences with peripheral artery disease (33%), and social support (13%). Negative attitudes were discussed in 10 of the 18 (56%) posts related to poor experiences with health care providers; 15.1% of the medical advice was inconsistent with clinical treatment guidelines. • No P values stated.
Video				

Study	Study type	Content	Outcomes pertaining to quality of web-based health care patient content	Main results
Bae et al (2018) [30]	Observational study	YouTube videos in English as a patient education resource for cataract surgery	• 14 criteria important for educating patients about the procedure.	• The mean number of usefulness criteria satisfied was only 2.28 (SD 1.80) out of 14. • There was no significant difference in view counts between the most useful videos and other videos (P=.94). • Videos from medical bodies such as the National Health Service were found more useful in terms of patient education (P<.001).
Pedersen et al (2019) [31]	Feasibility study	Development of a preventive educational video for patients exposed to whiplash trauma	• The development followed a systematic approach and was theory driven, supplemented with available empirical knowledge.	• The participants (n=4) felt that the content was "relevant, helpful, and reassuring to watch." • All four preferred video content instead of written material. • No P values stated.
Finnegan et al (2018) [32]	Case study	A web-based vaccine communication project (textual, videos, and infographics)	• Case study of a provaccine information hub launched in 2011. Vaccines Today provides high-quality information about vaccines and diseases, expert interviews, answers to frequently asked questions, parent or patient stories, and videos or infographics.	• Two categories of informing patients were found to work well: (1) the storytelling approach and (2) answering questions posed by patients. • No P values stated.
Button et al (2018) [33]	Mixed methods study (qualitative and feasibility study)	A web-based intervention (TRAK[c]) that provides knee patients with health information	• Testing the TRAK intervention in patients undergoing physiotherapy to gain their subjective insights into its use	• The participants reported that TRAK was easy to use overall. Basic internet skills were all that were required. • Using TRAK as an adjunct to physiotherapist management increased the patients' understanding and confidence. • No P values stated.
Vogel et al (2018) [34]	User survey	VaPE[d] in anesthesia	• The content of the videos, the technique of video presentation, usefulness of VaPE • Interviews carried out with patients and physicians	• In total, 98% (78/80) of the anesthetists found VaPE useful for patient education. • In total, 93% (74/80) observed time saved for the following interview. • In total, 96% (77/80) stated that watching the video left patients better informed. • Increased anxiety caused by VaPE was noted by 46% (37/80); 54% (43/80) found no such effect. • No P values stated.
Pithadia et al (2019) [35]	Cross-sectional study	YouTube videos as a source of patient information about phototherapy and excimer laser for psoriasis	• Assess the educational quality of YouTube videos about phototherapy and excimer laser for psoriasis	• In total, 11.2% (15/135) of the videos contained high-quality patient educational information, 2.5% (3/135) were fair quality, and 66.1% (89/135) were low quality. • A total of 28.2% (35/135) of videos provided background information regarding psoriasis. Of these 35 videos, 28 (80%) contained evidence-based content about the epidemiology, systemic involvement, genetics, and immune nature of psoriasis. • Of the 35 videos, 7 (20%) presented nonevidence-based claims and high mortality rates associated with psoriasis. • No P values stated.
Ferhatoglu et al (2019) [36]	Observational study			

Study	Study type	Content	Outcomes pertaining to quality of web-based health care patient content	Main results
		Sleeve gastrectomy videos shared on YouTube	• The popularity of the videos was evaluated with the Video Power Index. • The educational quality of the videos was measured using the DISCERN score, JAMAe benchmark criteria, and GQSf. The technical quality was measured by the SGSSg.	• The DISCERN, JAMA benchmark criteria, GQS, and SGSS evaluations of academic-sourced videos yielded significantly higher scores than patient-sourced videos (*P*<.001, *P*<.001, *P*=.001, and *P*<.001, respectively). • However, the Video Power Index evaluation of patient-sourced videos yielded significantly higher scores than academic- and physician-sourced videos (*P*<.001 and *P*=.003, respectively). • Negative correlations between the Video Power Index and the JAMA, GQS, and SGSS scores were found.
Erdem et al (2018) [37]	Observational study	Bariatric surgery videos (n=175) on YouTube	• Usefulness of bariatric surgery videos on YouTube: A usefulness score (very useful, useful, or not useful)	• Of the 175 videos, 94 (53.7%) were useful, and 43 (24.6%) were very useful. No videos were found containing misleading information. • A Spearman rank correlation found no significant correlation between the usefulness score and the number of views (ρ=−0.118; *P*=.12), number of likes (ρ=−0.038; *P*=.61), number of dislikes (ρ=−0.003; *P*=.97) or video length (ρ=−0.106; *P*=.16).
Biggs et al (2013) [38]	Observational study	YouTube as a source of information on rhinosinusitis	• Videos (n=100) were graded on their ability to inform the layperson about rhinosinusitis.	• 45% of the videos were deemed to provide some useful information. 55% of the videos contained little or no useful facts, 27% of which contained potentially misleading or even dangerous information. Videos uploaded by medical professionals or those from health information websites contained more useful information than those uploaded by independent users. • No *P* values stated.
Kwok et al (2017) [39]	Observational study	Videos available on YouTube pertaining to interventional treatment for varicose veins	• Informational and scientific quality (good, fair, and poor) and stance (for, neutral, against, and unclear) toward the treatment option discussed, treatment type, and video source.	• The largest group of videos (47.3%) received a quality rating of fair, meaning that they discussed one or two aspects of a treatment option, such as procedural technique and indications. Among those videos rated poor (25%), nearly all (98.2%) failed to mention a specific treatment. • No *P* values stated.
Bademci et al (2017) [40]	Observational study	YouTube videos on deep venous thrombosis	• Scientific content, accuracy, and currency	• Although most of the videos uploaded (25/111, 22.9%) were created by physicians, the number of views for website-based videos was significantly higher (*P*=.002). When the uploaded videos were assessed in terms of their usefulness, the videos from physicians and hospitals were statistically more useful than the other videos (*P*<.001).
Visuals or pictographs				

Study	Study type	Content	Outcomes pertaining to quality of web-based health care patient content	Main results
Christensen et al (2017) [41]	Pilot study	Doodle Health: A crowdsourcing web-based game for the co-design and testing of pictographs to reduce disparities in health care communication	• To test the usability of the game and its appeal to health care consumers in the co-design and evaluation of pictographs.	• Initial testing indicates that crowdsourcing is a promising approach to pictograph development and testing for relevancy and comprehension. More than 596 drawings were collected, and 1758 guesses were performed to date with 70%-90% accuracy. • No *P* values stated.
Device on which content is accessed				

Study	Study type	Content	Outcomes pertaining to quality of web-based health care patient content	Main results
Gogovor et al (2017) [42]	Literature review and qualitative focus group study	Development of an internet-based chronic pain self-management program	• Information needs and gaps in chronic pain management as well as technology features to inform the development of an internet-based self-management program	• The gaps identified in terms of chronic pain management included lack of knowledge, limited access to health care, substandard care, and scarce self-management support. • The focus group themes included patient education on chronic pain care, attitude-belief-culture, financial and legal issues, and motivational content. • No *P* values stated.
Lüchtenberg et al (2008) [43]	Observational study	Websites containing medical information addressing laymen or patients (n=139)	• Accessibility using a quantitative checklist based upon the Web Content Accessibility Guidelines of the World Wide Web Consortium	• Of the 139 sites, 25 (17.9%) of the sites were categorized as WAIh guidelines level A or AA. • WAI guidelines level AA was reached by 0.7% (1/139) of website. • None of the websites reached level AAA. • Of the 139 sites, 82% (114) of the assessed consumer websites were not completely accessible to persons who are visually impaired. • No *P* values stated.
Bashi et al (2018) [44]	Systematic review	Patient educational interventions using mobile apps	• The reporting quality of the studies was assessed according to the mHealthi evidence and predefined reporting assessment criteria.	• Of the 15 studies, none reported on the data security, privacy, and confidentiality measures. • No *P* values stated.
Noel et al (2017) [45]	Prospective cohort study	A mobile medical app was developed to improve postoperative care of patients who had undergone plastic surgery	• The content, design, and efficacy of the app were assessed with a questionnaire (n=60).	• The participants reported that the questions regarding postoperative management were addressed effectively, with a mean score of 4.1/5. • Most of the participants recommended the app to patients who had undergone plastic surgery, with a mean score of 4.6/5. • The app's information prevented 12 patients from calling the plastic surgeon or the emergency department unnecessarily. • No *P* values stated.
Nicholas et al (2015) [46]	Systematic review	Mobile apps for bipolar disorder	• The comprehensiveness and quality of information was assessed against core psychoeducation principles and current bipolar disorder treatment guidelines. • The management tools were evaluated with reference to the best practice resources for the specific area. • General app features and privacy and security.	• Informative apps covered more than a third (4/11, 36%) of core psychoeducation principles and 15% (2/13) of best practice guidelines. • A third (10/32, 31%) cited their sources. • "Neither comprehensiveness of psychoeducation information (ρ=−0.11; *P*=.80) nor adherence to best practice guidelines (ρ=−0.02; *P*=.96) were significantly correlated with average user ratings."
Jamison et al (2017) [47]	Randomized controlled trial	To test an app that enables patients with chronic pain to assess, monitor, and communicate their status to their health care provider.	• Frequency of app use and app satisfaction scores	• In total, 78.1% (82/105) of the participants reported daily using the app. • Patient satisfaction survey results: • Ease of use: 1.8/10 (0=very easy to use, 10=unusable) • Willingness to use after the study: 2.4/10 (0=very willing; 10=unwilling. • Participants with more daily assessments reported higher app satisfaction (*P*<.05) than those who used the app less.

Study	Study type	Content	Outcomes pertaining to quality of web-based health care patient content	Main results
Schulz et al (2007) [48]	Randomized controlled trial	Website designed to enhance self-management in chronic lower back pain	• Change in pain levels, change in knowledge, behavioral changes, and medication use	• Users accessed the website an average of 11.5 times during the 5-month study. • Mean pain levels fell in the control group from 5 to 3.9 (10=most severe pain imaginable, 1=no pain), whereas the mean pain levels in the control group remained largely the same (6.1 to 6.3). • No *P* values stated.
Caiata Zufferey et al (2009) [49]	Observational study	Website *Oneself* designed to promote self-management and inform patients on lower back pain management	• Self-comprehension • Improvement of vocabulary, knowledge of exercises, self-confidence, and motivation	• Of the 129 survey participants, 32 (24.8%) reported that Oneself increased their knowledge about back pain. • Successful testimonials indicated that self-management was encouraged. • No *P* values stated.
Hagerman et al (2017) [50]	Observational study	DAs[j] for patients with low-risk PCa[k]	• What are the informational needs of patients with low-risk PCa, and what are the essential aspects of treatment DAs that increase the likelihood of physicians recommending them to the patient?	• Semistructured interviews found that "Physicians highlighted the importance of patient education and described the characteristics of a low-risk PCa DA that would increase the likelihood of its use in clinical practice." • Encourage patients to take their time in decision-making. • Frankly inform on posttreatment side effects. • Incorporate physician recommendations on content and mode of delivery. • No *P* values stated.
Kim et al (2002) [51]	Observational study	Website comprising a situational approach to the organization of disease-specific patient information	• Interface usability, personal relevance of retrieved information, comprehension of retrieved information.	• Responses (n=37) yielded high ratings for the following: interface usability (4.6/5); personal relevance of information found (4.7/5); comprehension of information (4.8/5). • No *P* values stated.
Meppelink et al (2015) [52]	Observational study	Colorectal cancer screening messages divided into high-literacy and low-literacy groups, with and without illustrations	• Information recall, attitudes, intention to undergo screening	• Spoken messages about colorectal cancer screening improved recall (*P*=.03) and attitudes (*P*=.02) compared with written messages in individuals with lower health literacy. • Animations alone failed to improve recall, but when combined with spoken text, they significantly improved recall (*P*=.02).
Mayer et al (2003) [53]	Literature review	A theory designed to format multimedia content to optimize patient education	• Overloading, speed of content delivery, and misalignment of textual and visual cues	• Narration has better transfer of information than on-screen text. • Learner-controlled segments increase transfer of information. • Graphics and corresponding text should be aligned visually. • Signals also improve transfer. • No *P* values stated.
Stage of patient journey				

Study	Study type	Content	Outcomes pertaining to quality of web-based health care patient content	Main results
Biernatzki et al (2018) [54]	Cross-sectional descriptive	Evaluation of the informational needs of patients with diabetes	• Treatment process, course of disease, abnormalities of glucose metabolism, and diabetes through the life cycle	• Information needs among patients with diabetes is poorly investigated, although in high demand. • No *P* values stated.
Boyde et al (2009) [55]	Observational study	An investigation of the learning style and learning needs of patients with HF[l]	• Questionnaire identifying preferred learning modalities	• In total, 64% (55/86) of the participants reported a preference for multimodal learning style; 18% (15/86) preferred textual information; 11% (9/86) preferred auditory; and 7% (6/86) preferred kinesthetic. • Signs and symptoms were ranked as the most important topics to learn about. • No *P* values stated.
Hagenhoff et al (1994) [56]	Systematic review	Evaluation of the perceptions of both patients and nurses on the importance of educational content for patients with congestive HF	• Questionnaire evaluating the importance of the following categories: anatomy and physiology; psychology; risk factors; medications, diet, and activity; and other	• Patients and nurses rated all information as moderately to very important to learn. • Patients often rated information as more important than nurses did. • No *P* values stated.
Wehby et al (1999) [57]	Descriptive comparative study	Perceptions of RNs[m] and patients concerning educational content on HF were analyzed	• Ranking of categories of HF education in order of importance by patients and RNs	• "Patients perceived the subscales of general HF information, risk factors, medications, prognosis, and signs and symptoms as more important to learn than the RNs (*P*<.05)." • "Patients perceived diet information as less important to learn than the RNs (*P*<.05)." • "Patients perceived all eight subscales as more realistic to learn than the RNs (*P*<.05). Although not in identical order, both groups ranked education related to medication and signs and symptoms as the 2 priority areas." • "Diet information was ranked eighth by the patients and third by the RNs." • No *P* values stated.
Clark et al (2004) [58]	Descriptive correlational study	Examination of perceived learning needs of patients with heart failure after discharge	• "The Outpatient Heart Failure Learning Needs Inventory was used to rate the participants' perceptions of the importance of educational topics on a 5-point Likert scale."	• "The findings indicated that the subjects perceived signs and symptoms and medications as most important to learn and diet, activity, and psychological factors as least important to learn. These findings are consistent with previous research and provide a framework on which to base the development of educational programs for patients with heart failure. A significant finding was that nearly 25% of the screened patients were unable to participate because they were unaware that they had been diagnosed with heart failure." • No *P* values stated.
Kiliç B et al (2015) [59]	Descriptive comparative study	Qualitative analysis of questionnaires examining perceptions of RNs and patients concerning educational content on HF were analyzed.	• Themes related to the educational needs of patients about use of drugs. • Themes related to lifestyle changes. • Themes about the educational needs of the patients related to the characteristics of the disease ranking of categories of HF education in order of importance by patients and RNs.	• "In this study, HF patients stated that they mainly need information about the effects and purposes of the drugs they used. The need for information about the management of the symptoms that affect daily activities are considered 2nd and the educational needs about the disease itself are considered 3rd in importance." • No *P* values stated.

Study	Study type	Content	Outcomes pertaining to quality of web-based health care patient content	Main results
Solomon et al (2018) [60]	Qualitative study	To build an evidence-based web-based patient information resource for patients with HIV	• Transcribed interviews of stakeholders underwent qualitative content analysis	• The interviewees suggested that descriptions of all members of the health care team involved with HIV care be included on the website. • It was also suggested to organize the menu into health challenge categories for ease of navigation. • No *P* values stated.
Liu et al (2017) [61]	Umbrella review	Aimed to identify the current evidence on health education–related interventions for patients with acute coronary syndrome or type 2 diabetes	• Clinical outcomes, behavioral outcomes, psychosocial outcomes, and medical service use	• Nurses and multidisciplinary teams were the most frequent health care professionals to provide education, and most educational interventions were delivered after discharge. • Face-to-face sessions were the most common delivery formats of the patient educational interventions. • The psychoeducational interventions were found to be effective in reducing smoking and admissions for patients with acute coronary syndrome. • No *P* values stated.
Credibility and completeness				

Study	Study type	Content	Outcomes pertaining to quality of web-based health care patient content	Main results
Boyer et al (1998) [62]	Review	Review of *HONcode*[n], a guideline designed to raise the quality of web-based patient education data	• Guidelines to information providers, with the aim of raising the quality of web-based data available and helping to identify websites that are maintained by qualified people and contain reliable data.	• The HONcode mainly includes the following ethical aspects: the author's credentials, the date of the last modification with respect to clinical documents, confidentiality of data, source data reference, funding, and the advertising policy. • No *P* values stated.
Priyanka et al (2018) [63]	Observational study	Evaluation and analysis of web-based content pertaining to esophageal duodenoscopy for patients	• GQS, Health on Net, Flesch-Kincaid Reading Ease, and Flesch-Kincaid Grade Level	• Three websites were found to have high-quality, comprehensive, and authentic information: Healthline, Uptodate, and Emedicine. • In total, 13 sites yielded moderate quality of information. The mean Flesch-Kincaid Reading Ease score was 46.92. • The mean Flesch-Kincaid Grade Level was 11th grade. • No *P* values stated.
Couper et al (2010) [64]	Cross-sectional survey	Analysis of the perceived importance of sources of health information on the web	• Ranking of sources in terms of reliability and influence; use of the internet in age groups.	• Internet use was more common at younger ages, increasing from 14% among those aged 70 years or older to 38% for those aged 40-49 years. • Internet users rated health care providers as the most influential source of information for medical decisions, followed by the internet, family and friends, and media. • No *P* values stated.
Volk et al (2013) [65]	Cross-sectional survey	Evaluation of ongoing studies regarding what the standards for DAs for patients should be	• Voting system to develop criteria for DA standards	• The review comprised 13 manuscripts on topics including current frameworks used to create health care content, health literacy, and the role of patient stories. • No *P* values stated.

[a]SMOG: Simple Measure of Gobbledygook.

[b]PEMAT: Patient Education Materials Assessment Tool.

[c]TRAK: Taxonomy for the Rehabilitation of Knee Conditions.

[d]VaPE: Video-Assisted Patient Education.

[e]JAMA: Journal of the American Medical Association.

[f]GQS: Global Quality Scores.

[g]SGSS: Sleeve Gastrectomy Scoring System.

[h]WAI: Web Accessibility Initiative.

[i]mHealth: mobile health.

[j]DA: decision aid.

[k]PCa: prostate cancer.

[l]HF: heart failure.

[m]RN: Regional nurse.

[n]HONcode: Health on the Net Foundation Code of Conduct.

Textual Information

A study by Ernest et al [12] used the *DISCERN* Tool to evaluate the quality of written information regarding hormone treatment and surgery, which was found to be poor [66]. DISCERN is a validated tool developed by an expert panel through a process of panel debate and health care information analysis. It is noted to be the "first standardised index of quality of consumer health information" [66]. The DISCERN Tool comprises 15 key questions that investigate publication reliability and details of treatment choices, followed by overall judgment of quality. Each question is answered on a 5-point scale ranging from no to yes.

The study by Vivenkanatham et al [18] evaluated textual information on polymyalgia rheumatica. The study concluded that for web-based health care content to be effective, readability must be accessible to people of all literacy levels. Similar

conclusions on readability were also reached in the study by Maciolek et al [20], which reported that patient education materials are most effective when simple language accessible to a wide patient population is used. A quality assessment study of web-based content on rheumatoid arthritis analyzed the readability, applicability, and accessibility of patient education websites [21]. The mean reading level was found to be 12.1 (SD 2.3), according to the Flesch-Kincaid Readability Tool. This tool gives a measure of how difficult a passage written in English is to understand through analyzing factors such as word length, sentence length, and total number of syllables. It provides a grade level according to the US educational system ranging from fifth grade to college graduate [67]. The same study reported that 78% of the websites assessed were easy to navigate [21]. Importantly, only 33% of the websites were assessed to be user friendly for people who are visually or hearing impaired.

An evaluation of web-based information on disorders of sex development noted strengths, including the tendency of webpages to present focused information in chunks and in a logical sequence [12]. A semistructured interview study of general practitioners conducted by Hjelmager et al [13] revealed that health information technology apps require textual information that is targeted to patients and written with the input of health care professionals.

The study by Rofaiel et al [14] assessed the quality of website information about inflammatory bowel disease using the DISCERN score to assess reliability and the relevance of pertinent details. The mean DISCERN score for patient-searched websites was not statistically different from that for physician-recommended websites (3.21 vs 3.62, respectively; *P*=.16). Numerous studies [12,13,15-17] identified that readability, as analyzed by the Flesch-Kincaid Tool, of web-based information exceeded the recommended sixth to eighth grade reading levels [68].

In all, 3 studies evaluated the understandability and actionability of web-based patient information by using the validated Patient Education Materials Assessment Tool [12,17,20]. The tool comprises inventories (one for print and another for audiovisual content) that list desirable and undesirable characteristics of information and produce a numeric value for understandability as well as actionability (ie, how easy it is to act on given information). The scores range from 0% to 100%, and a higher score indicates that the text is more understandable or actionable. Web-based educational content pertaining to disorders of sexual development and tympanostomy tube placement yielded low actionability scores (mean scores 23% and 44.6%, respectively). Interestingly, a study by Maciolek et al [20] found that the Patient Education Materials Assessment Tool understandability score for academic institution patient educational material exceeded that for content created by private institutions (*P*=.02) and content from websites unaffiliated with a urologist (*P*=.01).

Discussion Boards or Web-Based Groups

The study by Cedars et al [27] reported on a study of patients with urethral strictures who participate in web-based discussions and share experiences to gain emotional support and find answers. Patients participating in these web-based groups were more often than not satisfied with their postoperative outcomes. These findings are supported by the study by Teaford et al [28], which explored the experiences of new mothers using a web-based postpartum forum. The study found that web-based forums provided a sense of community and a platform for sharing information. They identified five themes pertaining to the participants' concept of a web-based community: social support, anonymity, in-groups, drama, and entertainment.

However, the findings of a study by Castaneda et al [29] highlight the importance of exercising caution with web-based forums. The study evaluated the content of peripheral arterial disease eHealth forums and found that 15.1% of the medical advice given on such platforms was inconsistent with guidelines. Furthermore, the study found that 10 of the 18 posts related to negative personal experiences with health care providers.

It is important to note that there may be differences in patient engagement when comparing groups that are run by health care (or affiliated) professional service providers with those set up by patients or lay individuals themselves. A study comparing the difference between the effects of peer-led and moderated groups found that moderated groups were often more active and therefore had higher patient engagement.

Video-Based Content

The study by Pedersen et al [31] found that 4 participants who were interviewed after watching a 14-minute educational video on whiplash injuries felt reassured, particularly because the video aligned with information that they had received at the hospital. The study by Finnegan et al [32] reported that people who visited an information hub where videos were embedded in the webpage spent longer than 2 minutes on that webpage, indicating that visitors with average reading capability watch and listen to the video while browsing the text. The study also reported that this information hub's YouTube channel, which featured videos explaining the concept of herd immunity, was particularly successful, with visitors spending more than 6 minutes on that page [32].

In the context of managing health conditions, the study by Button et al [33] found that video-based content was particularly helpful for patients having physiotherapy for knee injuries because they were able to visualize the correct technique. Patient understanding and confidence were found to be improved by this intervention. In a separate study of 80 anesthetists using a Video-Assisted Patient Education intervention, 96% reported that patients felt that they had a better understanding of the information provided through the Video-Assisted Patient Education intervention, and 97.5% of the anesthetists felt that it was a useful form of patient education [34]. However, 46% noted increased anxiety caused by the intervention [34].

Interestingly, the study by Ferhatoglu et al [36] found in an evaluation of the quality of YouTube content on sleeve gastrectomy that patient experience and advertisement videos were significantly more popular than academic videos created by medical professionals, according to the Video Power Index (*P*<.001 and *P*<.003, respectively). The Video Power Index assesses video performance by comparing the video with leaders in its respective industry, measures its impact and effectiveness across all platforms, and facilitates strategies to cater to target

audiences [69]. The study by Erdem and Sisik [37] reported similar findings and found no significant association between video traction (*likes*, *dislikes*, or *views*) and usefulness of the content (Spearman rank correlation ρ=−0.038, *P*=.61; ρ=−0.003, *P*=.97; and ρ=−0.118, *P*=.12, respectively). Interestingly, the study found no significant correlation in usefulness to video length (ρ=−0.106; *P*=.16) in contrast to the findings of the study by Biggs et al [38], which concluded that medical videos categorized as useful had a mean length of 6 minutes and 51 seconds, with the videos rated in the top 10 having a mean length of 14 minutes and 47 seconds.

Similar to previous findings, a study conducted to assess the quality of YouTube videos on cataract surgery concluded that videos created by medical organizations such as the National Health Service were significantly more useful in terms of educating patients about the procedure than videos sourced by independent medical professionals and other sources (*P*<.001) [30]. The study by Bademci et al [40] similarly concluded that medical topic videos on deep vein thrombosis sourced from medical professionals and hospitals were significantly more useful than videos from other sources (*P*<.001).

In a study evaluating YouTube content on varicose veins, it was found that most of the videos were dominated by the private health care sector and that the video content presented a distorted view of treatment options, consequently leading to a skewed patient perception of the therapeutic options available to them [39]. Not only were 32% of these videos found to be of poor quality, but videos from private medical companies were also significantly more likely to favor minimally invasive surgery over ligation and stripping than videos from other sources. Once again, there was no significant association found between quality of content and viewing frequency.

Visuals or Pictographs

Pictographs are pictorial representations of words or phrases. The study by Christensen et al [41] described the value of pictographs in health communication. The study presents the results of building and testing the game *Doodle Health*, which is designed to produce pictographs through crowdsourcing. It found that this method of pictograph creation yielded positive feedback from focus groups with regard to usability and comprehension. Analysis of the feedback found that 62.2% of the participants praised the game, with a respondent describing it as “engaging and easy to use.” However, the study also found that people from diverse cultural backgrounds have different styles of communication, which may render visually presented information ineffective for minority groups. Crowdsourcing pictographs was suggested as a potential solution to this concern.

Device Accessibility

A study by Gogovor et al [42] concluded that the next generation of web-based educational health care programs should integrate apps for reasons of (1) accessibility, (2) flexibility, and (3) security and trustworthiness. Almost all the studies in our review used web-based platforms and required devices that accessed the web, with some studies requiring devices that accessed the video-sharing platform, YouTube, specifically.

A study by Lüchtenberg et al [43] found that only 18% (25/139) of the health information websites evaluated achieved a high standard of accessibility for users who are visually impaired as assessed by the Web Accessibility Initiative guidelines. The study concluded that web-based educational health care content should at least meet the requirements of priority 1 (level A) and preferably priority 2 (level AA) of the Web Accessibility Initiative guidelines. Developers can ensure accessibility by, for instance, having text alternatives for all nontext content and using high contrast ratios of text and images [36].

Few studies used smartphones exclusively to deliver information. Describing the benefits of using a smartphone, the study by Bashi et al [44] noted that smartphone adoption is becoming widespread, meaning more people can be reached; however, studies on how apps should deliver information in terms of interface and content is particularly lacking. The study by Noel et al [45] assessed the impact of a mobile medical app on plastic surgery patient care. A total of 60 patients answered questionnaires pertaining to the app, and the results supported the app’s utility from the patient’s perspective. The mean score for the app’s ability to answer patient questions was 4.1/5, and 20% of the patients were prevented from unnecessarily calling the emergency department. The use of a smartphone also enabled patients to access information from different locations and allowed for real-time disease management. A systematic review by Nicholas et al [46] revealed that of the 11 studies of apps providing patient information, only a third covered core psychoeducational principles. Furthermore, the average user ratings were not correlated with either comprehensiveness or adherence to best practice guidelines (ρ=−0.11, *P*=.80 and ρ=−0.02, *P*=.96, respectively).

Stage of Patient Journey

A qualitative study by Gogovor et al [42] using health care professionals and patient focus groups found that health care professionals suggested stratifying access to information according to *stage of advancement* in terms of their chronic pain management. A qualitative study of *Oneself*, a website designed for use by patients with chronic lower back pain, found it to be an effective adjunct to doctor-patient consultations because patients found that they were unable to have all their queries answered in the limited time of a consultation [49].

At different stages of a patient journey, decision aids may be used as adjunctive decision-making tools to support patients. In the study by Hagerman et al [50], the authors conducted semistructured interviews of 33 physicians to identify the desirable characteristics of decision aids. Of the 33 physicians, 20 (61%) stated that patients should be educated on the lack of urgency with regard to making a treatment decision. Of the 33 physicians, 28 (85%) agreed that decision aids should be provided to the patient after the consultation when the patient is at home. Furthermore, 36% (12/33) of the physicians deemed it *very* or *extremely* important that decision aids are designed to be used during and after consultations.

Tailoring information delivery to the stage of the patient journey is further supported by the results of the study by Kim et al [51] who developed a system to organize web-based disease-specific information according to a situational knowledge base model.

The approach categorizes information about a specific disease (eg, thyroid cancer) into sections corresponding to discrete clinical events (eg, presentation, fine-needle aspiration biopsy, and diagnosis). In all, 75 patients completed a questionnaire evaluating the website, which found mean usability to be 4.6/5, personal relevance of received information 4.7/5, and comprehension of received information 4.8/5.

Credibility and Completeness of Information

A study by Jamison et al [47] found that 86% of the apps for pain conditions reported were created with no involvement by health care professionals. A comprehensive study by Bae et al [30], assessing the quality of the content of YouTube videos for cataract surgery patient education, found that there was an abundance of videos simply showing patients undergoing a live procedure. More than 20% of the educational videos were commercial and hence potentially misleading. This may make it challenging to find high-quality, comprehensive educational videos on the web.

A study by Pithadia et al [35] used the American Academy of Dermatology guidelines as a benchmark to evaluate the accuracy of patient information YouTube videos on psoriasis treatments. It concluded that 12% of the videos contained high-quality patient education content, and most of them were not patient-centric. Similarly, the study by Ferhatoglu et al [36] used the *Journal of American Medical Association* benchmark criteria to assess the educational quality of sleeve gastrectomy YouTube videos and found that this score was significantly higher in university-affiliated physician videos than in other videos (P<.001).

The Health on the Net Foundation Code of Conduct (HONcode) presents a set of eight principles designed to set the quality standard for web-based patient information [62] (Textbox 1). A study by Laversin et al [70] compared 165 HONcode-certified websites with 165 noncertified websites. Only 0.6% of the noncertified websites conformed to the principles of the HONcode compared with 89% of the certified sites (P<.10). As the study followed certified websites 6 months after certification, the effect of the HONcode certification shows short-term sustainability.

Textbox 1. Quality standard (adapted from the study by Laversin et al [70]).

Eight Principles Designed to Set the Quality Standard for Web-Based Patient Information

1. Authoritative: qualifications of the authors indicated
2. Complementarity: information should support the doctor-patient relationship
3. Privacy: personal data collected by the site kept private
4. Attribution: cite all references
5. Justifiability: back up claims relating to benefits and performance
6. Transparency: accessible presentation, accurate email contact
7. Financial disclosure: identify funding sources
8. Advertising policy: clearly distinguish advertising from editorial content

Discussion

Principal Findings

In today's world of access to knowledge often being initially web-based, it is of importance for health care professionals to be able to create effective content. This is further emphasized in the current environment where minimum contact between patients and health care providers is required. We performed a wide scoping review of the literature to identify the features of web-based content and other telemedicine requirements that may improve quality of engagement with web-based health care content in this growing field. Using these results, we have developed a framework (Figure 2) to facilitate the development of web-based patient health care content. All the categories except for one (visual or pictograph) were reported on by 3 or more papers. We included the visual or pictograph category after discussion because it was felt to be a relevant and important means of communicating information. Although the features identified are, on reflection, intuitive, the framework arms the content creator with the best available strategies in making the content engaging and hence effective.

Figure 2. Infographic framework for modes of engagement for web-based health care content. HONcode: Health on the Net Foundation Code of Conduct; PEMAT: Patient Education Materials Assessment Tool.

Other similar frameworks such as the International Patient Decision Aid Standards (IPDAS) and the Standards for Universal Reporting of Patient Decision Aid Evaluation Studies (SUNDAE) checklists [65,71] have been developed for the evaluation of decision aids. However, to the best of our knowledge, this is the first framework to specifically focus on engagement with web-based content rather than a checklist approach to ensuring that decisions are made with appropriate consideration given to all relevant issues and options. A study design published by Knerr et al [72] aimed to evaluate patient behavior in response to a patient decision aid based on the IPDAS standards but has not reported results. Although patient decision aids can be a form of web-based content, efforts up to now have been directed toward ensuring transparency and trust in imparting information rather than ensuring the ability of web-based material to engage the user [73]. The need to engage

people in health care content has been overlooked. We propose that this framework be used to improve engagement, which in turn will improve adherence with medical treatment and hence improve outcomes. Nevertheless, the rigorous process used by the IPDAS and SUNDAE developers is one that we would be interested in following in refining and adapting the framework arrived at through this literature review.

Further work will involve a co-design process with all stakeholders (including patients) to refine the insights we have gained from the studies regarding engagement with web-based content. The personalization of health care content may have bearing on the issue of engagement. Studies have shown that it is possible to *segment* the population according to the likelihood of responding to health care messages [74,75]. Although our framework provides the best evidence available relating to engagement with web-based content, the holy grail may lie in developing further the field of psychographics for health care. Although researchers have investigated the way that segmentation affects a defined intervention, the effect of segmentation itself is yet to be assessed in a meaningful manner [75]. The internet is able to bring together varied but related content using the concept of the semantic web and the application of folksonomies [76]. The confluence of Web 3.0 (to crowdsource content relevant to a desired health care behavior), psychographic segmentation (including segmentation based on the proposed framework), and machine learning may provide a way forward. We have developed a Web 3.0 health care content platform (Health Shared) and intend to use it for this purpose.

Limitations

The findings of this review should be considered in the context of several limitations. The principal limitation is that most of the components of this suggested framework are not supported by strong evidence. The studies were heterogeneous in their aims, interventions, and outcomes, and some were of poor methodological quality. Few studies discussed the effect of the platform used on patient engagement—for example, smartphone app versus website versus commercial health care information—which may play a role in patient engagement. However, given that most of the studies discussed in the review were general scans of available websites, the framework developed is largely applicable to the website development platform. In addition, few studies describe the differences between informational sites and other modes-of-engagement systems that provide the ability for patient input and enable patients to contact their provider or providers; therefore, we were not able to compare these patient information platforms.

Furthermore, because only the PubMed database was searched, studies are likely to have been omitted from this review. Despite this limitation, the components of the framework are intuitive, and we believe that its application may be beneficial to health care providers and content creators. Evaluation and subsequent validation of the proposed framework by key stakeholders, including patients, clinicians, and content creators, would increase the robustness.

Conclusions

There is a paucity of high-quality data relating to the factors that improve users' quality of engagement with web-based health care content. Our framework summarizes the reported studies, which may be useful to health care content creators. Evaluation of the utility of web-based content to engage users is of significant importance and may be accessible through tools such as the Net Promoter score. Web 3.0 technology and development of the field of psychographics for health care offer further potential for development [75]. Future work may also involve improvement of the framework through a co-design process.

Although there are often specific health care issues needing to be addressed in response to crisis situations, we believe that this work is more generally important in facilitating patient activation and patient-supported self-management, which are two major pillars in how health care systems need to realign to keep up with increasing demand.

Authors' Contributions

UJ conceived the idea; LAO, EA, AA, NH, SL, and PN wrote the manuscript; and UJ, LAO, and PN reviewed and edited the manuscript.

Conflicts of Interest

UJ is the founder of Health Shared.

References

1. Dutton WH, Blank G. Next generation users: The internet in Britain. SSRN J 2012:1960655. [doi: 10.2139/ssrn.1960655]
2. O'Neill B, Ziebland S, Valderas J, Lupiáñez-Villanueva F. User-generated online health content: a survey of internet users in the United Kingdom. J Med Internet Res 2014 Apr 30;16(4):e118 [FREE Full text] [doi: 10.2196/jmir.3187] [Medline: 24784798]
3. Pellisé F, Sell P, EuroSpine Patient Line Task Force. Patient information and education with modern media: the Spine Society of Europe Patient Line. Eur Spine J 2009 Aug 19;18 Suppl 3(S3):395-401 [FREE Full text] [doi: 10.1007/s00586-009-0973-1] [Medline: 19381695]
4. Outreville JF, Kongstvedt PR. The managed health care handbook. J Risk Ins 1998 Jun;65(2):351. [doi: 10.2307/253545]

5. Entwistle VA, France EF, Wyke S, Jepson R, Hunt K, Ziebland S, et al. How information about other people's personal experiences can help with healthcare decision-making: a qualitative study. Patient Educ Couns 2011 Dec;85(3):291-298. [doi: 10.1016/j.pec.2011.05.014] [Medline: 21652162]
6. Kamel BM, Wheeler S. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Info Libr J 2007 Mar;24(1):2-23 [FREE Full text] [doi: 10.1111/j.1471-1842.2007.00701.x] [Medline: 17331140]
7. Bader JL, Strickman-Stein N. Evaluation of new multimedia formats for cancer communications. J Med Internet Res 2003 Aug 29;5(3):e16 [FREE Full text] [doi: 10.2196/jmir.5.3.e16] [Medline: 14517107]
8. Marcus C. Strategies for improving the quality of verbal patient and family education: a review of the literature and creation of the EDUCATE model. Health Psychol Behav Med 2014 Jan 01;2(1):482-495 [FREE Full text] [doi: 10.1080/21642850.2014.900450] [Medline: 25750796]
9. Dekkers T, Melles M, Groeneveld BS, de Ridder H. Web-based patient education in orthopedics: systematic review. J Med Internet Res 2018 Apr 23;20(4):e143 [FREE Full text] [doi: 10.2196/jmir.9013] [Medline: 29685869]
10. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
11. Chedid RA, Terrell RM, Phillips KP. Best practices for online Canadian prenatal health promotion: a public health approach. Women Birth 2018 Aug;31(4):223-231. [doi: 10.1016/j.wombi.2017.10.005] [Medline: 29113753]
12. Ernst M, Chen D, Kennedy K, Jewell T, Sajwani A, Foley CB, DSD-TRN Psychosocial WorkgroupAccord Alliance. Disorders of sex development (DSD) web-based information: quality survey of DSD team websites. Int J Pediatr Endocrinol 2019;2019:1 [FREE Full text] [doi: 10.1186/s13633-019-0065-x] [Medline: 31149017]
13. Hjelmager DM, Vinther LD, Poulsen SH, Petersen LS, Jensen MB, Riis A. Requirements for implementing online information material for patients with low back pain in general practice: an interview study. Scand J Prim Health Care 2019 Mar 31;37(1):60-68 [FREE Full text] [doi: 10.1080/02813432.2019.1569413] [Medline: 30700191]
14. Rofaiel R, Chande N. Biologic agents in inflammatory bowel disease – quality of internet website information. Saudi J Gastroenterol 2018;24(6):336. [doi: 10.4103/sjg.sjg_55_18]
15. Alfonso AR, DeMitchell-Rodriguez EM, Ramly EP, Noel DY, Levy-Lambert D, Wang MM, et al. Assessment of American cleft palate-craniofacial association-approved teams' websites for patient-oriented content and readability. Cleft Palate Craniofac J 2019 Oct 26;56(9):1213-1219. [doi: 10.1177/1055665619850441] [Medline: 31129984]
16. Ayyaswami V, Padmanabhan D, Patel M, Prabhu AV, Hansberry DR, Agarwal N, et al. A readability analysis of online cardiovascular disease-related health education materials. Health Lit Res Pract 2019 Apr 01;3(2):74-80 [FREE Full text] [doi: 10.3928/24748307-20190306-03] [Medline: 31049489]
17. Fajardo MA, Weir KR, Bonner C, Gnjidic D, Jansen J. Availability and readability of patient education materials for deprescribing: an environmental scan. Br J Clin Pharmacol 2019 Jul 07;85(7):1396-1406 [FREE Full text] [doi: 10.1111/bcp.13912] [Medline: 30848837]
18. Vivekanantham A, Protheroe J, Muller S, Hider S. Evaluating on-line health information for patients with polymyalgia rheumatica: a descriptive study. BMC Musculoskelet Disord 2017 Jan 26;18(1):43 [FREE Full text] [doi: 10.1186/s12891-017-1416-5] [Medline: 28122554]
19. Harris VC, Links AR, Hong P, Walsh J, Schoo DP, Tunkel DE, et al. Consulting Dr. Google: Quality of online resources about tympanostomy tube placement. Laryngoscope 2018 Dec;128(2):496-501. [doi: 10.1002/lary.26824] [Medline: 28842989]
20. Maciolek KA, Jarrard DF, Abel EJ, Best SL. Systematic assessment reveals lack of understandability for prostate biopsy online patient education materials. Urology 2017 Nov;109:101-106. [doi: 10.1016/j.urology.2017.07.037] [Medline: 28780302]
21. Siddhanamatha HR, Heung E, Lopez-Olivo MD, Abdel-Wahab N, Ojeda-Prias A, Willcockson I, et al. Quality assessment of websites providing educational content for patients with rheumatoid arthritis. Semin Arthritis Rheum 2017 Jun;46(6):715-723. [doi: 10.1016/j.semarthrit.2017.01.006] [Medline: 28258768]
22. Institute of Medicine, Board on Neuroscience and Behavioral Health, Committee on Health Literacy. In: Nielsen-Bohlman L, Panzer A, Kindig D, editors. Health Literacy: A Prescription to End Confusion. Washington, DC: National Academies Press; 2004.
23. Chin J, Moeller D, Johnson J, Duwe EA, Graumlich JF, Murray MD, et al. A multi-faceted approach to promote comprehension of online health information among older adults. Gerontologist 2018 Jul 13;58(4):686-695 [FREE Full text] [doi: 10.1093/geront/gnw254] [Medline: 28329840]
24. Boudewyns V, O'Donoghue AC, Kelly B, West SL, Oguntimein O, Bann CM, et al. Influence of patient medication information format on comprehension and application of medication information: a randomized, controlled experiment. Patient Educ Couns 2015 Jul 10;98(12):1592-1599. [doi: 10.1016/j.pec.2015.07.003] [Medline: 26198545]
25. Morrow D, Weiner M, Young J, Steinley D, Deer M, Murray M. Improving medication knowledge among older adults with heart failure: a patient-centered approach to instruction design. Gerontologist 2005 Aug;45(4):545-552. [doi: 10.1093/geront/45.4.545] [Medline: 16051918]

26. Pander Maat H, Lentz L. Improving the usability of patient information leaflets. Patient Educ Couns 2010 Jul;80(1):113-119. [doi: 10.1016/j.pec.2009.09.030] [Medline: 19854022]
27. Cedars BE, Cohen AJ, Fergus KB, Baradaran N, Ndoye M, Kamal P, et al. Qualitative analysis of the content found in online discussion boards for urethral stricture disease and urethroplasty. Urology 2019 Aug;130:155-161. [doi: 10.1016/j.urology.2019.03.033] [Medline: 31129196]
28. Teaford D, McNiesh S, Goyal D. New mothers' experiences with online postpartum forums. MCN Am J Matern Child Nurs 2019;44(1):40-45. [doi: 10.1097/NMC.0000000000000489] [Medline: 30444739]
29. Castaneda P, Sales A, Osborne NH, Corriere MA. Scope, themes, and medical accuracy of eHealth peripheral artery disease community forums. Ann Vasc Surg 2019;54:92-102. [doi: 10.1016/j.avsg.2018.09.004] [Medline: 30267913]
30. Bae SS, Baxter S. YouTube videos in the English language as a patient education resource for cataract surgery. Int Ophthalmol 2018 Oct 28;38(5):1941-1945. [doi: 10.1007/s10792-017-0681-5] [Medline: 28849436]
31. Pedersen MM, Fink P, Kasch H, Frostholm L. Development of an internet-delivered educational video for acute whiplash injuries. Pilot Feasibility Stud 2019 Apr 27;5(1):60 [FREE Full text] [doi: 10.1186/s40814-019-0445-2] [Medline: 31061715]
32. Finnegan G, Holt D, English PM, Glismann S, Thomson A, Salisbury DM, et al. Lessons from an online vaccine communication project. Vaccine 2018 Oct 22;36(44):6509-6511. [doi: 10.1016/j.vaccine.2018.05.007] [Medline: 29921491]
33. Button K, Nicholas K, Busse M, Collins M, Spasić I. Integrating self-management support for knee injuries into routine clinical practice: TRAK intervention design and delivery. Musculoskelet Sci Pract 2018 Feb;33:53-60 [FREE Full text] [doi: 10.1016/j.msksp.2017.11.002] [Medline: 29172113]
34. Vogel H, Schaaf W, Jacob M. [Video-assisted patient education in anesthesia : how do medical users assess the procedure?]. Anaesthesist 2018 Nov 9;67(11):829-836. [doi: 10.1007/s00101-018-0496-4] [Medline: 30302517]
35. Pithadia DJ, Reynolds KA, Lee EB, Wu JJ. A cross-sectional study of YouTube videos as a source of patient information about topical psoriasis therapies. J Dermatolog Treat 2019:1-4. [doi: 10.1080/09546634.2019.1597247] [Medline: 30885022]
36. Ferhatoglu MF, Kartal A, Ekici U, Gurkan A. Evaluation of the reliability, utility, and quality of the information in sleeve gastrectomy videos shared on open access video sharing platform YouTube. Obes Surg 2019 May;29(5):1477-1484. [doi: 10.1007/s11695-019-03738-2] [Medline: 30706318]
37. Erdem H, Sisik A. The reliability of bariatric surgery videos in YouTube platform. Obes Surg 2018 Mar 5;28(3):712-716. [doi: 10.1007/s11695-017-2911-3] [Medline: 28875396]
38. Biggs TC, Bird JH, Harries PG, Salib RJ. YouTube as a source of information on rhinosinusitis: the good, the bad and the ugly. J Laryngol Otol 2013 Aug;127(8):749-754. [doi: 10.1017/S0022215113001473] [Medline: 23866821]
39. Kwok TM, Singla AA, Phang K, Lau AY. YouTube as a source of patient information for varicose vein treatment options. J Vasc Surg Venous Lymphat Disord 2017 Mar;5(2):238-243. [doi: 10.1016/j.jvsv.2016.10.078] [Medline: 28214492]
40. Bademci M, Yazman S, Güneş T, Ocakoglu G, Tayfur K, Gokalp O. YouTube as a potential source of information on deep venous thrombosis. Phlebology 2017 Sep 12;32(8):543-547. [doi: 10.1177/0268355516673795] [Medline: 27738240]
41. Christensen C, Redd D, Lake E, Shipman JP, Aiono H, Altizer R, et al. Doodle Health: A crowdsourcing game for the co-design and testing of pictographs to reduce disparities in healthcare communication. AMIA Annu Symp Proc 2017;2017:585-594 [FREE Full text] [Medline: 29854123]
42. Gogovor A, Visca R, Auger C, Bouvrette-Leblanc L, Symeonidis I, Poissant L, et al. Informing the development of an internet-based chronic pain self-management program. Int J Med Inform 2017 Jan;97:109-119. [doi: 10.1016/j.ijmedinf.2016.10.005] [Medline: 27919369]
43. Lüchtenberg M, Kuhli-Hattenbach C, Sinangin Y, Ohrloff C, Schalnus R. Accessibility of health information on the internet to the visually impaired user. Ophthalmologica 2008;222(3):187-193. [doi: 10.1159/000126082] [Medline: 18497528]
44. Bashi N, Fatehi F, Fallah M, Walters D, Karunanithi M. Self-management education through mhealth: review of strategies and structures. JMIR Mhealth Uhealth 2018 Oct 19;6(10):e10771 [FREE Full text] [doi: 10.2196/10771] [Medline: 30341042]
45. Noel W, Bosc R, Jabbour S, Kechichian E, Hersant B, Meningaud J. Smartphone-based patient education in plastic surgery. Ann Plast Surg 2017 Dec;79(6):529-531. [doi: 10.1097/SAP.0000000000001241] [Medline: 29053521]
46. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. J Med Internet Res 2015 Aug;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
47. Jamison RN, Jurcik DC, Edwards RR, Huang C, Ross EL. A pilot comparison of a smartphone app with or without 2-way messaging among chronic pain patients: who benefits from a pain app? Clin J Pain 2017 Aug;33(8):676-686 [FREE Full text] [doi: 10.1097/AJP.0000000000000455] [Medline: 27898460]
48. Schulz PJ, Rubinell S, Hartung U. An internet-based approach to enhance self-managementof chronic low back pain in the italian-speaking population of Switzerland: results from a pilot study. Int J Public Health 2007 Oct;52(5):286-294. [doi: 10.1007/s00038-007-5127-9] [Medline: 18030944]
49. Zufferey MC, Schulz PJ. Self-management of chronic low back pain: an exploration of the impact of a patient-centered website. Patient Educ Couns 2009 Oct;77(1):27-32. [doi: 10.1016/j.pec.2009.01.016] [Medline: 19321286]
50. Hagerman C, Bellini P, Davis K, Hoffman RM, Aaronson DS, Leigh DY, et al. Physicians' perspectives on the informational needs of low-risk prostate cancer patients. Health Educ Res 2017 Apr 01;32(2):134-152 [FREE Full text] [doi: 10.1093/her/cyx035] [Medline: 28380628]

51. Kim MI, Ladenson P, Johnson KB. A situational approach to the design of a patient-oriented disease-specific knowledge base. Proc AMIA Symp 2002:385-389 [FREE Full text] [Medline: 12463852]
52. Meppelink CS, Smit EG, Buurman BM, van Weert JCM. Should we be afraid of simple messages? The effects of text difficulty and illustrations in people with low or high health literacy. Health Commun 2015 Sep 15;30(12):1181-1189. [doi: 10.1080/10410236.2015.1037425] [Medline: 26372031]
53. Mayer RE, Moreno R. Nine ways to reduce cognitive load in multimedia learning. Edu Psychol 2003 Mar;38(1):43-52. [doi: 10.1207/S15326985EP3801_6]
54. Biernatzki L, Kuske S, Genz J, Ritschel M, Stephan A, Bächle C, et al. Information needs in people with diabetes mellitus: a systematic review. Syst Rev 2018 Dec 14;7(1):27 [FREE Full text] [doi: 10.1186/s13643-018-0690-0] [Medline: 29444711]
55. Boyde M, Tuckett A, Peters R, Thompson DR, Turner C, Stewart S. Learning style and learning needs of heart failure patients (The Need2Know-HF patient study). Eur J Cardiovasc Nurs 2009 Dec;8(5):316-322. [doi: 10.1016/j.ejcnurse.2009.05.003] [Medline: 19520614]
56. Hagenhoff BD, Feutz C, Conn VS, Sagehorn KK, Moranville-Hunziker M. Patient education needs as reported by congestive heart failure patients and their nurses. J Adv Nurs 1994 Apr;19(4):685-690. [doi: 10.1111/j.1365-2648.1994.tb01139.x] [Medline: 8021389]
57. Wehby D, Brenner PS. Perceived learning needs of patients with heart failure. Hear Lung J Acute Crit Care 1999 Jan;28(1):31-40. [doi: 10.1016/s0147-9563(99)70041-5]
58. Clark JC, Lan VM. Heart failure patient learning needs after hospital discharge. Appl Nurs Res 2004 Aug;17(3):150-157. [doi: 10.1016/j.apnr.2004.06.009] [Medline: 15343548]
59. Kiliç B, Sütçü Çiçek H. Assessment of educational needs of heart failure patients in intensive care units: qualitative study. Intensive Care Med Exp 2015 Oct 1;3(S1):A208. [doi: 10.1186/2197-425x-3-s1-a208]
60. Solomon P, Salbach NM, O'Brien KK, Nixon S, Worthington C, Baxter L, et al. Collaborative development of an educational resource on rehabilitation for people living with HIV. Disabil Rehabil 2018 Nov 12;40(22):2671-2676. [doi: 10.1080/09638288.2017.1347719] [Medline: 28701083]
61. Liu X, Shi Y, Willis K, Wu C, Johnson M. Health education for patients with acute coronary syndrome and type 2 diabetes mellitus: an umbrella review of systematic reviews and meta-analyses. BMJ Open 2017 Oct 16;7(10):e016857 [FREE Full text] [doi: 10.1136/bmjopen-2017-016857] [Medline: 29042383]
62. Boyer C, Selby M, Appel R. The health on the net code of conduct for medical and health web sites. In: Cesnik B, McCray A, Scherrer J, editors. MEDINFO '98 - Vol 52 - Studies in Health Technology and Informatics. Amsterdam: IOS Press; 1998.
63. Priyanka P, Hadi Y, Reynolds G. Analysis of the patient information quality and readability on Esophagogastroduodenoscopy (EGD) on the internet. Am J Gastroenterol 2018;113(Supplement):289-290. [doi: 10.14309/00000434-201810001-00504]
64. Couper MP, Singer E, Levin CA, Fowler FJ, Fagerlin A, Zikmund-Fisher BJ. Use of the internet and ratings of information sources for medical decisions: results from the DECISIONS survey. Med Decis Making 2016 Dec 04;30(5_suppl):106-114. [doi: 10.1177/0272989x10377661]
65. Volk RJ, Llewellyn-Thomas H, Stacey D, Elwyn G. Ten years of the International Patient Decision Aid Standards Collaboration: evolution of the core dimensions for assessing the quality of patient decision aids. BMC Med Inform Decis Mak 2013;13 Suppl 2:S1 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S1] [Medline: 24624947]
66. Discern online. URL: http://www.discern.org.uk/ [accessed 2020-03-24]
67. Flesch–Kincaid readability tests. Wikipedia. URL: https://en.wikipedia.org/wiki/Flesch%E2%80%93Kincaid_readability_tests#Flesch%E2%80%93Kincaid_grade_level [accessed 2020-03-24]
68. Badarudeen S, Sabharwal S. Assessing readability of patient education materials: current role in orthopaedics. Clin Orthop Relat Res 2010 Oct;468(10):2572-2580 [FREE Full text] [doi: 10.1007/s11999-010-1380-y] [Medline: 20496023]
69. Improve video performance with Shareablee's new Video Power Index (VPI). Shareablee. 2020. URL: https://www.shareablee.com/blog/2017/06/27/improve-video-performance-with-shareablee-s-new-video-power-index-vpi [accessed 2020-05-05]
70. Laversin S, Boujard V, Gaudinat A, Simonet M, Boyer C. Improving the transparency of health information found on the internet through the honcode: a comparative study. In: Moen A, Andersen S, Aarts J, Hurlen P, editors. User Centred Networked Health Care - Volume 169 - Studies in Health Technology and Informatics. Amsterdam: IOS Press; 2011:654-658.
71. Sepucha KR, Abhyankar P, Hoffman AS, Bekker HL, LeBlanc A, Levin CA, et al. Standards for UNiversal reporting of patient Decision Aid Evaluation studies: the development of SUNDAE checklist. BMJ Qual Saf 2018 May 21;27(5):380-388 [FREE Full text] [doi: 10.1136/bmjqs-2017-006986] [Medline: 29269567]
72. Knerr S, Wernli KJ, Leppig K, Ehrlich K, Graham AL, Farrell D, et al. A web-based personalized risk communication and decision-making tool for women with dense breasts: Design and methods of a randomized controlled trial within an integrated health care system. Contemp Clin Trials 2017 May;56:25-33 [FREE Full text] [doi: 10.1016/j.cct.2017.02.009] [Medline: 28257920]
73. Krol MW, de Boer D, Delnoij DM, Rademakers JJ. The Net Promoter Score--an asset to patient experience surveys? Health Expect 2015 Dec;18(6):3099-3109. [doi: 10.1111/hex.12297] [Medline: 25345554]

74. Kannan VD, Veazie PJ. Who avoids going to the doctor and why? Audience segmentation analysis for application of message development. Health Commun 2015 Jul 25;30(7):635-645. [doi: 10.1080/10410236.2013.878967] [Medline: 25062466]
75. Gomez A, Loar R, England Kramer A. The impact of market segmentation and social marketing on uptake of preventive programmes: the example of voluntary medical male circumcision. A literature review. Gates Open Res 2018 Dec 11;2:68 [FREE Full text] [doi: 10.12688/gatesopenres.12888.1] [Medline: 31131368]
76. Jaffer U, Vaughan-Huxley E, Standfield N, John NW. Medical mentoring via the evolving world wide web. J Surg Educ 2013 Jan;70(1):121-128. [doi: 10.1016/j.jsurg.2012.06.024] [Medline: 23337681]

Abbreviations

HONcode: Health on the Net Foundation Code of Conduct
IPDAS: International Patient Decision Aid Standards
SUNDAE: Standards for Universal Reporting of Patient Decision Aid Evaluation Studies

Edited by G Eysenbach, R Kukafka; submitted 07.05.20; peer-reviewed by M Nomali, R Jamison, E Ross; comments to author 15.06.20; revised version received 06.08.20; accepted 02.11.20; published 23.09.21.

Please cite as:
Oktay LA, Abuelgasim E, Abdelwahed A, Houbby N, Lampridou S, Normahani P, Peters N, Jaffer U
Factors Affecting Engagement in Web-Based Health Care Patient Information: Narrative Review of the Literature
J Med Internet Res 2021;23(9):e19896
URL: https://www.jmir.org/2021/9/e19896
doi:10.2196/19896
PMID:34554104

Review

Effect of Digital Care Platforms on Quality of Care for Oncological Patients and Barriers and Facilitators for Their Implementation: Systematic Review

Jana S Hopstaken[1*], MD, MA; Lynn Verweij[2*], MSc; Cees J H M van Laarhoven[1], MD, MSc, PhD; Nicole M A Blijlevens[2], MD, PhD; Martijn W J Stommel[1], MD, PhD; Rosella P M G Hermens[3], MSc, PhD

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

[2]Department of Hematology, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands

[3]Department of IQ Healthcare, Radboud University Medical Center, Nijmegen, Netherlands

[*]these authors contributed equally

Corresponding Author:
Jana S Hopstaken, MD, MA
Department of Surgery
Radboud Institute for Health Sciences
Radboud University Medical Center
Geert Grooteplein 10
Nijmegen, 6525 GA
Netherlands
Phone: 31 024 36 680 86
Email: jana.hopstaken@radboudumc.nl

Abstract

Background: Oncological health care services are challenged by the increasing number of cancer survivors, long-term follow-up care, and fragmentation of care. Digital care platforms are potential tools to deliver affordable, patient-centered oncological care. Previous reviews evaluated only one feature of a digital care platform or did not evaluate the effect on enhancement of information, self-efficacy, continuity of care, or patient- and health care provider–reported experiences. Additionally, they have not focused on the barriers and facilitators for implementation of a digital care platform in oncological care.

Objective: The aim of this systematic review was to collect the best available evidence of the effect of a digital care platform on quality of care parameters such as enhancement of available information, self-efficacy, continuity of care, and patient- and health care provider–reported experiences. Additionally, barriers and facilitators for implementation of digital care platforms were analyzed.

Methods: The PubMed (Medline), Embase, CINAHL, and Cochrane Library databases were searched for the period from January 2000 to May 2020 for studies assessing the effect of a digital care platform on the predefined outcome parameters in oncological patients and studies describing barriers and facilitators for implementation. Synthesis of the results was performed qualitatively. Barriers and facilitators were categorized according to the framework of Grol and Wensing. The Mixed Methods Appraisal Tool was used for critical appraisal of the studies.

Results: Seventeen studies were included for final analysis, comprising 8 clinical studies on the effectiveness of the digital care platform and 13 studies describing barriers and facilitators. Usage of a digital care platform appeared to enhance the availability of information and self-efficacy. There were no data available on the effect of a digital care platform on the continuity of care. However, based on focus group interviews, digital care platforms could potentially improve continuity of care by optimizing the exchange of patient information across institutes. Patient-reported experiences such as satisfaction with the platform were considerably positive. Most barriers for implementation were identified at the professional level, such as the concern for increased workload and unattended release of medical information to patients. Most facilitators were found at the patient and innovation levels, such as improved patient-doctor communication and patient empowerment. There were few barriers and facilitators mentioned at the economic and political levels.

Conclusions: The use of digital care platforms is associated with better quality of care through enhancement of availability of information and increased self-efficacy for oncological patients. The numerous facilitators identified at the patient level illustrate that patients are positive toward a digital care platform. However, despite these favorable results, robust evidence concerning the

effectiveness of digital care platforms, especially from high-quality studies, is still lacking. Future studies should therefore aim to further investigate the effectiveness of digital care platforms, and the barriers and facilitators to their implementation at the economic and political levels.

(J Med Internet Res 2021;23(9):e28869) doi:10.2196/28869

KEYWORDS

digital care platforms; cancer care; eHealth; telemedicine; health care services; fragmentation of care; health care fragmentation; oncology; quality of care; barriers; facilitators; patient experience

Introduction

Background

Over the past few decades, the management of patients with cancer has considerably changed. Owing to earlier detection of cancer and improved treatment strategies, the number of cancer survivors has increased [1]. For this reason, patients with cancer currently require long-term follow-up care, similar to patients with chronic diseases. It is expected that the increased need of follow-up care, in combination with an overall increase of cancer patients due to an aging society, will intensify the use of health care services and increase health care costs [2]. Current health care systems cannot accommodate these increased demands and are deemed unsustainable [3,4]. An additional problem that requires a reevaluation of oncological health care services is the multidisciplinary, and sometimes multi-institutional, aspect of current care. This refers to the fact that cancer patients are usually cared for by multiple health care providers (HCPs), and that, as a result of centralization of complex care to high-volume centers [5], patients may receive parts of their treatment in multiple hospitals [6-9]. These aspects result in the fragmentation of cancer care [10].

The challenges facing cancer care are recognized by the World Health Organization [11] and the Health Program of the European Union [12], and have prompted them to think of different health care structures that enable the delivery of affordable, coordinated, patient-centered oncological care. A solution in restructuring health care for oncological patients could be found in telemedicine or electronic health (eHealth). Moreover, the COVID-19 pandemic, which led to the sudden forced implementation of telemedicine, has demonstrated the potential of telemedicine and eHealth to many patients and doctors [13]. These technologies therefore seem to be indispensable for sustainable care.

For these reasons, the time seems right to deploy eHealth much more widely in oncological care. An updated, accepted definition of eHealth is lacking, but eHealth can be perceived as an umbrella term for all digital communication and information technologies that aid in health care or health care services [14,15]. Over the past decades, multiple eHealth interventions have been developed to support oncological care [16,17]. The eHealth-based intervention of interest in this review is a digital care platform. We define a digital care platform as an eHealth-based tool that aims to increase coordinated and patient-centered care. A digital care platform incorporates several different features, which separately have been the subject of study elsewhere [18-21]. First, digital care platforms may provide patients with information specific to their situation. Second, they can provide an overview of the patients' personal health records, including appointments, medical results, and correspondence. Third, digital care platforms may also offer direct, secure messaging with HCPs (eg, electronic consultation [e-Consult]). We consider these three characteristics as key features of a digital care platform. Some additional features may include the registration of patient-reported outcomes (PROMs), a patient forum, the possibility to exchange patient-related information between different health care institutes, and enable communication between HCPs and general practitioners.

Studies on the use of eHealth-based tools and their effect on care processes have been previously described for patients with chronic diseases [22,23]. A recent systematic review by Tighe et al [23] indicated that digital platform–like interventions such as self-management tools have a positive effect on physical activity and disease-related quality of life. A systematic review by Kooij et al [22] indicated that digital interventions had positive effects on patient confidence and HCP satisfaction, but that firm conclusions on its clinical effects could not be drawn. Studies on the use of eHealth in oncological care have reported favorable results, as it has been associated with improved patient-provider communication [24]; improved coping with cancer-related symptoms such as fatigue, depression, anxiety, and physical activity [16,25-28]; and improved medication adherence and higher patient satisfaction [20,27]. Some studies have also reported favorable effects on quality of life [29,30]. However, these studies investigated the effect of only one feature of a digital care platform, for instance the registration of PROMs. Previous reviews did not study the digital care platform as a central part of the oncological health care service or the effect of digital care platforms on enhancement of information, continuity of care, or patient- and HCP-reported experiences. The effect on patient self-efficacy has only been studied for one feature of a digital care platform, namely self-management programs [25]. The outcome parameters of interest (ie, enhancing availability of information, self-efficacy, continuity of care, and patient-reported and HCP-reported experiences) can also be placed in the Quality of Care framework provided by the Institute of Medicine [31]. The six domains of quality of care are safe, effective, patient-centered, timely, efficient, and equitable care. Enhancing availability of information is a measure of effective and equitable care, self-efficacy is a measure of patient-centered care, patient-reported and HCP-reported experiences are measures of safe and patient-centered care, and continuity of care is a measure of efficient care and arguably also timely care.

Additionally, we did not identify any existing systematic reviews that focused on the barriers and facilitators that exist for the

successful implementation of a digital care platform in oncological care. This is important to assess so as to adequately determine the feasibility of a digital care platform as part of routine oncological care.

Objectives

The aim of this systematic review was two-fold: (1) to collect the best available evidence of the effect of a digital care platform on quality of care for oncological patients by focusing on enhancement of available information, self-efficacy, continuity of care (including communication), and patient-reported and HCP-reported experiences; and (2) to analyze the currently reported barriers and facilitators for implementation of a digital platform in oncological health care.

Methods

Protocol and Registration

This systematic review protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO; registration number CRD42020199282) [32] and was carried out according to the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analyses) guidelines [33].

Search Strategy

The PubMed (Medline), Embase, CINAHL, and Cochrane Library databases were searched for the period from January 1, 2000 until May 27, 2020. Search terms used included "neoplasms" and affiliated terms combined with "patient portal" or "digital care" or "eHealth." The aim of this search query was to provide search results on literature involving both the effect of a digital care platform on quality of care for oncological patients as well as barriers and facilitators for implementation. The exact search query is shown in Textbox 1.

Textbox 1. Search query.

(neoplasms [mesh] OR cancer* [tiab] OR tumor* [tiab] OR tumour* [tiab] OR neoplasm* [tiab] OR malignan* [tiab])

AND

("Patient Portals"[Mesh] OR "Telemedicine" [Mesh] OR Patient portal*[tiab] OR Patient platform*[tiab] OR Patient web portal*[tiab] OR Patient internet portal*[tiab] OR virtual care*[tiab] OR digital care [tiab] OR ehealth [tiab] OR e-health [tiab] OR econsult [tiab] OR e-consult [tiab])

Eligibility Criteria

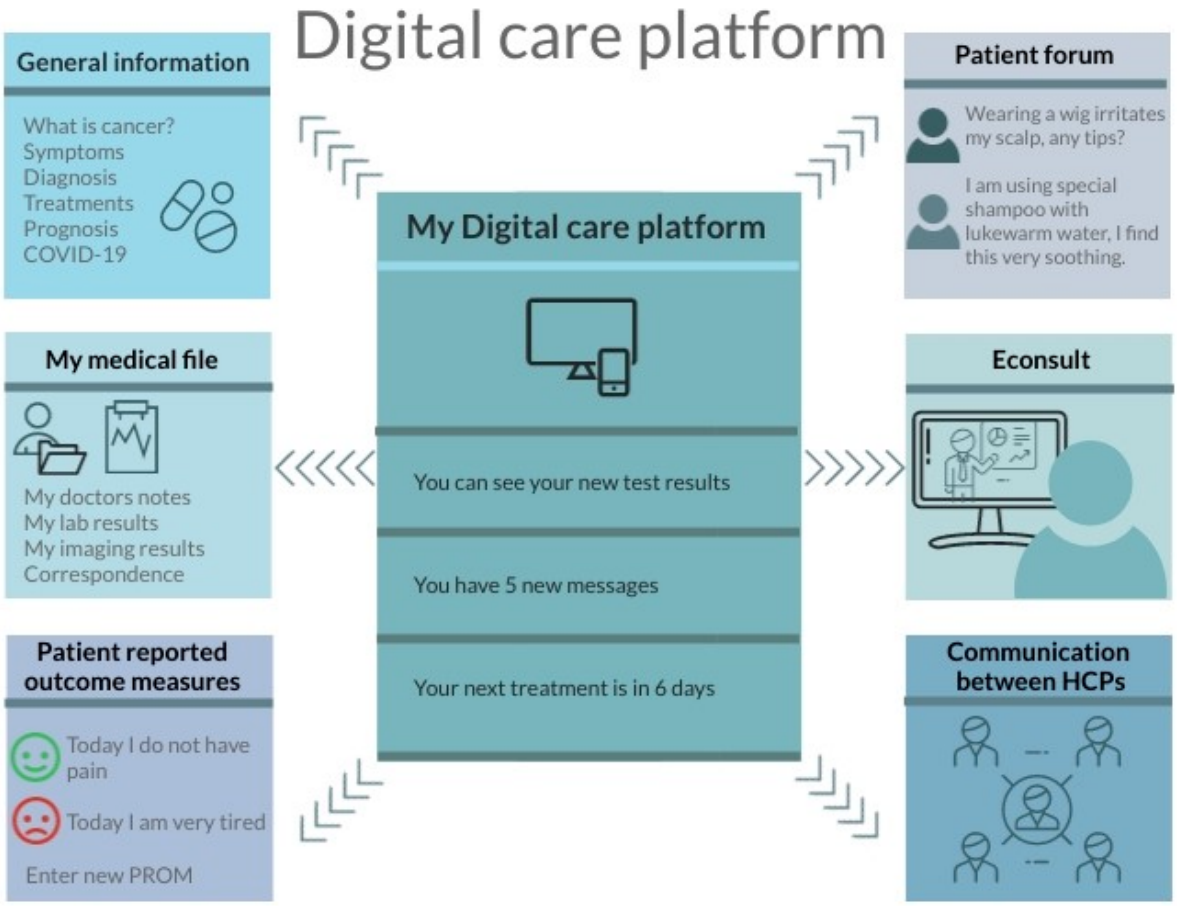
Studies of interest included randomized controlled trials (RCTs), prospective studies, and retrospective cohort studies. In addition, qualitative studies in which questionnaires were distributed or stakeholders were interviewed to investigate barriers and facilitators were also included. All studies were required to generate empirical data. Studies not written in English were also screened on the condition that they presented an English abstract. A translator could translate the full text if the abstract seemed to be eligible, thus avoiding a language bias. Systematic and narrative reviews, conference abstracts, and single case reports were excluded.

For our first objective, to collect the best available evidence of the effect of a digital care platform on quality of care, studies were required to involve oncological patients ≥18 years old. These clinical studies had to assess the effectiveness of a digital care platform. Although we provided our definition of a digital care platform in the Introduction, considering the exploratory stage of research of digital platforms in oncological care, we suspected that there would be a rather limited number of studies investigating such an extensive digital care platform. For this reason, for inclusion in our systematic review, a digital care platform was required to have at least two of the following key features: (1) provide general information concerning the disease as education (eg, symptoms, treatment, follow-up, prognosis); (2) provide patient-specific information concerning their medical file, such as planned appointments, treatments, and lab results; and (3) enable patients to communicate with their physician or specialized nurse via chatting, e-Consult, or email. Figure 1 depicts these and other features of a digital care platform.

Studies that did not involve a digital care platform with at least two of the three above-mentioned features and studies that provided insufficient or vague details concerning the digital intervention were excluded from the analysis to assess our primary objective.

For the second objective, identification of barriers and facilitators for implementation of a digital care platform in oncological care, studies were not required to involve a digital care platform that was already implemented. They could comprise studies that actually implemented a digital care platform and subsequently described the barriers and facilitators for implementation, but they could also comprise studies that identified barriers and facilitators based on a hypothetical discussion with stakeholders. In the latter case, the digital care platform in question was not yet implemented or developed.

Figure 1. Explanatory illustration of a digital care platform. Studies included in this systematic review were required to investigate a digital care platform with at least two of the following features: (1) general information provision, (2) electronic patient file, or (3) electronic consult (Econsult) with health care providers (HCPs). PROM: patient-reported outcome measure.



Study Selection

Two separate reviewers (JH and LV) screened the search output on titles and abstracts using Rayyan software [34]. During this screening process, the reviewers were blinded to each other's decisions. Studies with a contradictory judgment were discussed. In case of a remaining discrepancy, senior reviewers (RH and MS) were asked to arbitrate. The full texts of the selected studies were screened for eligibility. Reference lists of studies that were included after full-text screening were checked for additional eligible studies (snowball method). In cases in which multiple eligible studies reported on the same dataset (≥50% overlap of sample size), we excluded the article with the shortest follow-up.

Data Extraction and Analysis

Two authors (JH and LV) extracted data using a shared template. These data included: (1) author, year of publication; (2) country; (3) study design; (4) patient population, sample size; (5) platform that was the subject of the study; 6) features of the platform; and (7) the outcome parameters, including enhancement of available information, self-efficacy, continuity of care, and patient- and HCP-reported experiences. If possible, data were pooled; otherwise, synthesis of the results was performed qualitatively.

Barriers and facilitators for implementation of digital care platforms were categorized according to the framework of Grol and Wensing [35]. This framework categorizes barriers and facilitators at six different levels: (1) innovation, which involves advantages in practice, feasibility, credibility, accessibility, and attractiveness; (2) individual professional, which concerns the awareness, knowledge, attitude, motivation to change, and behavioral routines of the involved professionals; (3) patient, which involves knowledge, skills, attitude, and compliance of the patients; (4) social context, which concerns opinion of colleagues, culture of the network, collaboration, and leadership; (5) organizational context, which includes the organization of the care processes, staff, capacities, resources, and structures; and (6) economic and political contexts, which involve regulations, financial arrangements, and policies. The frequency of the barriers and facilitators mentioned, and the quality of the studies were used to prioritize barriers and facilitators.

Critical Appraisal of Evidence

The Mixed Methods Appraisal Tool (MMAT) version 2018 was used to critically appraise the included studies [36]. The MMAT is designed to help reviewers appraise the quality of empirical studies with different methods. These include quantitative RCTs, quantitative nonrandomized trials, quantitative descriptive studies, as well as qualitative and mixed methods studies. Each category includes five different quality parameters, all requiring to be assessed by answering "yes," "no," or "cannot tell" (maximum total score=5). Two authors (JH and LV) independently appraised all studies. In case of disagreement, RH and MS were asked to arbitrate. Studies with

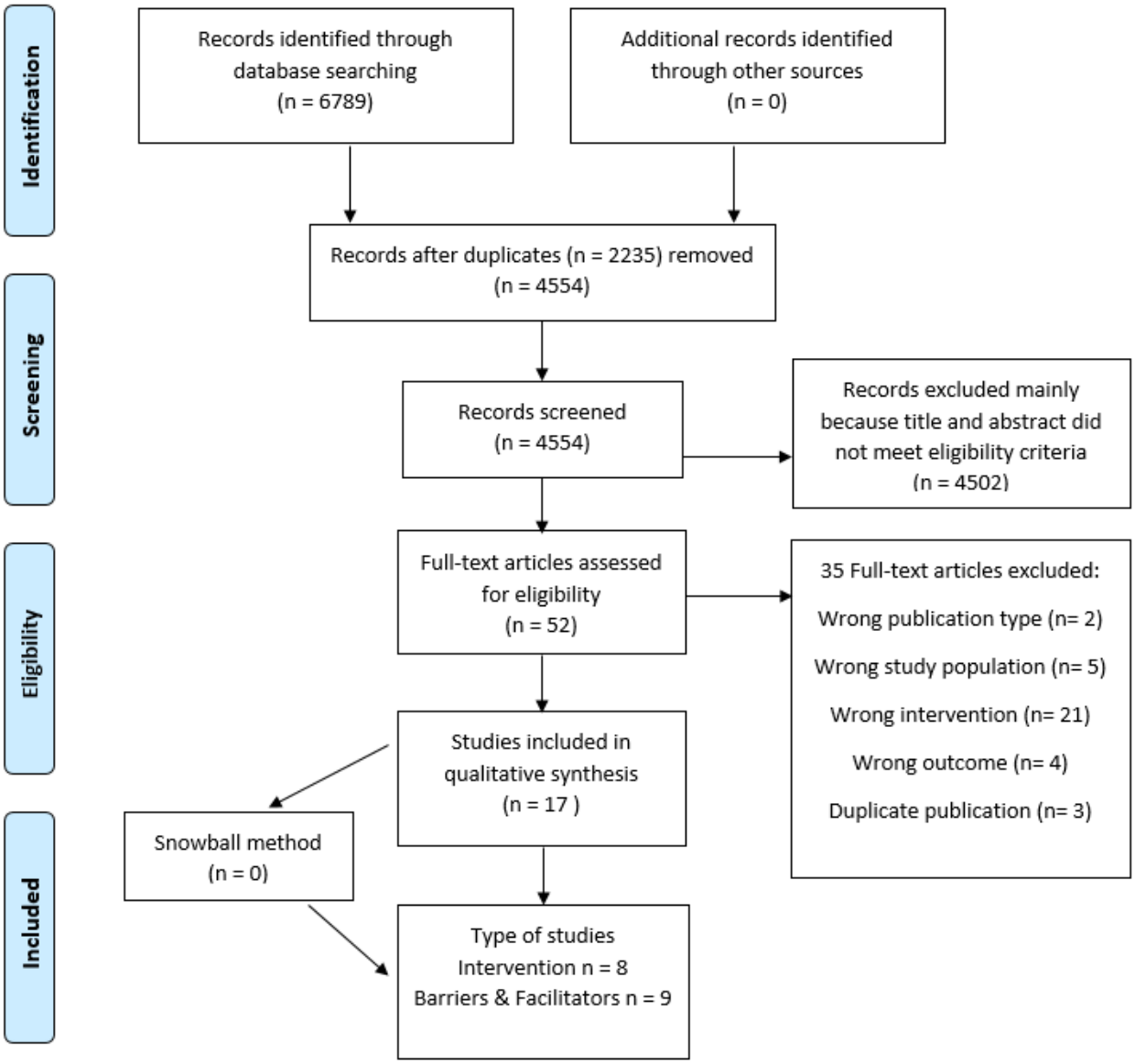
a total score of 1-2 were considered to be of low quality, a study with a score of 3 was considered to be of moderate quality, and studies with total scores of 4-5 were considered high-quality studies.

Results

Search Results

The initial search generated 6789 articles. After removal of duplicates and initial screening of titles and abstracts, 52 studies remained for full-text assessment. Of these 52 studies, 17 articles met our eligibility criteria and were included in this systematic review. Figure 2 depicts the PRISMA flowchart for the study screening process.

Figure 2. PRISMA flowchart.



Study Characteristics

Overview

Of the 17 included studies, eight were clinical studies that implemented a digital care platform and assessed its effect on at least one of the predefined outcome parameters [37-44]. Four of these clinical studies [39,42-44] also described barriers and facilitators for implementation. The other nine studies [45-53] only investigated barriers and facilitators for digital care platforms. None of the included studies was written in a non-English language. Table 1 and Table 2 show the baseline characteristics of the clinical studies and the barrier and facilitator studies included in this review, respectively.

Table 1. Baseline characteristics of clinical studies.

Reference	Country	Study design	Patient population and sample size	Platform studied	Features of digital care platform			Outcome parameters				Q1[a]	Q2[b]
					General information	Access to EMR[c]	e-Consult[d]	Availability of information	Self-efficacy	COC[e]	PRE[f]/ HCPE[g]		
Baker et al [37]	US	RCT[h]	450 patients with BCA[i]	CHESS[j]	✔	X	✔	✔	✔	X	✔	✔	X
Børøsund et al [38]	Norway	RCT	167 patients with BCA (<12 months)	Web-choice	✔	X	✔	X	✔	X	X	✔	X
De Regge et al [39]	Belgium	Mixed methods	23 patients with metastatic kidney or bone cancer	DOP	✔	✔	✔	✔	X	X	✔	✔	✔
Groen et al [40]	Netherlands	Feasibility study	37 patients with NSCLC[k] (<12 months)	MyAVL	✔	✔	X	✔	✔	X	✔	✔	X
Gustafson et al [41]	US	RCT	257 patients with BCA	CHESS	✔	X	✔	X	✔	X	X	✔	X
Kuijpers et al [42]	Netherlands	Mixed methods	92 patients with BCA (<12 months treatment)	MyAVL	✔	✔	X	✔	✔	X	✔	✔	✔
Ruland et al [43]	Norway	Prospective cohort	103 patients with BCA (n=56) and PCA[l] (n=47)	Web-choice	✔	X	✔	✔	X	✔	✔	✔	✔
Tiong et al [44]	Australia	Prospective pilot study	50 patients with BCA	Healthy.me	✔	X	✔	✔	✔	✔	✔	✔	✔

[a]Q1: addresses research question 1 (to collect the best available evidence of the effect of a digital care platform on quality of care for oncological patients).

[b]Q2: addresses research question 2 (to analyze the currently reported barriers and facilitators for implementation of a digital platform in oncological health care).

[c]EMR: electronic medical record.

[d]e-Consult: electronic consult.

[e]COC: continuity of care.

[f]PRE: patient-reported experiences.

[g]HCPE: health care provider–reported experiences.

[h]RCT: randomized controlled trial.

[i]BCA: breast cancer.

[j]CHESS: Comprehensive Health Enhancement and Support System.

[k]NSCLC: nonsmall cell lung cancer.

[l]PCA: prostate cancer.

Table 2. Baseline characteristics of barrier and facilitator studies.

Reference	Country	Study design	Patient population and sample size	Description of platform	Q1[a]	Q2[b]
Alpert et al [45]	US	In-depth semistructured interviews with cancer patients and oncologists	35 cancer patients (breast n=9, hematologic n=6, gastrointestinal n=5, genitourinary n=4, lung n=3, sarcoma n=3, skin n=3, gynecologic n=2, other n=3) and 13 oncologists	Patient portals offer functional benefits to patients as they enable convenient patient access to EMR[c] data from devices such as personal computers or smartphones; they allow the ability to request medication refills, schedule appointments, and they support secure messaging	X	✔
Baudendistel et al [46]	Germany	Qualitative explorative study using focus groups	12 colorectal cancer patients, 17 physicians, and 26 other health care professionals	Patient-controlled "personal electronic health record" (PEPA) (in development)	X	✔
Ector et al [47]	Netherlands	Design thinking development study	8 CML[d] patients, 14 hematologists	CMyLife, a web-based, patient-centered intervention	X	✔
Geerts et al [48]	Netherlands	Mixed methods design: focus group and literature for patient survey, and physician survey based on literature	Patients with hematologic malignancy (questionnaire: n=204, focus group: n=6) and their physicians (questionnaire: n=13)	Not yet developed	X	✔
Gerber et al [49]	US	Qualitative study using 2 focus groups	13 nurses from a cancer center	Not yet developed	X	✔
Girault et al [50]	Canada	Questionnaire-based survey	1072 patients treated in a comprehensive cancer center	Internet-based technologies such as patient portals, websites, and applications managed by health care institutions to improve cancer care coordination	X	✔
Kildea et al [51]	Canada	Participatory stakeholder design	Focus group and survey among 361 cancer patients	Electronic patient portal accessible through smartphones (not yet developed)	X	✔
Kuijpers et al [52]	Netherlands	Focus group interviewing (9 groups)	21 BCA[e] patients, 31 lung cancer patients, and 31 health professionals	Interactive portal	X	✔
McCleary et al [53]	US	Focus group sessions and surveys	Focus group: 20 patients and family and 5 advisory council members; survey: 1019 cancer patients	Patient gateway application as patient portal: a secure, web-based database enabling patient access to their health and disease information embedded within an EMR, managed by individual health care organizations and accessible via the internet	X	✔
De Regge et al [39]	Belgium	Mixed method triangulation design	Interviews with 23 patients, 2 physicians, 1 nurse specialist, 2 nurse consultants, 9 nurses, and 1 psychologist	The digital oncology platform includes the individualized care path, reliable treatment-related information, contact details for the treatment team, secure conversations with the treatment team, self-registration of complaints, and a diary	✔	✔
Kuijpers et al [42]	Netherlands	Mixed methods design: pretest-posttest design and focus group	6/92 BCA patients were included in focus group discussions; 24 health care providers were asked to fill out a questionnaire	MyAvL includes personalized educational material, overview of past and upcoming appointments, EMRs, medication overview, questionnaire concerning PROMs[f]	✔	✔
Ruland et al [43]	Norway	Prospective cohort	103 patients with breast cancer (n=56) and prostate cancer (n=47) received questionnaires with space for additional comments and suggestions	WebChoice is an interactive electronic health application that includes personalized information, a communication tool where patients can receive support from peers or professionals, a diary, and a self-management component	✔	✔

Reference	Country	Study design	Patient population and sample size	Description of platform	Q1[a]	Q2[b]
Tiong et al [44]	Australia	Prospective pilot study	50 patients with BCA were asked to use the platform and fill out a questionnaire; 9 patients were invited for a face-to-face feedback session	Healthy.me is a secure personally controlled health management website that features the patient journey with tailored information, appointment overviews, interactive forums, and messaging with peers and private messaging with health care providers	✔	✔

[a]Q1: addresses research question 1 (to collect the best available evidence of the effect of a digital care platform on quality of care for oncological patients).

[b]Q2: addresses research question 2 (to analyze the currently reported barriers and facilitators for implementation of a digital platform in oncological health care).

[c]EMR: electronic medical record.

[d]CML: chronic myeloid leukemia.

[e]BCA: breast cancer.

[f]PROM: patient-reported outcome measure.

Study Design

Among the clinical studies, three were RCTs [37,38,41], two were mixed methods studies [39,42], and three were pilot or feasibility studies [40,43,44] investigating the use of the platform in a small cohort. The three RCTs included 450, 167, and 257 patients, respectively, and all assessed the effectiveness of different variations of a digital care platform [37,38,41]. For example, Baker et al [37] assessed the effectiveness of different components of a digital care platform by comparing patient groups using a platform with information only; a platform with information and support; and a platform with information, support, and coaching. The control group did not use a digital care platform but used the internet. An example of one of the mixed methods studies is that performed by Kuijpers et al [42], in which a digital care platform was implemented in a small cohort with a postintervention questionnaire, followed by focus group discussions. Studies solely investigating barriers and facilitators were mainly qualitative in nature, with the exception of three studies applying a mixed methods methodology [48,51,53] and one study using a quantitative methodology [50].

Patient Population

Breast cancer patients formed the majority of the study population of the clinical studies, with 75% of all studies including solely this patient population. The patient population was therefore mostly female, highly educated, and young (mean age 50 years) [37,38,42,43]. The other studies included patients with nonsmall cell lung cancer or a diverse group a of cancer patients such as those with renal cell cancer and sarcoma. One study [43] included two different patient groups, namely breast cancer and prostate cancer patients, and compared the use and effectiveness of the digital care platform between these two groups. The barrier and facilitators studies (Table 2) included patients with colorectal cancer, lung cancer, breast cancer, hematological cancer, and a variety of other cancer types.

Interventions

Within the eight clinical studies, five web-based platforms were distinguished. These platforms were all web-based and were not integrated in the electronic medical record (EMR). Seventy-five percent of the studies assessed a platform that also intended to improve self-management by, for instance, symptom monitoring, physical activity advice, or self-therapy [37,38,40-43]. Although each digital care platform had at least two of the three predefined criteria, heterogeneity concerning the platforms was observed. For example, the digital care platform described by De Regge et al [39] enabled HCPs, including general practitioners, to read their patients' medical records from other health care centers. This allowed them to gain insight into their patients' treatment trajectory and decision-making by other involved HCPs. Other platforms such as those described by Groen et al [40] and Kuijpers et al [42] did not have this feature, and provided personalized patient education material, an overview of upcoming hospital appointments, and tailored physical activity.

Methodological Quality

Quality assessment of the included studies is shown in Table 3. Of the clinical studies, four studies were of low quality [37,41,43,44], three of moderate quality [38,39,42], and one study was assessed as a high-quality study [40]. Regarding the barrier and facilitator studies, three studies were of low quality [49,51,53], one of moderate quality [48], and five were of high quality [45-47,50,52].

Table 3. Mixed Methods Appraisal Tool scoring of the included studies (N=17).

Reference	Qualitative					Quantitative RCTs[a]					Quantitative nonrandomized					Quantitative descriptive					Mixed methods[b]					Total score[c]
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	
Clinical studies																										
Baker et al [37]	—[d]	—	—	—	—	N[e]	?[f]	?	?	?	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
Børøsund et al [38]	—	—	—	—	—	?	Y[g]	Y	?	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
Groen et al [40]	—	—	—	—	—	—	—	—	—	—	Y	Y	Y	?	Y	—	—	—	—	—	—	—	—	—	—	4
Gustafson et al [41]	—	—	—	—	—	1	?	?	?	?	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
Clinical studies with barriers and facilitators																										
de Regge et al [39]	Y	Y	?	N	Y	—	—	—	—	—	—	—	—	—	—	Y	Y	Y	Y	Y	Y	?	Y	Y	N	3
Kuijpers et al [42]	—	—	—	—	—	—	—	—	—	—	N	Y	N	Y	Y	—	—	—	—	—	—	—	—	—	—	3
Ruland et al [43]	—	—	—	—	—	—	—	—	—	—	Y	N	?	?	?	—	—	—	—	—	—	—	—	—	—	1
Tiong et al [44]	—	—	—	—	—	—	—	—	—	—	N	?	Y	N	?	—	—	—	—	—	—	—	—	—	—	1
Barrier and facilitator studies																										
Alpert et al [45]	Y	Y	Y	Y	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	5
Baudendistel et al [46]	Y	Y	Y	Y	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	5
Ector et al [47]	Y	Y	Y	Y	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	5
Geerts et al [48]	Y	Y	?	N	?	—	—	—	—	—	—	—	—	—	—	Y	Y	Y	?	Y	Y	Y	Y	?	N	3
Gerber et al [49]	Y	Y	?	?	N	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2
Girault et al [50]	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Y	Y	?	Y	Y	—	—	—	—	—	4
Kildea et al [51]	Y	Y	?	N	?	—	—	—	—	—	—	—	—	—	—	Y	Y	?	?	?	Y	?	?	?	N	1
Kuijpers et al [52]	Y	Y	?	Y	Y	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	4
McCleary et al [53]	Y	Y	?	N	?	—	—	—	—	—	—	—	—	—	—	Y	Y	?	Y	Y	?	N	?	Y	N	1

[a]RCT: randomized controlled trial.

[b]For mixed methods studies, a score of 5.5 was required to be evaluated using category 1 and category 3 or 4.

[c]Total score: 0-2=low quality, 3=intermediate quality, 4-5=high quality.

[d]Not relevant.

[e]N: no.

[f]?: cannot tell.

[g]Y: yes.

Outcome Measures

Enhancing Availability of Information

Five studies [39,40,42-44] reported on enhancing availability of general information on a digital care platform. In a high-quality feasibility study [40], the option to receive information was scored by patients with a mean of 7.1 (SD 1.5) on a scale of 1-10. The actual usage of information sections in the digital care platform differed among studies. One high-quality study reported the patient EMR to be the most used section of the digital care platform (mean 6.7, SD 4.7 logins during the 4-month study period) [40] and another study of moderate quality reported the opposite finding that the information section was rarely viewed (mean 0.75, SD 1.4 number of times participants consulted the information section) [39]. Needs of patients also differed, as Ruland et al [43] reported a significant difference (625 vs 271, *P*=.01) in the number of visits to the information section between breast cancer and prostate cancer patients, with breast cancer patients visiting this section more often. In this study, which was appraised as low quality, information comprehensibility of the digital care platform was rated with an overall mean score of 7.2 (SD 1.4) and usefulness was rated with a mean score of 6.5 (SD 1.7), both on a scale of 1-10 [43]. The main reasons for consulting the general disease-related information section of the digital care platform were to get help with problems and to get assurance. In addition to the information section, the platform's ability to directly communicate with a nurse was perceived as very useful, because it was perceived as easy-to-understand information. The nurses' contribution to a patient forum was reported as a trustworthy source of information [43] and therefore was highly valued by patients [44].

Self-Efficacy

Six studies investigated the effect of a digital care platform on self-efficacy [37,38,40-42,44] and all of them reported positive effects. Although the effects were not significant, a clear trend toward a positive effect on self-efficacy was reported. In the RCT performed by Børøsund et al [38], a moderate-quality study, patients using a digital care platform tended to score higher on self-efficacy compared with patients in the usual care group (mean difference 8.81, range 33-297, 95% CI –0.92 to 18.53, *P*=.08). Patients reported that a digital care platform was helpful in managing their health [44]. Groen et al [40] and Kuijpers et al [42], studies with high to intermediate quality, reported that access to general disease-related information and patients' EMRs enhanced patients' knowledge of their disease and their sense of control, based on questionnaires. Two RCTs showed that using a digital care platform increased patients' perceived ability to obtain and use health care information. The digital care platform also increased their comfort and activation level of dealing with physicians and health care situations compared to the patient group only using the internet [37,41]. However, both of these RCTs were categorized as low quality.

Continuity of Care

None of the included clinical studies reported specifically on continuity of care as a primary or secondary outcome parameter. However, three low/intermediate-quality studies [38,39,43] reported on the aspect of communication between patients and HCPs and among HCPs within the platform. A questionnaire-based pilot study [43] reported that the most valued feature of the platform (ie, Webchoice) as expressed by the study population was the ability to send messages to their HCPs. That same research group investigated the effects of their platform in an RCT and reported 40% use of the messaging service in this platform [38]. Patients perceived this feature to be useful and easy to understand. De Regge et al [39] reported that HCPs perceived the ability to exchange research results between HCPs (eg, HCPs in other hospitals and primary care) as valuable to patient care and a means to optimize continuity of care across institutes.

Patient- and HCP-Reported Experiences

Seven studies [37-40,42-44] reported on patient- or HCP-reported experiences. Patient satisfaction with the studied platform was considerably high in three intermediate/high-quality studies, with a mean rating of 3.9 (range 3.8-4.09) on a 1-5 scale [39,40,42]. A low-quality RCT that aimed to compare the effects of three different types of interventions and one control observed significantly higher patient satisfaction scores with their HCPs compared with those of the control group at 6 weeks (3.46 vs 3.17, *P*=.01) and 6 months (3.48 vs 3.28, *P*=.03) [37]. The majority of patients (75%-93%) stated that the platform was easy to use and that it was a valuable addition to their health care experience [40,42]. The most used features of the platform varied among studies, but mainly consisted of the personal medical records in intermediate/high-quality studies [39,40,42]. Only two clinical studies reported the experiences by HCPs. De Regge et al [39], an intermediate-quality study, reported that HCPs perceived the digital care platform as valuable compared to current care because it provides reliable, easy-to-access information for patients and because it enables the exchange of patient-related information between care providers. Despite these positive experiences, one-quarter of the physicians interviewed by Kuijpers et al [42], an intermediate-quality study, reported an increase in workload after installment of the digital care platform, varying from a few extra minutes to more than 10 minutes per patient, for additional explanations on information made available on the platform.

Barriers and Facilitators Levels

Overview

Table 2 presents the baseline characteristics of the studies describing barriers and facilitators. Tables S1 and S2 in Multimedia Appendix 1 present all barriers and facilitators identified in the included studies. The most prominent barriers and facilitators are discussed below. Figure 3 depicts the illustrative quotations found in the studies for each level.

Figure 3. Illustrative quotations mentioned in included studies concerning barriers (in red) and facilitators (in green) for implementation of a digital platform.

Innovation Level

Four studies [44,46,48,51] described technical challenges as a possible barrier for implementation. These included software development, functionality of the website, bugs, and log-on and internet access issues. A second barrier, concerning accessibility of the innovation, was patients' unawareness of the existence of the digital care platform, which led to limited or no use of the platform [53]. Despite these barriers, three studies showed that the use of a digital care platform is feasible [40,42,44]. Facilitators in practice were numerous: access to general and tailored medical information [42,45,47-49,52], full EMRs [39,42,45-48,50-52], and educational materials [42,47]. Moreover, the possibility to consult HCPs through interactive tools was mentioned to be a positive feature of a digital care platform that encouraged usage [43,44,47,48,50].

Individual Professional Level

The most frequently described barrier at the professional level was the concern of HCPs regarding the release of medical information on the digital care platform and the comprehensibility of medical jargon for patients [39,45,46,48,49,51,52]. Patients learning about a change in their health and well-being via a digital care platform was a fear expressed by all oncologists included in two studies [45,48]. The second most frequently described barrier, described in seven studies [39,42,45-47,49,51], concerned the fear of increased workload such as an increased burden of documentation, monitoring, and direct digital communication with patients [39,42,45,49].

Concerning facilitators, one study suggested that a digital care platform would not increase workload but would rather reduce workload by easing data management [46]. Another facilitator

mentioned was the ability to exchange information or medical results between different HCPs, including those in primary care such as general practitioners [39,46]. The optimization of data exchange between institutes could avoid or reduce repeated health assessments [46].

Patient Level

Three studies indicated that patients had concerns about the readability of the information (medical jargon) displayed on the digital care platform and whether medical results could be adequately understood without professional interpretation [45,46,48]. Some patients mentioned that this could increase anxiety [48,52].

Facilitators at the patient level were numerous, as indicated in Table S2 of Multimedia Appendix 1. In several studies, patients and oncologists mentioned that patients having access to the platform and their EMR could better prepare them for their doctor visit and enabled them to take on an active role during the consultation [45-47]. In addition, patients' willingness to communicate via digital resources facilitated implementation [53]. Lastly, improved patient-doctor communication was described as a facilitator, either by improved preparedness for doctor visits or by direct messaging [52].

Social Context Level

Three studies reported concerns on how a digital care platform could change the doctor-patient relationship [48] such as by increasing the patient's autonomous handling of information [46]. The digital care platform could make patients increasingly reliant on technology with decreased reliance on face-to-face communication [46,49]. This was feared as it may negatively influence the outpatient clinic visits as patients become more focused on discussing (irrelevant) details regarding available biomedical results rather than discussing values and preferences important for a follow-up treatment [48].

The change of the doctor-patient relationship was also mentioned as a facilitator. By enabling access to medical information, patients could become more actively involved in the management of their care and feel more comfortable to interact with their HCPs during consultations [42,45].

Organizational Context Level

Two studies mentioned the integration of the platform into existing systems as a barrier [44,47]. This would be especially relevant in cases in which the digital care platforms were also developed for other chronic diseases [47]. Limitations in time and resources were reported as another type of barrier [44].

A clear facilitator at the organizational level was the digital care platform's ability to exchange important patient information across health care institutes. HCPs emphasized the advantage of a digital care platform in improving the cross-sectoral availability of information about the patient to HCPs who are involved in the patient's treatment. It is suggested that the digital care platform could thereby diminish information loss, be of use in emergency situations where the rapid release of medical information is vital, and avoid repeated diagnostic investigations [46]. Organization of care could be improved as the improved availability of information promotes cross-boundary continuity of care and diminishes fragmentation of care [39,46]. Another facilitator at the organizational level was the early introduction of the digital care platform to the patient, preferably at diagnosis [44], and sufficient instructions concerning appropriate use of the platform for patients as well as HCPs [49].

Economic and Political Contexts Level

Data security and protection regarding the digital care platform were concerns reported by numerous studies [46,50,51,53]. During the development of the digital care platform described by Kildea et al [51], cybersecurity testing and legal issues were the most time-consuming processes. These legal issues involved the ownership of intellectual property [51] and liability [46,51].

Secure access was therefore identified as an important facilitator for a digital care platform [48]. Another facilitator, in the economical context, was that conferred by the optimized exchange of patient information between institutes, enabled with a digital care platform, and that unnecessary repeated diagnostic procedures and health assessments could be avoided [46].

Discussion

Principal Findings

A digital care platform for oncological patients has the potential to improve quality of care through the improved availability of information and positive effect on self-efficacy. Although continuity of care was not studied as a primary outcome in the identified studies, based on focus group interviews with HCPs, a digital care platform potentially improves continuity of care by optimizing the exchange of patient information across institutes. Patient-reported experiences such as satisfaction with the platforms were considerably positive. Our barrier and facilitator analysis indicated that the majority of barriers exist at the professional level. This included a concern of increased workload for HCPs and release of unattended medical information to patients. The majority of facilitators were identified at the patient and innovation levels. The patient's ability to become more informed, empowered, and involved in their care was identified as a prominent facilitator. Another relevant facilitator, at the organization level, is the digital care platform's potential to improve information exchange between HCPs across different institutes. This is relevant as this may improve continuity of care and diminish fragmentation of oncological care. Among the clinical studies, the majority were of low to intermediate quality. Regarding the barrier and facilitator studies, the majority were of high quality.

The positive effects of a digital care platform described in this review are consistent with prior research studying isolated features of digital care platforms [16,29,54]. A digital care platform seems to be a helpful medium in providing patients with general disease-related and personal information. This is a welcoming result as previous studies have described that a considerable number of patients are dissatisfied with information provision [55,56]. Moreover, 40%-80% of all medical information provided by the HCP during a consultation is forgotten or remembered incorrectly [57]. Therefore, the need of patients to read or reread medical information that is relevant

for their situation is an important facilitator for a digital care platform. In this systematic review, we identified studies that showed positive effects on self-efficacy, but no convincing clinically significant effects. In another systematic review, which studied the effectiveness of eHealth-based self-management tools, significant yet small effects on self-efficacy were described [25]. The authors explain that the tool was able to increase self-efficacy by enabling patients to enhance participation in their care trajectory. A more recent randomized trial refuted this finding [58]. In this RCT, the effect of an eHealth-based self-management tool (Oncokompas) was investigated in 600 cancer patients. At 3 and 6 months, patients were assessed on self-efficacy with the General Self-Efficacy scale. No significant effects on self-efficacy were observed (mean difference 0.5, 95% CI –0.4 to 1.4, *P*=.31). The authors partially attributed the lack of effect to the included patient population, which mostly consisted of patients diagnosed with cancer 2 years prior to the study. The authors suggested that this population is perhaps less in need of a self-management tool to increase self-efficacy as they have more experience and "know-how" compared with their newly diagnosed peers. The hypothesis that the effect of an eHealth-based tool may vary considerably depending on the patient population is important to consider for assessing the effectiveness of a digital care platform. In our systematic review, we included studies with patients in different stages of their disease trajectory and with different cancer types, and all effects regarding self-efficacy were positive but not significant. On the one hand, this aspect can be perceived as a limitation because the heterogeneity of patients may underestimate the effect of a digital care platform. On the other hand, it is a strength as we now have an overview of the effect of a digital care platform on self-efficacy of a broad cancer patient population, thereby increasing the generalizability of our results.

Similarity with prior work concerns the limited number of studies investigating combined features of a digital care platform. During the screening process of this review, we encountered numerous studies that only investigated one feature of the platform. This finding is similar to the findings of a review by Kruse et al [20], which concluded that many studies assessed the effect of one feature of a digital care platform and that a full platform was rarely studied in clinical practice. To adequately ascribe outcomes as an effect of the intervention, it is reasonable to first study an isolated feature of a digital care platform. However, it is absolutely vital to also study the combined features in a digital care platform for two key reasons: (1) because this is likely to be the eHealth intervention that is implemented in practice, and (2) because the effect of multiple features may not be equivalent to the sum of effects of a single feature.

The fact that we assessed the effects of digital care platforms that had at least two key features can therefore be considered a strength of this review. In addition, we assessed the effect of a digital care platform on quality of care parameters such as continuity of care, and described the patient- and HCP-related experiences. Other reviews did not, and instead only assessed the effect of a digital care platform with one feature or chose to study different endpoints such as fatigue, physical activity, depression, quality of life, and self-management abilities [20,25,29,30,54,59]. An additional strength is that we analyzed barriers and facilitators for implementation. This allows for a good understanding of what is needed for successful implementation. A final strength is that the search query was quite elaborate without exclusion of non-English articles, thereby minimizing the possibility of missing valuable studies. For these reasons, this systematic review provides a comprehensive overview of the best available evidence of the effect of a digital care platform on quality of care for oncological patients.

However, this systematic review also has some limitations. First, the included clinical studies were mostly early-stage, single-arm prospective studies where feasibility and acceptability were investigated. Although we did include three studies with a more advanced research methodology (ie, RCTs), these studies were all appraised as studies with low to intermediate quality. This makes interpretation of these RCTs quite challenging. One RCT reported a significant difference in patient satisfaction scores favoring a type of digital care platform [37]; however, the minimal difference (3.48 vs 3.28 on a 1-5 scale) puts into question the clinical relevance of this finding. A second limitation of this review concerns the heterogeneity of study designs and of the digital care platforms. Despite our predefined criteria of what a digital care platform should include, they were still quite different from each other with respect to the specific web format, options, and intended use, among other aspects. This heterogeneity precluded a pooled analysis with quantitatively measured outcome parameters. A third limitation concerns the study population and its representativeness. Three-quarters of the clinical studies included breast cancer patients. This study population mainly consists of well-educated females [60]. Indeed, 72%-73% of the breast cancer population in the studies included in this review had a college or university degree. It is plausible that a high education is an important factor for successful use of a digital care platform. For this reason, results of this review should be interpreted with caution and cannot simply be generalized to other cancer patients. A final limitation concerns the barrier and facilitator analysis. Studies included in this review almost exclusively focused on the barriers and facilitators mentioned by the end users of the platform (ie, the patients and HCPs). Evidently, their perspectives are crucial, but barriers and facilitators proposed by other stakeholders should also be investigated, such as stakeholders with an organizational, economic, and political background (eg, health insurers). This latter group can provide insight into what is required to realize the structural financing of a digital care platform.

Although there is a growing body of literature that describes the positive effects of digital care platforms, high-quality studies describing the effectiveness of these platforms integrated in oncological care are currently lacking. In addition, most studies implemented a digital care platform in one target patient population such as breast cancer patients. More evidence is required concerning the desirability and use of a platform in patients with other types of cancers and education levels. More specifically, it is important to investigate whether patients with a lower level of education can benefit from digital care platforms

to the same extent as their well-educated peers to avoid widened health disparities. The same applies to the patient population with limited internet access or internet skills. Studies in this review were performed in Australia, North America, and western Europe. The usefulness of digital care platforms in countries with low internet access is likely very different.

Concerning barriers and facilitators, future studies should aim to further elucidate barriers and facilitators at the organizational level and the economic and political levels. Ultimately, this is required for a digital care platform to become a successful eHealth-based tool in the improvement of quality of care for patients living with cancer.

Conclusion

Digital care platforms have a favorable effect on availability of information and enhancement of self-efficacy. Additionally, they could potentially serve as a valuable medium to improve continuity of care by optimizing communication between patients and HCPs and among HCPs. The vast majority of patients are positive about a digital care platform and its ability to meet their needs in improving the availability of information and patient involvement. Although these results are favorable, they were mostly generated by early-stage, nonrandomized studies with a specific patient population. To fully understand whether a digital care platform is able to increase quality of care by supporting the delivery of coordinated, patient-centered oncological care, more advanced studies such as RCTs are required, as well as studies investigating the barriers and facilitators at the economic and political levels.

Authors' Contributions

JH and LV were primarily responsible for the review and data extraction, and share first authorship. MS and RH contributed equally as senior reviewers and share last authorship. All authors contributed to and approved the final version of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Tables S1 and S2.
[DOCX File , 22 KB - jmir_v23i9e28869_app1.docx]

References

1. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, et al. Cancer treatment and survivorship statistics, 2019. CA Cancer J Clin 2019 Sep 11;69(5):363-385. [doi: 10.3322/caac.21565] [Medline: 31184787]
2. Yabroff KR, Lund J, Kepka D, Mariotto A. Economic burden of cancer in the United States: estimates, projections, and future research. Cancer Epidemiol Biomarkers Prev 2011 Oct 05;20(10):2006-2014. [doi: 10.1158/1055-9965.epi-11-0650]
3. Mayer DK, Nasso SF, Earp JA. Defining cancer survivors, their needs, and perspectives on survivorship health care in the USA. Lancet Oncol 2017 Jan;18(1):e11-e18. [doi: 10.1016/s1470-2045(16)30573-3]
4. Mariotto AB, Yabroff KR, Shao Y, Feuer EJ, Brown ML. Projections of the cost of cancer care in the United States: 2010-2020. J Natl Cancer Inst 2011 Jan 19;103(2):117-128 [FREE Full text] [doi: 10.1093/jnci/djq495] [Medline: 21228314]
5. Stitzenberg KB, Sigurdson ER, Egleston BL, Starkey RB, Meropol NJ. Centralization of cancer surgery: implications for patient access to optimal care. J Clin Oncol 2009 Oct 01;27(28):4671-4678. [doi: 10.1200/jco.2008.20.1715]
6. Reade C, Elit L. Trends in gynecologic cancer care in North America. Obstet Gynecol Clin North Am 2012 Jun;39(2):107-129. [doi: 10.1016/j.ogc.2012.02.003] [Medline: 22640706]
7. Gouma DJ, Busch ORC. Quality of care a matter of experience. Ned Tijdschr Geneeskd 2007 Sep 22;151(38):2082-2086. [Medline: 17948822]
8. Hannan EL, Radzyner M, Rubin D, Dougherty J, Brennan MF. The influence of hospital and surgeon volume on in-hospital mortality for colectomy, gastrectomy, and lung lobectomy in patients with cancer. Surgery 2002 Jan;131(1):6-15. [doi: 10.1067/msy.2002.120238] [Medline: 11812957]
9. Kilsdonk MJ, Siesling S, van Dijk BAC, Wouters MW, van Harten WH. What drives centralisation in cancer care? PLoS One 2018 Apr 12;13(4):e0195673 [FREE Full text] [doi: 10.1371/journal.pone.0195673] [Medline: 29649250]
10. Institute of Medicine, National Cancer Policy Board, Committee on Cancer Survivorship: Improving Care and Quality of Life. In: Hewitt M, Greenfield S, Stovall E, editors. From Cancer Patient to Cancer Survivor: Lost in Transition. Washington, DC: National Academies Press; 2006.
11. Arraras JI, Greimel E, Sezer O, Chie W, Bergenmar M, Costantini A, et al. An international validation study of the EORTC QLQ-INFO25 questionnaire: an instrument to assess the information given to cancer patients. Eur J Cancer 2010 Oct;46(15):2726-2738. [doi: 10.1016/j.ejca.2010.06.118] [Medline: 20674333]

12. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. Health Serv Res 2005 Dec;40(6 Pt 1):1918-1930 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00438.x] [Medline: 16336556]
13. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital health strategies to fight COVID-19 worldwide: challenges, recommendations, and a call for papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
14. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun 18;3(2):e20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
15. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005 Feb 24;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
16. Penedo FJ, Oswald LB, Kronenfeld JP, Garcia SF, Cella D, Yanez B. The increasing value of eHealth in the delivery of patient-centred cancer care. Lancet Oncol 2020 May;21(5):e240-e251 [FREE Full text] [doi: 10.1016/S1470-2045(20)30021-8] [Medline: 32359500]
17. Clauser SB, Wagner EH, Aiello Bowles EJ, Tuzzio L, Greene SM. Improving modern cancer care through information technology. Am J Prev Med 2011 May;40(5 Suppl 2):S198-S207 [FREE Full text] [doi: 10.1016/j.amepre.2011.01.014] [Medline: 21521595]
18. Asan O, Nattinger AB, Gurses AP, Tyszka JT, Yen TW. Oncologists' views regarding the role of electronic health records in care coordination. JCO Clinical Cancer Informatics 2018 Dec(2):1-12. [doi: 10.1200/cci.17.00118]
19. de Wit ML, van Uden-Kraan CF, Lissenberg-Witte BI, Melissant HC, Fleuren MA, Cuijpers P, et al. Adoption and implementation of a web-based self-management application "Oncokompas" in routine cancer care: a national pilot study. Support Care Cancer 2019 Aug 18;27(8):2911-2920 [FREE Full text] [doi: 10.1007/s00520-018-4591-5] [Medline: 30564933]
20. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015 Feb 10;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
21. Melissant HC, Verdonck-de Leeuw IM, Lissenberg-Witte BI, Konings IR, Cuijpers P, Van Uden-Kraan CF. 'Oncokompas', a web-based self-management application to support patient activation and optimal supportive care: a feasibility study among breast cancer survivors. Acta Oncologica 2018 Feb 16;57(7):924-934. [doi: 10.1080/0284186x.2018.1438654]
22. Kooij L, Groen WG, van Harten WH. The effectiveness of information technology-supported shared care for patients with chronic disease: a systematic review. J Med Internet Res 2017 Jun 22;19(6):e221 [FREE Full text] [doi: 10.2196/jmir.7405] [Medline: 28642218]
23. Tighe SA, Ball K, Kensing F, Kayser L, Rawstorn JC, Maddison R. Toward a digital platform for the self-management of noncommunicable disease: systematic review of platform-like interventions. J Med Internet Res 2020 Oct 28;22(10):e16774 [FREE Full text] [doi: 10.2196/16774] [Medline: 33112239]
24. Hong YA, Hossain MM, Chou WS. Digital interventions to facilitate patient-provider communication in cancer care: A systematic review. Psychooncology 2020 Apr 13;29(4):591-603. [doi: 10.1002/pon.5310] [Medline: 31834650]
25. Xu A, Wang Y, Wu X. Effectiveness of e-health based self-management to improve cancer-related fatigue, self-efficacy and quality of life in cancer patients: Systematic review and meta-analysis. J Adv Nurs 2019 Dec 20;75(12):3434-3447. [doi: 10.1111/jan.14197] [Medline: 31566769]
26. Ayyoubzadeh SM, Kalhori SRN, Shirkhoda M, Mohammadzadeh N, Esmaeili M. Supporting colorectal cancer survivors using eHealth: a systematic review and framework suggestion. Support Care Cancer 2020 Aug 09;28(8):3543-3555. [doi: 10.1007/s00520-020-05372-6] [Medline: 32152763]
27. McCann L, McMillan KA, Pugh G. Digital interventions to support adolescents and young adults with cancer: systematic review. JMIR Cancer 2019 Jul 31;5(2):e12071 [FREE Full text] [doi: 10.2196/12071] [Medline: 31368438]
28. Slev VN, Mistiaen P, Pasman HRW, Verdonck-de Leeuw IM, van Uden-Kraan CF, Francke AL. Effects of eHealth for patients and informal caregivers confronted with cancer: A meta-review. Int J Med Inform 2016 Mar;87:54-67. [doi: 10.1016/j.ijmedinf.2015.12.013] [Medline: 26806712]
29. Triberti S, Savioni L, Sebri V, Pravettoni G. eHealth for improving quality of life in breast cancer patients: A systematic review. Cancer Treat Rev 2019 Mar;74:1-14. [doi: 10.1016/j.ctrv.2019.01.003] [Medline: 30658289]
30. Larson JL, Rosen AB, Wilson FA. The effect of telehealth interventions on quality of life of cancer patients: a systematic review and meta-analysis. Telemed J E Health 2018 Jun;24(6):397-405. [doi: 10.1089/tmj.2017.0112] [Medline: 29112484]
31. Institute of Medicine (US) Committee on Quality of Healh Care in America. In: Richardson W, Berwick D, Bisgard J, Bristow L, Buck C, Cassel C, editors. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press; 2001.
32. Bamidis PD, Paraskevopoulos E, Konstantinidis E, Spachos D, Billis A. Multimodal e-Health services for smoking cessation and public health: The SmokeFreeBrain Project approach. Stud Health Technol Inform 2017;245:5-9. [Medline: 29295041]
33. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009 Oct;62(10):1006-1012. [doi: 10.1016/j.jclinepi.2009.06.005] [Medline: 19631508]

34. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
35. Grol R, Wensing M. What drives change? Barriers to and incentives for achieving evidence-based practice. Med J Aust 2004 Mar 15;180(S6):S57-S60. [doi: 10.5694/j.1326-5377.2004.tb05948.x] [Medline: 15012583]
36. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. EFI 2018 Dec 18;34(4):285-291. [doi: 10.3233/efi-180221]
37. Baker TB, Hawkins R, Pingree S, Roberts LJ, McDowell HE, Shaw BR, et al. Optimizing eHealth breast cancer interventions: which types of eHealth services are effective? Transl Behav Med 2011 Mar 29;1(1):134-145 [FREE Full text] [doi: 10.1007/s13142-010-0004-0] [Medline: 21709810]
38. Børøsund E, Cvancarova M, Moore SM, Ekstedt M, Ruland CM. Comparing effects in regular practice of e-communication and Web-based self-management support among breast cancer patients: preliminary results from a randomized controlled trial. J Med Internet Res 2014 Dec 18;16(12):e295 [FREE Full text] [doi: 10.2196/jmir.3348] [Medline: 25525672]
39. De Regge M, Decoene E, Eeckloo K, Van Hecke A. Development and evaluation of an integrated digital patient platform during oncology treatment. J Patient Exp 2020 Feb 21;7(1):53-61 [FREE Full text] [doi: 10.1177/2374373518825142] [Medline: 32128372]
40. Groen WG, Kuijpers W, Oldenburg HS, Wouters MW, Aaronson NK, van Harten WH. Supporting lung cancer patients with an interactive patient portal: feasibility study. JMIR Cancer 2017 Aug 08;3(2):e10 [FREE Full text] [doi: 10.2196/cancer.7443] [Medline: 28790025]
41. Gustafson D, Hawkins R, McTavish F, Pingree S, Chen W, Volrathongchai K, et al. Internet-based interactive support for cancer patients: are integrated systems better? J Commun 2008 Jun;58(2):238-257 [FREE Full text] [doi: 10.1111/j.1460-2466.2008.00383.x] [Medline: 21804645]
42. Kuijpers W, Groen WG, Oldenburg HS, Wouters MW, Aaronson NK, van Harten WH. eHealth for breast cancer survivors: use, feasibility and impact of an interactive portal. JMIR Cancer 2016 May 10;2(1):e3 [FREE Full text] [doi: 10.2196/cancer.5456] [Medline: 28410178]
43. Ruland CM, Maffei RM, Børøsund E, Krahn A, Andersen T, Grimsbø GH. Evaluation of different features of an eHealth application for personalized illness management support: cancer patients' use and appraisal of usefulness. Int J Med Inform 2013 Jul;82(7):593-603. [doi: 10.1016/j.ijmedinf.2013.02.007] [Medline: 23507561]
44. Tiong SS, Koh E, Delaney G, Lau A, Adams D, Bell V, et al. An e-health strategy to facilitate care of breast cancer survivors: A pilot study. Asia Pac J Clin Oncol 2016 Jun 03;12(2):181-187. [doi: 10.1111/ajco.12475] [Medline: 26935343]
45. Alpert JM, Morris BB, Thomson MD, Matin K, Brown RF. Identifying how patient portals impact communication in oncology. Health Commun 2019 Nov 06;34(12):1395-1403 [FREE Full text] [doi: 10.1080/10410236.2018.1493418] [Medline: 29979886]
46. Baudendistel I, Winkler E, Kamradt M, Brophy S, Längst G, Eckrich F, et al. Cross-sectoral cancer care: views from patients and health care professionals regarding a personal electronic health record. Eur J Cancer Care (Engl) 2017 Mar 03;26(2):e12429. [doi: 10.1111/ecc.12429] [Medline: 26840784]
47. Ector GI, Westerweel PE, Hermens RP, Braspenning KA, Heeren BC, Vinck OM, et al. The development of a web-based, patient-centered intervention for patients with chronic myeloid leukemia (CMyLife): design thinking development approach. J Med Internet Res 2020 May 15;22(5):e15895 [FREE Full text] [doi: 10.2196/15895] [Medline: 32412424]
48. Geerts PAF, van der Weijden T, Loeffen PGM, Janssen LEF, Almekinders C, Wienhold TA, et al. Developing a patient portal for haematology patients requires involvement of all stakeholders and a customised design, tailored to the individual needs. BMC Med Inform Decis Mak 2019 Jul 11;19(1):129 [FREE Full text] [doi: 10.1186/s12911-019-0868-y] [Medline: 31296210]
49. Gerber D, Beg M, Duncan T, Gill M, Craddock Lee SJ. Oncology nursing perceptions of patient electronic portal use: a qualitative analysis. Oncol Nurs Forum 2017 Mar 01;44(2):165-170 [FREE Full text] [doi: 10.1188/17.ONF.165-170] [Medline: 28222081]
50. Girault A, Ferrua M, Lalloué B, Sicotte C, Fourcade A, Yatim F, et al. Internet-based technologies to improve cancer care coordination: current use and attitudes among cancer patients. Eur J Cancer 2015 Mar;51(4):551-557. [doi: 10.1016/j.ejca.2014.12.001] [Medline: 25661828]
51. Kildea J, Battista J, Cabral B, Hendren L, Herrera D, Hijal T, et al. Design and development of a person-centered patient portal using participatory stakeholder co-design. J Med Internet Res 2019 Feb 11;21(2):e11371 [FREE Full text] [doi: 10.2196/11371] [Medline: 30741643]
52. Kuijpers W, Groen WG, Loos R, Oldenburg HSA, Wouters MWJM, Aaronson NK, et al. An interactive portal to empower cancer survivors: a qualitative study on user expectations. Support Care Cancer 2015 Sep 27;23(9):2535-2542 [FREE Full text] [doi: 10.1007/s00520-015-2605-0] [Medline: 25876157]
53. McCleary NJ, Greenberg TL, Barysauskas CM, Guerette EJ, Hassan M, Jacobson JO, et al. Oncology patient portal enrollment at a comprehensive cancer center: a quality improvement initiative. J Oncol Pract 2018 Aug;14(8):e451-e461. [doi: 10.1200/jop.17.00008]

54. Escriva Boulley G, Leroy T, Bernetière C, Paquienseguy F, Desfriches-Doria O, Préau M. Digital health interventions to help living with cancer: A systematic review of participants' engagement and psychosocial effects. Psychooncology 2018 Dec 24;27(12):2677-2686. [doi: 10.1002/pon.4867] [Medline: 30152074]
55. Lamers RED, Cuypers M, Husson O, de Vries M, Kil PJM, Ruud Bosch JLH, et al. Patients are dissatisfied with information provision: perceived information provision and quality of life in prostate cancer patients. Psychooncology 2016 Jun 25;25(6):633-640. [doi: 10.1002/pon.3981] [Medline: 26403417]
56. Sim J, Shin J, Park S, Chang Y, Shin A, Noh D, et al. Association between information provision and decisional conflict in cancer patients. Ann Oncol 2015 Sep;26(9):1974-1980 [FREE Full text] [doi: 10.1093/annonc/mdv275] [Medline: 26116430]
57. Kessels RPC. Patients' memory for medical information. J R Soc Med 2017 Nov 23;96(5):219-222. [doi: 10.1177/014107680309600504]
58. van der Hout A, van Uden-Kraan CF, Holtmaat K, Jansen F, Lissenberg-Witte BI, Nieuwenhuijzen GAP, et al. Role of eHealth application Oncokompas in supporting self-management of symptoms and health-related quality of life in cancer survivors: a randomised, controlled trial. Lancet Oncol 2020 Jan;21(1):80-94. [doi: 10.1016/S1470-2045(19)30675-8] [Medline: 31838009]
59. Lorca-Cabrera J, Grau C, Martí-Arques R, Raigal-Aran L, Falcó-Pegueroles A, Albacar-Riobóo N. Effectiveness of health web-based and mobile app-based interventions designed to improve informal caregiver's well-being and quality of life: A systematic review. Int J Med Inform 2020 Feb;134:104003. [doi: 10.1016/j.ijmedinf.2019.104003] [Medline: 31790857]
60. Heck KE, Pamuk ER. Explaining the relation between education and postmenopausal breast cancer. Am J Epidemiol 1997 Feb 15;145(4):366-372. [doi: 10.1093/oxfordjournals.aje.a009114] [Medline: 9054241]

Abbreviations

e-Consult: electronic consultation
eHealth: electronic health
EMR: electronic medical record
HCP: health care professional
MMAT: Mixed Methods Appraisal Tool
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROM: patient-reported outcome measure
PROSPERO: Prospective Register of Systematic Reviews
RCT: randomized controlled trial

Edited by R Kukafka; submitted 17.03.21; peer-reviewed by W van Harten, AW Zahoor; comments to author 07.04.21; revised version received 17.05.21; accepted 30.06.21; published 24.09.21.

Please cite as:
Hopstaken JS, Verweij L, van Laarhoven CJHM, Blijlevens NMA, Stommel MWJ, Hermens RPMG
Effect of Digital Care Platforms on Quality of Care for Oncological Patients and Barriers and Facilitators for Their Implementation: Systematic Review
J Med Internet Res 2021;23(9):e28869
URL: https://www.jmir.org/2021/9/e28869
doi:10.2196/28869
PMID:34559057

Viewpoint

A Road Map for Remote Digital Health Technology for Motor Neuron Disease

Ruben P A van Eijk[1,2], PhD; Anita Beelen[3,4], PhD; Esther T Kruitwagen[3,4], MD; Deirdre Murray[5,6], PhD; Ratko Radakovic[7,8,9,10,11], PhD; Esther Hobson[12], PhD; Liam Knox[12], PhD; Jochem Helleman[3,4], MSc; Tom Burke[5,13], PhD; Miguel Ángel Rubio Pérez[14], PhD; Evy Reviers[15], MSc; Angela Genge[16], PhD; Frederik J Steyn[17,18,19], PhD; Shyuan Ngo[18,19,20,21], PhD; John Eaglesham[22], PhD; Kit C B Roes[23], PhD; Leonard H van den Berg[1], PhD; Orla Hardiman[24,25], MD; Christopher J McDermott[12], PhD

[1]UMC Utrecht Brain Centre, University Medical Centre Utrecht, Utrecht, Netherlands
[2]Biostatistics & Research Support, Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht, Netherlands
[3]Department of Rehabilitation, University Medical Centre Utrecht, Utrecht, Netherlands
[4]Center of Excellence for Rehabilitation Medicine, University Medical Centre Utrecht and De Hoogstraat Rehabilitation, Utrecht, Netherlands
[5]Academic Unit of Neurology, Trinity College Dublin, Dublin, Ireland
[6]Department of Physiotherapy, Beaumont Hospital, Dublin, Ireland
[7]Faculty of Medicine and Health Sciences, University of East Anglia, Norwich, United Kingdom
[8]Euan MacDonald Centre for Motor Neuron Disease Research, University of Edinburgh, Edinburgh, United Kingdom
[9]Norfolk and Norwich University Hospital, Norwich, United Kingdom
[10]Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh, United Kingdom
[11]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, United Kingdom
[12]Department of Neuroscience, Sheffield Institute for Translational Neuroscien, University of Sheffield, Sheffield, United Kingdom
[13]Department of Psychology, Beaumont Hospital, Dublin, Ireland
[14]Servicio de Neurología, Hospital del Mar, Barcelona, Spain
[15]European Organization for Professionals and Patients with ALS (EUpALS), Leuven, Belgium
[16]Department of Neurology, Montreal Neurological Institute, McGill University, Montreal, QC, Canada
[17]School of Biomedical Sciences, Faculty of Medicine, University of Queensland, Brisbane, Australia
[18]The Royal Brisbane and Women's Hospital, Herston, Australia
[19]Wesley Medical Research, the Wesley Hospital, Auchenflower, Australia
[20]Centre for Clinical Research, University of Queensland, Brisbane, Australia
[21]Australian Institute for Bioengineering and Nanotechnology, University of Queensland, Brisbane, Australia
[22]Advanced Digital Innovation (UK) Ltd, Salts Mill, United Kingdom
[23]Department of Health Evidence, Section Biostatistics, Radboud Medical Centre Nijmegen, Nijmegen, Netherlands
[24]Department of Neurology, National Neuroscience Centre, Beaumont Hospital, Dublin, Ireland
[25]FutureNeuro SFI Research Centre, Royal College of Surgeons in Ireland, Dublin, Ireland

Corresponding Author:
Ruben P A van Eijk, PhD
UMC Utrecht Brain Centre
University Medical Centre Utrecht
Heidelberglaan 100
Utrecht, 3584 CX
Netherlands
Phone: 31 88 75 579 39
Fax: 31 88 75 554 94
Email: r.p.a.vaneijk-2@umcutrecht.nl

Abstract

Despite recent and potent technological advances, the real-world implementation of remote digital health technology in the care and monitoring of patients with motor neuron disease has not yet been realized. Digital health technology may increase the

accessibility to and personalization of care, whereas remote biosensors could optimize the collection of vital clinical parameters, irrespective of patients' ability to visit the clinic. To facilitate the wide-scale adoption of digital health care technology and to align current initiatives, we outline a road map that will identify clinically relevant digital parameters; mediate the development of benefit-to-burden criteria for innovative technology; and direct the validation, harmonization, and adoption of digital health care technology in real-world settings. We define two key end products of the road map: (1) a set of reliable digital parameters to capture data collected under free-living conditions that reflect patient-centric measures and facilitate clinical decision making and (2) an integrated, open-source system that provides personalized feedback to patients, health care providers, clinical researchers, and caregivers and is linked to a flexible and adaptable platform that integrates patient data in real time. Given the ever-changing care needs of patients and the relentless progression rate of motor neuron disease, the adoption of digital health care technology will significantly benefit the delivery of care and accelerate the development of effective treatments.

(J Med Internet Res 2021;23(9):e28766) doi:10.2196/28766

KEYWORDS

amyotrophic lateral sclerosis; digital health care technology; e-health

Introduction

Remote digital health technology, ranging from simple mobile apps to implantable devices, will reform the delivery of care. Despite the recent technological advances in biosensors, mobile communications, and cloud computing, their real-world implementation remains to be fully realized [1]. Motor neuron disease (MND) is a debilitating disorder in which digital health care technology will benefit delivery of care [2,3] and also expedite the development of effective treatments [4-6]. Patients living with MND rapidly lose their functional independence, making travel, communication, and visits to specialist clinics for assessment of vital clinical parameters challenging [7]. This has been particularly the case during the COVID-19 pandemic [8,9]. The effects of the pandemic have focused attention on technological advances that increase the accessibility and personalization of care, and remote biosensors that optimize the collection of vital clinical parameters, irrespective of the patients' ability to visit the clinic, are poised to revolutionize the clinical encounter. These developments hold benefits for both research and care settings, driving the implementation of real-world use of remote digital health care technology in MND [2,10].

Notwithstanding, the following key barriers delay the adoption and wide-scale implementation of digital health care technology in MND [2]: (1) technical complexities, (2) low compliance rates, (3) time investment required, (4) high costs, and (5) legislation challenges. For clinical trials, adoption is further complicated by regulatory hurdles because of reduced auditability, consistency, and data quality [6,11,12]. Although there is considerable work underway in these areas that will facilitate the validation and initial adoption of digital technology in MND [4,5,13-15], it is also the case that device or platform incompatibility might further delay their wide-scale adoption and large-scale data aggregation. Therefore, there is a recognized need to open collaborations among clinicians, engineers, and technology companies that will support early implementation.

Here, we outline a road map that aims to (1) facilitate the identification of clinically relevant digital parameters; (2) mediate the development of benefit-to-burden criteria for innovative technology; and (3) direct the validation, harmonization, and adoption of digital health care technology in real-world settings.

The Need for Remote Technology in MND

Disease heterogeneity and rapid physical deterioration are key drivers in the delivery of care and monitoring of disease progression in clinical trials. Although MND is recognized as a clinical entity, it is better considered as a number of different subgroups with variable clinical presentations, different causes, and different disease trajectories. A flexible approach toward management is required such that the patient's clinical condition is monitored at intervals that best reflect both the needs of the patient and their family and the trajectory of the disease. As progression ultimately leads to severe disability, clinical attendance is increasingly burdensome [16], notwithstanding the known benefits of multidisciplinary clinics that include survival and optimal access to assistive devices or proven treatments [17]. Unfortunately, even in developed countries, attendance at multidisciplinary clinics is as low as 43%, and many of those who can attend are unable to return [18].

We illustrate the extent of this attrition process, as observed in a large clinical trial [19], in Figure 1. From the point of trial enrollment, almost 1 in 4 (184/750, 24.5%) patients failed to visit the clinic after 12 months, and more than one-third (271/750, 36.1%) were unable to provide reliable respiratory information. These figures are likely an underestimation of the attrition rates occurring at multidisciplinary clinics because clinical trials select for patients with better prognoses [20,21]. These high attrition rates affect the estimation of treatment benefit within clinical trials, reduce statistical power, and may lead to the continuation of an ineffective compound into a subsequent phase of clinical development.

Figure 1. Percentage of EMPOWER patients with missing information on clinical outcome measures. For each clinical visit in the EMPOWER study, we determined the number of patients that were on study medication and were alive (ie, number at risk). Subsequently, we calculated the proportion of missing observations per outcome measure. If the patient was no longer able to visit the clinic (as indicated by the number of missing in-clinic muscle strength assessments), ALSFRS-R data and blood samples were collected remotely by home visit or by phone. ALSFRS-R: Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised.

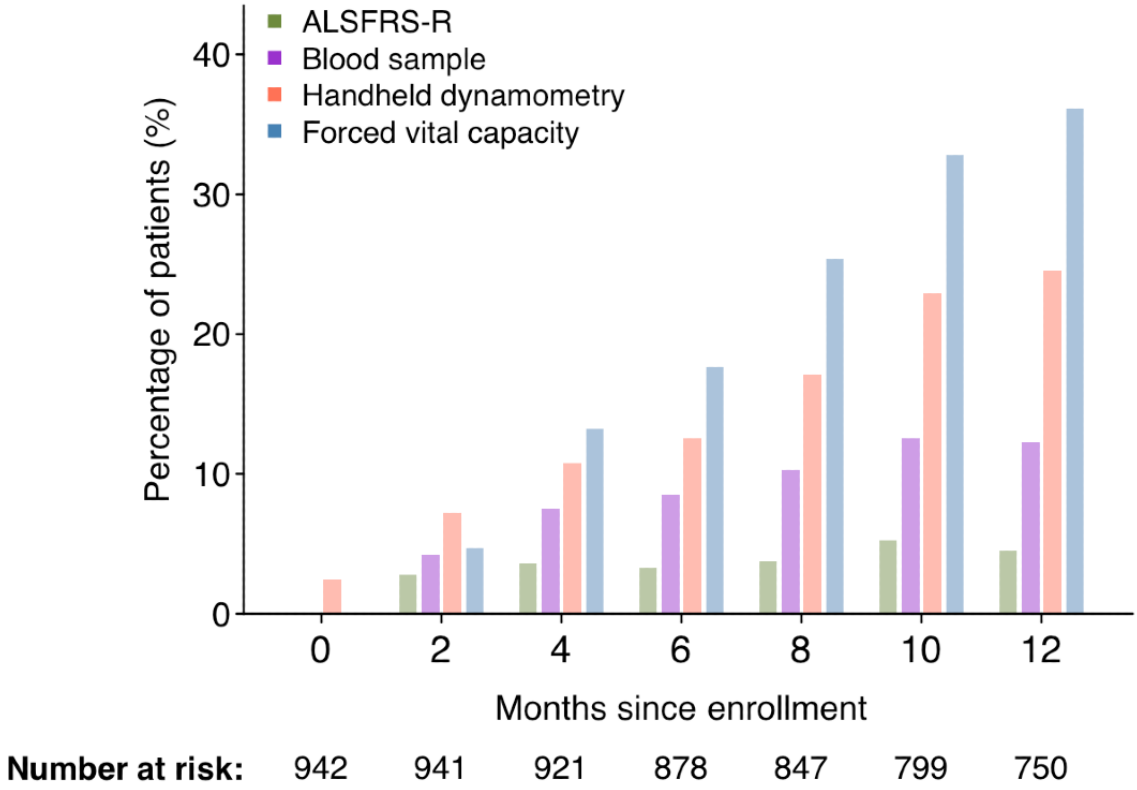


Figure 1 also demonstrates that outcomes that can be measured remotely, either by phone for the Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised (ALSFRS-R) or by a home visit for a blood sample, can reduce attrition and may mitigate many of the current challenges. Remote data collection can be tailored to identify and accommodate the ongoing needs of patients and caregivers, thus shortening the time spent at the hospital or traveling to the hospital and personalizing visiting schemes [14,22]. This in turn can provide a communication channel between patients and health care providers that builds an iterative and shared decision-making process regarding the timing and management of interventions and the approaches to end of life [23]. For clinical trials, home-based monitoring of disease progression is likely to more accurately reflect the patients' true physical condition in free-living (ie, nonclinical) settings and lower the burden of monitoring. This in turn is likely to increase the number of patients who can participate in clinical research and provide real-world insight into the therapeutic value of experimental drugs [6,11,24]. Designing an information source that reflects the patients' true physical condition in near real time is therefore of significant value for all stakeholders.

A Road Map for MND

In this paper, we considered technology that assists with remote monitoring of MND disease progression or identifies critical health issues such as the development of respiratory or nutritional failure. To that end, we defined two key end products of the road map, as follows:

1. A set of reliable digital parameters to capture data, collected under *free-living* conditions at the patient's home, that reflect patient-centric measures and facilitate clinical decision making
2. An integrated open-source system that provides personalized feedback to patients, health care providers, clinical researchers, and caregivers

To achieve this, there is a need for input from all stakeholders, including those living with MND and those who (informally) care for patients with MND. To promote the uptake of new technology in real-world settings, we used a user-centered co-design approach that involves input from stakeholders (eg, patients, caregivers, physicians, ethicists, regulators, programmers, and researchers) at every stage of development [25]. There will be specific attention paid to patient, caregiver, and researcher access as well as international regulatory, financial, and licensing hurdles to ensure widespread implementation, sustainability, and growth [26]. For this road map, we have identified the following three key questions: (1) what to measure; (2) how to measure; and (3) how to implement? We outlined the requirements to answer each of these questions, considering the needs of the various stakeholders of the technology.

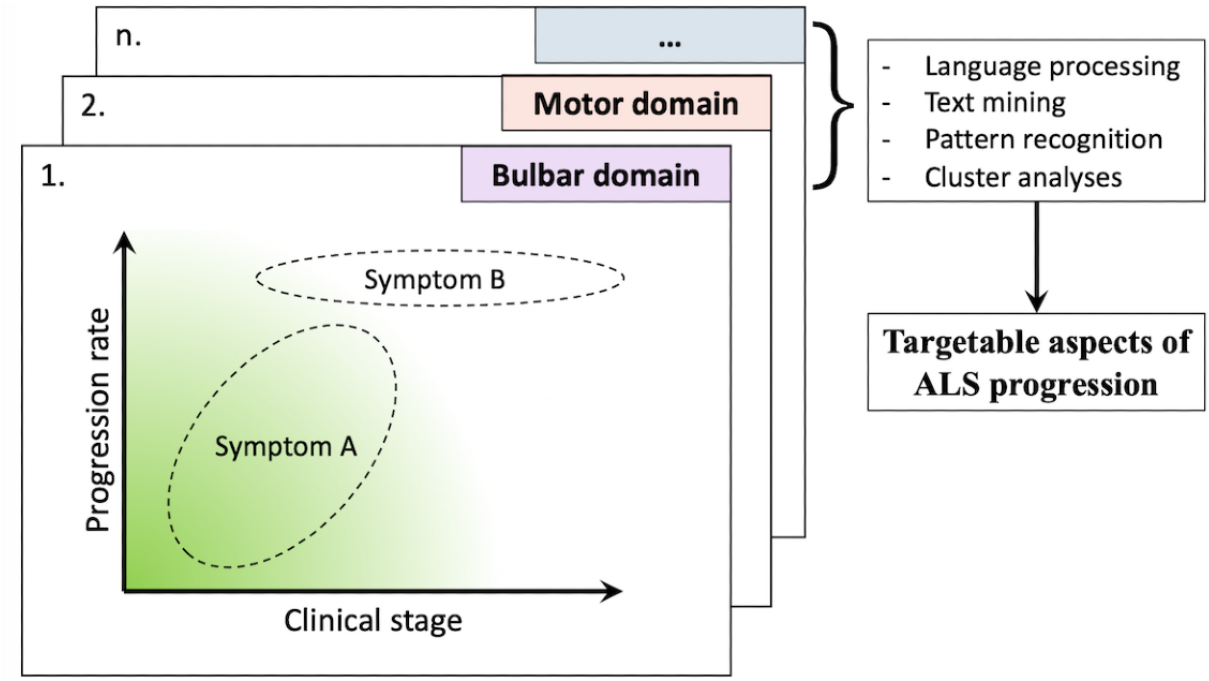
What to Measure?

Patient centricity is crucial to promote long-term adherence, where end points should focus on how aspects of MND affect daily living and distinguish between symptoms (ie, impairments) and activities (ie, disabilities). The clinician's input may help to prioritize end points in clinically relevant domains or supplement other aspects of MND that patients may not recognize as clinically important (eg, weight loss or cognitive and behavioral changes). This approach maximizes the likelihood of acquiring data that facilitate clinically relevant decision making and minimizes the collection of redundant information.

Workplan: What to Measure?

Information will be collected from three independent sources: (1) a systematic literature review summarizing and meta-analyzing all published symptomology associated with MND; (2) face-to-face in-depth interviews and focus group discussions with patients, caregivers, and physicians until data saturation is reached; and (3) a cross-sectional, population-based patient survey. We will apply a purposive and theoretical sampling strategy to ensure that we capture the complete range of disease patterns reported among patients with MND [27]. Videoconferencing will be used to ensure full patient participation at all meetings. In Figure 2, we illustrate two key elements that may help to disentangle the clinical heterogeneity observed among patients and guide sampling strategies or evaluate generalizability: clinical disease stage and progression rate. Some symptoms such as speech disturbances may only be prevalent in early clinical stages and will disappear over time owing to a complete loss of function (illustrated as *Symptom A*). Conversely, some symptoms may only be prevalent in patients with a certain disease type or progression rate (eg, symptoms related to spasticity in a primarily upper motor neuron, a slowly progressing subtype of disease, illustrated as *Symptom B*). Mapping the prevalence of symptoms in this 2D plane may help to disentangle a large part of the clinical heterogeneity in MND. Clinical stage can be defined by the King's clinical and Milano-Torino functional staging algorithms [28,29], whereas the progression rate is directly related to survival time and could be mapped by, for example, the patient's predicted prognosis [30].

Figure 2. Untangling clinical heterogeneity in motor neuron disease in a 3D framework. The prevalence of symptoms as reported by patients will be mapped according to their clinical stage and (predicted) progression rate. Clusters of symptoms will be defined into domains (eg, bulbar, motor, respiratory, and cognition) for creating a 3D map of motor neuron disease symptom heterogeneity. The map will be used to identify clinically relevant and targetable end points. ALS: amyotrophic lateral sclerosis.



A complete list of symptoms will be extracted from the interviews and focus group meetings, supplemented with data from the literature review, and by language processing and text mining from a population-based, cross-sectional survey asking open questions such as "What bothers you the most about your MND" and "In what way does this affect your daily living?" Information collection will be conducted in a multicenter, international setting to account for potential geographical, social, and cultural differences. Finally, a large range of statistical methods will be used for pattern recognition to cluster targets and end points, identify redundant information, and classify subgroups of patients with specific needs. Ultimately, this results in a 3D framework, as depicted in Figure 2, that orders a large part of the variability in symptoms among patients. Additional classification using, for example, cognitive or behavioral impairment or the presence of dementia may be applied to further elucidate heterogeneity in MND [31,32]. The 3D framework could help to evaluate the representativeness of the included population and quickly identify clinically relevant and targetable end points. The final list of targetable end points will be evaluated by patients, caregivers, and physicians for further refinement and prioritization.

How to Measure?

The second step is to translate the list of clinically relevant aspects of MND into digital parameters. For example, speech is a clinically relevant outcome in which a mobile speech app could serve to improve its objective assessment in a standardized manner. Critical aspects to consider here are, among other things, whether the mobile app requires an action-dependent or passive task (eg, repeat a fixed sentence vs record a daily 5-min conversation), the uniformity and precision of the digital voice parameters (eg, speaking rate vs volume vs perturbations) [33], the required number and length of measurements (eg, daily, weekly, or monthly follow-up), and whether direct supervision or caregiver assistance is required. These factors not only determine the burden for patients or caregivers but also establish the value of the obtained information for researchers and health care providers. The aim is to define selection criteria for digital technology that evaluates burden-to-information ratios and, based on these criteria, results in a standardized and protocolized set of digital metrics for MND.

Workplan: How to Measure?

We will conduct a series of systematic reviews to define a list of candidate technologies for each targetable aspect of MND. Naturally, the real-world use of the suggested technology could reveal insights that may not have been reported previously in the literature or by the supplier. Longitudinal studies are preferred because they not only evaluate test-retest reliability and the validity to capture disease progression of a digital parameter but they also provide information on patient adherence and the adequacy of the monitoring protocol (eg, obtaining daily, weekly, or monthly measurements). The latter is an often-overlooked aspect, although it has important consequences for all stakeholders. Naturally, a daily monitoring scheme would provide the most information about the disease for health care providers and researchers [5] but could disproportionally increase patient burden and lead to high (differential) attrition. A periodic monitoring scheme may be more successful, and a data-driven optimization of this trade-off, while accounting for patient and health care provider preferences, should be used.

As a practical illustration, in Table 1, we show the benefit of changing the monitoring frequency from monthly to daily for two end points: ALSFRS-R and daily physical activity [4]. The benefit of more or less frequent monitoring strongly depends on the end point, and as demonstrated by the ALSFRS-R, it could be questionable whether high-frequency monitoring is always beneficial. In addition, because of a difference in progression rates among patients, the monitoring frequency could be further optimized for each individual patient (eg, a patient who slowly progresses may require less frequent monitoring). Ultimately, the optimal monitoring scheme is likely to differ from digital end point to end point, and longitudinal data are essential to estimate a data-driven optimal burden-to-information ratio for each device or mobile app [4].

Table 1. Sample size estimates for the Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised (ALSFRS-R) and daily activity with varying monitoring frequencies.[a]

Characteristic	ALSFRS-R	Daily activity
Monthly progression rate[b]	−0.06	−0.05
Between-patient variability, $\sigma^2_{between}$	0.06^2	0.04^2
Within-patient variability, σ^2_{within}	0.15^2	0.54^2
Required sample size (6-month follow-up), n (% difference)		
Monthly	226 (ref[c])	1208 (ref)
Biweekly	210 (−7)	805 (−33)
Weekly	199 (−12)	522 (−57)
Daily	187 (−17)	214 (−82)
Required sample size (12-month follow-up), n (% difference)		
Monthly	191 (ref)	322 (ref)
Biweekly	189 (−1)	251 (−22)
Weekly	187 (−2)	209 (−35)
Daily	185 (−3)	168 (−48)

[a]Daily activity was defined as the proportion of time that the patient was nonsedentary, as described elsewhere [4]. Sample size calculations are based on a standardized linear mixed model; data are based on 42 patients [4]. Sample size calculations assumed a target power of 80%, 2-sided α of 5%, and a 25% reduction in the progression rate; numbers are the required sample size per group [34].

[b]Progression rate is expressed as number of SD per month. Both the ALSFRS-R and daily activity were standardized to make a direct comparison possible. The ALSFRS-R has, on average, a faster progression rate, more variability between patients, and less variability within patients compared with daily activity.

[c]ref: reference.

There are circumstances in which digital technology may not appropriately quantify disease symptoms (eg, changes in mood or alterations in cognition). Therefore, an important aspect of the road map is to additionally develop scales and questionnaires that can be administered remotely to quantify clinically relevant, subjective symptoms [35]. It is important for these patient- and caregiver-reported outcome measures to consider elements such as the ease of engagement (eg, mobile apps with multiple verbal and nonverbal response formats), the length of questionnaires, the validity of questions (eg, differences between patient-reported and in-clinic outcomes), and the viability for digitization. Ultimately, it may prove critical to harmonize and integrate passive digital methods such as accelerometry with task-dependent methods and questionnaires to derive the most reliable and representative state of disease at any given time.

In addition, an important goal is to maximize the adherence and retention of the patient or caregiver. In two studies that reported on remote monitoring with personalized feedback in care settings [14,22], 80%-87% of the patients provided regular ALSFRS-R information. In two recent studies that solely monitored disease progression without providing care, this was reduced to 24%-56% [15,36]. It seems critical to provide an incentive to participate, where patients do not simply upload their data into the cloud but also receive something in return, such as care, tailored information, or personalized insights. An understanding of how best to retain patients in the digital environment (eg, by incorporating behavioral or motivational techniques) and which patient factors are related to digital uptake or adherence are fundamental to the success of the workplan. Finally, it is important to consider the geographical region and the local facilities and demographics. Older populations or patients from developing areas with minimal financial, technological, or educational facilities may encounter challenges when using devices. This could increase the risk of care asymmetries across populations and should be considered when comparing candidate technologies.

How to Implement?

A standardized format or platform for multiple digital metrics is not yet available, and a priority of the Treatment Research Initiative to Cure ALS will be to provide this technology [37]. This platform will solve the problem of collecting data relating to multiple domains while using multiple apps from different providers. Such a platform is essential because clinical decision making is often multidimensional and requires insight into various aspects of the patient simultaneously. The aim of this final part of the road map is to (1) integrate different data sources required by the end users and (2) deploy an open-source framework product in which one can flexibly integrate different third-party solutions, evaluate the performance of digital metrics compared with clinically relevant events, and allow in-depth analysis of health care expenditures and outcomes.

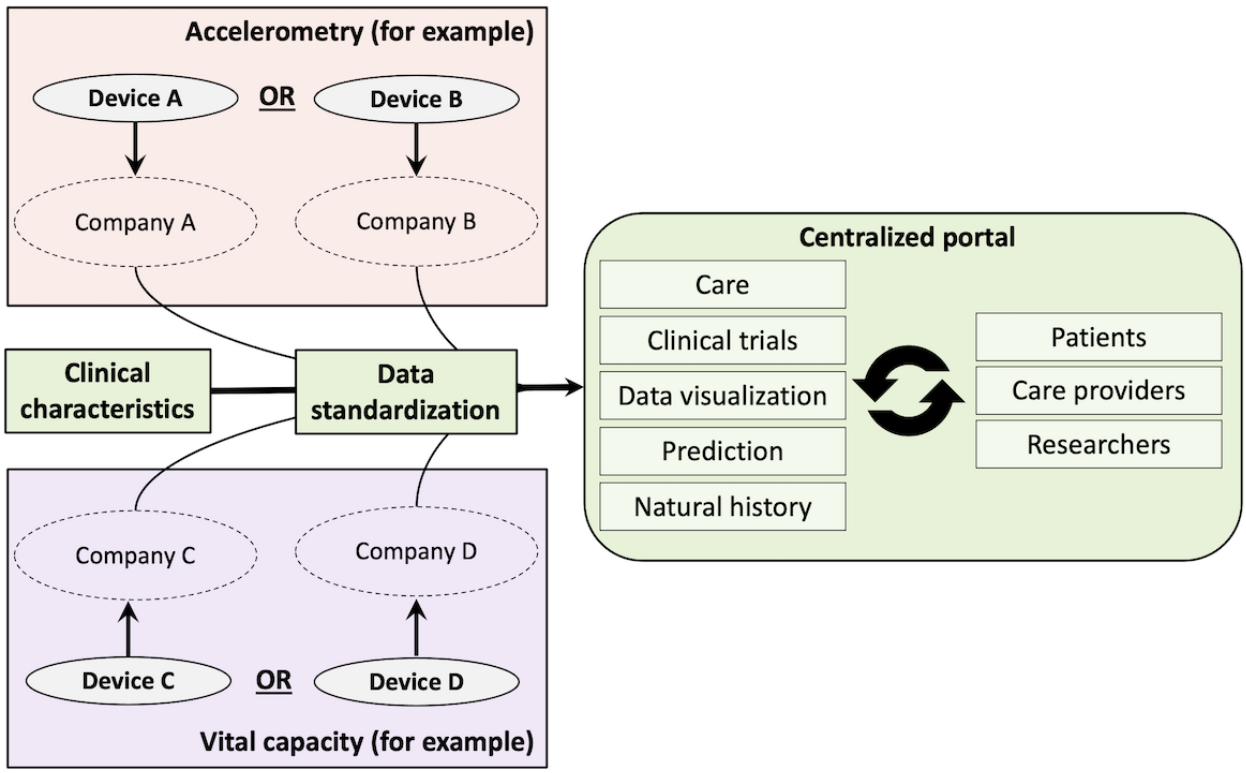
Workplan: How to Implement?

In Figure 3, we schematically illustrate the final product. The first step would be to standardize the metrics that are of interest across different devices. The standardized metrics are subsequently stored in a central database that acts as a single integrative platform. The capability of the device to allow uninterrupted, remote data access is therefore critical; the final selection of devices or mobile apps is likely an iterative process in close relationship with the *How to Measure* section. Although the concept is simple, this part of the road map may be the most challenging because of, among other things, privacy, regulatory, ethical, ownership, licensing, certification, and database challenges. A close collaboration among various industry, academic, funding, and regulatory parties has been initiated to provide a clear central and local governance and management structure with meticulous planning.

The next step is to translate the centralized data into useful information and disseminate insights to the relative end user. Given the extent of information acquired and the number of different stakeholders and settings, a customized approach will be required. Where a trialist may be primarily interested in a historical benchmark of group averages, clinicians and patients will be principally interested in individual disease patterns. These interests need to be carefully documented, and a dynamic, user-centered co-design approach is vital. It is important to evaluate how the implementation of digital parameters affects existing standard operating procedures and work practices or how health care professionals adhere to new MND technologies [38]. For care purposes, defining optimal *activation* or *flagging* rules is important for timely activation of additional care. For example, thresholds are needed for respiratory function (or decline in function) to alert the patient's health care provider or automatically refer the patient to a pulmonologist when the threshold is reached. Defining these rules can help to standardize remote monitoring of MND, and this requires the development of a disease-specific protocol based on MND guidelines and expert opinions [14].

In addition, efforts are required to evaluate different strategies for effective communication among users, for example, the implementation of chat functions between patients and health care providers or the provision of personalized information or feedback on the obtained measurements. By making use of open-source software platforms, tailored programs can be developed in a dynamic interface, while seamlessly integrating the continuously updating information from wearables and mobile apps. We have previously provided an example of such a platform for clinical trial design [39]. The platform and interactive platform will go through several stages before the product can be finalized, for example, prototype development, iterative testing phases, experimenting with varying key parameters (eg, user burden, monitoring frequency, different wording and timings, and different prompting regimes), pilot releases, and upgrading, taking into account the input from all users at every stage.

Figure 3. The harmonization and centralization of digital outcome data before dissemination. A schematic illustration of the integration of different devices from different third parties into a central database that acts as a single integrative platform is illustrated for vital capacity and accelerometry. Information and insights can subsequently be disseminated interactively to the respective end users.



Finally, only by real-world implementation can the true value of digital parameters and platforms be evaluated. For care settings, it is important to conduct dedicated studies to evaluate the cost-effectiveness of the digital health care technology and its effect on clinical outcomes. The gold standard would be randomized controlled trials, for example, allocating patients to usual care or digital assisted care or allocating hospitals in a cluster randomized study. An example of the former has been published previously, in which the authors evaluated the effect of home telemonitoring of noninvasive ventilation on the number of hospital visits [40], together with its cost-effectiveness [41]. Given the complexity of most eHealth interventions [13], such trials are challenging and it may not be immediately clear what an ideal efficacy end point would be (eg, survival time, number of hospital admissions, costs, or patient-reported indicators?). A key aspect for successful uptake is to have a financial model that is sustainable, while providing clinically relevant benefits to all stakeholders.

For clinical trials, the cost aspect may be of similar importance, and studies evaluating the gain in retention compared with cost would be highly insightful. Besides the potential reduction in attrition rates and gain in information quality, it will be necessary to determine whether digital metrics can ultimately replace the clinical efficacy end point (ie, surrogacy) and increase trial efficiency. The real-world systematic use of digital health care technology, alongside common clinical efficacy endpoints, is therefore a key first step. The potential use of real-world data originating from digital health care technology in clinical drug development has been recognized for regulatory purposes [42]. Nevertheless, important challenges remain, such as the design of interventional studies and challenges related to data availability, data quality, auditability, and the completeness of electronic phenotyping [43].

Data Security and Privacy Regulations

Given the current data security and privacy regulations for medical devices and personal data (eg, General Data Protection Regulation), data security and ethical hurdles will have a major impact on the developmental and operational aspects of the platform and on how the data may be used during (international) research projects or within care settings. These aspects are not only important for the platform itself but also apply to, for example, cloud services from third parties. How, when, and by whom data can be accessed need to be tailored and defined for each participating site individually to comply with local laws and regulations. Making use of existing infrastructures, for example, by integrating the platform into eHealth care services and requiring two-factor authentication could facilitate secure data access [14], but its feasibility depends on the available local facilities. In addition, considerations for using the data include obtaining consent to (1) share personally identifiable information for direct care only, (2) share personally identifiable information for research and care purposes, or (3) share anonymized data to be used for research only. Similarly, there is a need to address challenges in intellectual property and data ownership (eg, where the demarcation might be made between clinical and technical intellectual property).

An important consideration is the CE (Conformité Européenne) marking for medical devices and software. The disadvantage

of CE marking an entire ecosystem, as presented in Figure 3, is that it may significantly constrain the rate of change and evolution of the system. CE marking locks down the design, and any significant changes made to the system (eg, addition of new algorithms or devices) requires reapproval by external regulatory bodies. A consideration could be to use a mixed approach of using CE-marked devices and algorithms interconnected and facilitated by non–CE-marked software. This will require some design constraints to make it possible (eg, there can be no opacity in terms of non–CE-marked algorithms supporting clinical decision making), and achieving this goal will require careful design and partitioning from the outset. Another strategy is to consolidate and rationalize into a single CE-marked system once the ecosystem for ALS, including measurements, algorithms, and protocols, reaches maturity. These challenges will require dedicated strategies, which will need to be developed with the involvement of all relevant stakeholders and by obtaining expertise from ethicists, regulators, and data privacy officers. Ultimately, these discussions may affect the choice of measurement or device, indicating that the final set of digital parameters for MND is likely defined during a dynamic and iterative process rather than being fixed from the outset.

Conclusions

In conclusion, this road map aims to align the current developments in digital health care technology for MND and initiate a collaborative effort to mediate its wide-scale adoption across MND clinics and clinical trials. We defined the following three key questions, the answers to which are critical to achieve this aim: (1) what to measure; (2) how to measure; and (3) how to implement? Each question requires a dedicated study methodology to overcome potential implementation, adoption, or regulatory hurdles, which are summarized in Table 2.

Table 2. Summary of the road map for remote digital health technology for motor neuron disease (MND).

Question	Aim	Method	Result
What to measure?	To define clinically relevant, patient-centric, and targetable aspects of MND	• Inventorying symptomology associated with MND by systematic literature review, face-to-face in-depth interviews, focus group discussions, and population-based patient surveys	List of targetable end points for MND
How to measure?	To translate targetable end points for MND into digital parameters	• Candidate technologies are identified by systemic literature reviews. Standardized longitudinal studies are initiated to evaluate test-retest reliability and validity to capture disease progression and to assess protocol adherence.	Set of standardized and protocolized digital metrics
How to implement?	To centralize digital metrics in an integrative platform and disseminate user-dependent information	• Create a single database using open-source software platforms, reactive programming, and third-party cloud services • Initiation of dedicated studies to evaluate health care expenditures, cost-effectiveness outcomes, and surrogate biomarker value	Scalable and validated platform to provide tailored information for care and research settings

Given the ever-changing care needs of patients and the relentless progression rate of MND, aligning our current endeavors and facilitating the adoption of digital health care technology will significantly benefit the delivery of care and accelerate the development of effective treatments against this debilitating disease.

Conflicts of Interest

JE has a commercial interest in promoting the future use of remote digital technology platforms in Motor Neurone Disease. CJM has undertaken consultancy work for Biogen, Orion, Orphazyme, Merz and is supported by the NIHR Sheffield Biomedical Research Centre (BRC). SN is supported by the FightMND Mid-Career Research Fellowship. The remaining authors have no conflicts of interest to declare.

References

1. Espay AJ, Bonato P, Nahab FB, Maetzler W, Dean JM, Klucken J, Movement Disorders Society Task Force on Technology. Technology in Parkinson's disease: challenges and opportunities. Mov Disord 2016 Sep;31(9):1272-1282 [FREE Full text] [doi: 10.1002/mds.26642] [Medline: 27125836]
2. Helleman J, Kruitwagen ET, van den Berg LH, Visser-Meily JM, Beelen A. The current use of telehealth in ALS care and the barriers to and facilitators of implementation: a systematic review. Amyotroph Lateral Scler Frontotemporal Degener 2020 May;21(3-4):167-182. [doi: 10.1080/21678421.2019.1706581] [Medline: 31878794]
3. Paganoni S, Simmons Z. Telemedicine to innovate amyotrophic lateral sclerosis multidisciplinary care: the time has come. Muscle Nerve 2019 Jan;59(1):3-5. [doi: 10.1002/mus.26311] [Medline: 30066337]

4. van Eijk RP, Bakers JN, Bunte TM, de Fockert AJ, Eijkemans MJ, van den Berg LH. Accelerometry for remote monitoring of physical activity in amyotrophic lateral sclerosis: a longitudinal cohort study. J Neurol 2019 Oct;266(10):2387-2395 [FREE Full text] [doi: 10.1007/s00415-019-09427-5] [Medline: 31187191]
5. Rutkove SB, Qi K, Shelton K, Liss J, Berisha V, Shefner JM. ALS longitudinal studies with frequent data collection at home: study design and baseline data. Amyotroph Lateral Scler Frontotemporal Degener 2019 Feb;20(1-2):61-67 [FREE Full text] [doi: 10.1080/21678421.2018.1541095] [Medline: 30486680]
6. Kothare PA, Jadhav PR, Gupta P, Harrelson JC, Dickmann L. Harnessing the potential of emerging digital health and biological sampling technologies for clinical drug development: promise to reality. Clin Pharmacol Ther 2018 Dec;104(6):1125-1135. [doi: 10.1002/cpt.1100] [Medline: 29696625]
7. Hobson EV, Baird WO, Partridge R, Cooper CL, Mawson S, Quinn A, et al. The TiM system: developing a novel telehealth service to improve access to specialist care in motor neurone disease using user-centered design. Amyotroph Lateral Scler Frontotemporal Degener 2018 Aug;19(5-6):351-361. [doi: 10.1080/21678421.2018.1440408] [Medline: 29451026]
8. Bloem BR, Dorsey ER, Okun MS. The coronavirus disease 2019 crisis as catalyst for telemedicine for chronic neurological disorders. JAMA Neurol 2020 Aug 1;77(8):927-928. [doi: 10.1001/jamaneurol.2020.1452] [Medline: 32329796]
9. Andrews JA, Berry JD, Baloh RH, Carberry N, Cudkowicz ME, Dedi B, et al. Amyotrophic lateral sclerosis care and research in the United States during the COVID-19 pandemic: challenges and opportunities. Muscle Nerve 2020 Aug;62(2):182-186 [FREE Full text] [doi: 10.1002/mus.26989] [Medline: 32445195]
10. Hobson EV, Baird WO, Cooper CL, Mawson S, Shaw PJ, Mcdermott CJ. Using technology to improve access to specialist care in amyotrophic lateral sclerosis: a systematic review. Amyotroph Lateral Scler Frontotemporal Degener 2016;17(5-6):313-324. [doi: 10.3109/21678421.2016.1165255] [Medline: 27027466]
11. Gold M, Amatniek J, Carrillo MC, Cedarbaum JM, Hendrix JA, Miller BB, et al. Digital technologies as biomarkers, clinical outcomes assessment, and recruitment tools in Alzheimer's disease clinical trials. Alzheimers Dement (N Y) 2018;4:234-242 [FREE Full text] [doi: 10.1016/j.trci.2018.04.003] [Medline: 29955666]
12. Perry B, Herrington W, Goldsack JC, Grandinetti CA, Vasisht KP, Landray MJ, et al. Use of mobile devices to measure outcomes in clinical research, 2010-2016: a systematic literature review. Digit Biomark 2018;2(1):11-30 [FREE Full text] [doi: 10.1159/000486347] [Medline: 29938250]
13. Hobson EV, Baird WO, Bradburn M, Cooper C, Mawson S, Quinn A, et al. Using telehealth in motor neuron disease to increase access to specialist multidisciplinary care: a UK-based pilot and feasibility study. BMJ Open 2019 Oct 22;9(10):e028525 [FREE Full text] [doi: 10.1136/bmjopen-2018-028525] [Medline: 31640993]
14. Helleman J, van Eenennaam R, Kruitwagen ET, Kruithof WJ, Slappendel MJ, van den Berg LH, et al. Telehealth as part of specialized ALS care: feasibility and user experiences with 'ALS home-monitoring and coaching'. Amyotroph Lateral Scler Frontotemporal Degener 2020 May;21(3-4):183-192. [doi: 10.1080/21678421.2020.1718712] [Medline: 32003245]
15. Garcia-Gancedo L, Kelly ML, Lavrov A, Parr J, Hart R, Marsden R, et al. Objectively monitoring amyotrophic lateral sclerosis patient symptoms during clinical trials with sensors: observational study. JMIR Mhealth Uhealth 2019 Dec 20;7(12):e13433 [FREE Full text] [doi: 10.2196/13433] [Medline: 31859676]
16. Stephens HE, Young J, Felgoise SH, Simmons Z. A qualitative study of multidisciplinary ALS clinic use in the United States. Amyotroph Lateral Scler Frontotemporal Degener 2015;17(1-2):55-61. [doi: 10.3109/21678421.2015.1069851] [Medline: 26508132]
17. Rooney J, Byrne S, Heverin M, Tobin K, Dick A, Donaghy C, et al. A multidisciplinary clinic approach improves survival in ALS: a comparative study of ALS in Ireland and Northern Ireland. J Neurol Neurosurg Psychiatry 2015 May;86(5):496-501. [doi: 10.1136/jnnp-2014-309601] [Medline: 25550416]
18. Hogden A, Foley G, Henderson RD, James N, Aoun SM. Amyotrophic lateral sclerosis: improving care with a multidisciplinary approach. J Multidiscip Healthc 2017;10:205-215 [FREE Full text] [doi: 10.2147/JMDH.S134992] [Medline: 28579792]
19. Cudkowicz ME, van den Berg LH, Shefner JM, Mitsumoto H, Mora JS, Ludolph A, EMPOWER investigators. Dexpramipexole versus placebo for patients with amyotrophic lateral sclerosis (EMPOWER): a randomised, double-blind, phase 3 trial. Lancet Neurol 2013 Nov;12(11):1059-1067. [doi: 10.1016/S1474-4422(13)70221-7] [Medline: 24067398]
20. Chiò A, Canosa A, Gallo S, Cammarosano S, Moglia C, Fuda G, PARALS group. ALS clinical trials: do enrolled patients accurately represent the ALS population? Neurology 2011 Oct 11;77(15):1432-1437. [doi: 10.1212/WNL.0b013e318232ab9b] [Medline: 21956723]
21. van Eijk RP, Westeneng H, Nikolakopoulos S, Verhagen IE, van Es MA, Eijkemans MJ, et al. Refining eligibility criteria for amyotrophic lateral sclerosis clinical trials. Neurology 2019 Jan 9:- epub ahead of print [FREE Full text] [doi: 10.1212/WNL.0000000000006855] [Medline: 30626653]
22. Hobson E, Baird W, Bradburn M, Cooper C, Mawson S, Quinn A, et al. Process evaluation and exploration of telehealth in motor neuron disease in a UK specialist centre. BMJ Open 2019 Oct 22;9(10):e028526 [FREE Full text] [doi: 10.1136/bmjopen-2018-028526] [Medline: 31640994]
23. Galvin M, Carney S, Corr B, Mays I, Pender N, Hardiman O. Needs of informal caregivers across the caregiving course in amyotrophic lateral sclerosis: a qualitative analysis. BMJ Open 2018 Jan 27;8(1):e018721 [FREE Full text] [doi: 10.1136/bmjopen-2017-018721] [Medline: 29374665]

24. Steinhubl SR, Waalen J, Edwards AM, Ariniello LM, Mehta RR, Ebner GS, et al. Effect of a home-based wearable continuous ECG monitoring patch on detection of undiagnosed atrial fibrillation: the MSTOPS randomized clinical trial. J Am Med Assoc 2018 Jul 10;320(2):146-155 [FREE Full text] [doi: 10.1001/jama.2018.8102] [Medline: 29998336]
25. de Vito Dabbs A, Myers BA, Mc Curry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
26. Espay AJ, Hausdorff JM, Sánchez-Ferro A, Klucken J, Merola A, Bonato P, Movement Disorder Society Task Force on Technology. A roadmap for implementation of patient-centered digital outcome measures in Parkinson's disease obtained using mobile health technologies. Mov Disord 2019 May;34(5):657-663 [FREE Full text] [doi: 10.1002/mds.27671] [Medline: 30901495]
27. Moser A, Korstjens I. Series: practical guidance to qualitative research. Part 3: sampling, data collection and analysis. Eur J Gen Pract 2018 Dec;24(1):9-18 [FREE Full text] [doi: 10.1080/13814788.2017.1375091] [Medline: 29199486]
28. Roche JC, Rojas-Garcia R, Scott KM, Scotton W, Ellis CE, Burman R, et al. A proposed staging system for amyotrophic lateral sclerosis. Brain 2012 Mar;135(Pt 3):847-852 [FREE Full text] [doi: 10.1093/brain/awr351] [Medline: 22271664]
29. Chiò A, Hammond ER, Mora G, Bonito V, Filippini G. Development and evaluation of a clinical staging system for amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry 2015 Jan;86(1):38-44. [doi: 10.1136/jnnp-2013-306589] [Medline: 24336810]
30. Westeneng H, Debray TP, Visser AE, van Eijk RP, Rooney JP, Calvo A, et al. Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model. Lancet Neurol 2018 May;17(5):423-433. [doi: 10.1016/S1474-4422(18)30089-9] [Medline: 29598923]
31. Strong MJ, Abrahams S, Goldstein LH, Woolley S, Mclaughlin P, Snowden J, et al. Amyotrophic lateral sclerosis - frontotemporal spectrum disorder (ALS-FTSD): revised diagnostic criteria. Amyotroph Lateral Scler Frontotemporal Degener 2017 May;18(3-4):153-174 [FREE Full text] [doi: 10.1080/21678421.2016.1267768] [Medline: 28054827]
32. Crockford C, Newton J, Lonergan K, Chiwera T, Booth T, Chandran S, et al. ALS-specific cognitive and behavior changes associated with advancing disease stage in ALS. Neurology 2018 Oct 9;91(15):e1370-e1380 [FREE Full text] [doi: 10.1212/WNL.0000000000006317] [Medline: 30209236]
33. Chiaramonte R, Bonfiglio M. Acoustic analysis of voice in bulbar amyotrophic lateral sclerosis: a systematic review and meta-analysis of studies. Logoped Phoniatr Vocol 2020 Dec;45(4):151-163. [doi: 10.1080/14015439.2019.1687748] [Medline: 31760837]
34. Ard MC, Edland SD. Power calculations for clinical trials in Alzheimer's disease. J Alzheimers Dis 2011;26(Suppl 3):369-377 [FREE Full text] [doi: 10.3233/JAD-2011-0062] [Medline: 21971476]
35. Bakker LA, Schröder CD, Tan HH, Vugts SM, van Eijk RP, van Es MA, et al. Development and assessment of the inter-rater and intra-rater reproducibility of a self-administration version of the ALSFRS-R. J Neurol Neurosurg Psychiatry 2020 Jan;91(1):75-81. [doi: 10.1136/jnnp-2019-321138] [Medline: 31558653]
36. Rutkove SB, Narayanaswami P, Berisha V, Liss J, Hahn S, Shelton K, et al. Improved ALS clinical trials through frequent at-home self-assessment: a proof of concept study. Ann Clin Transl Neurol 2020 Jul;7(7):1148-1157 [FREE Full text] [doi: 10.1002/acn3.51096] [Medline: 32515889]
37. van Eijk RP, Kliest T, McDermott CJ, Roes KC, Van Damme P, Chio A, et al. TRICALS: creating a highway toward a cure. Amyotroph Lateral Scler Frontotemporal Degener 2020 Nov;21(7-8):496-501. [doi: 10.1080/21678421.2020.1788092] [Medline: 32643415]
38. Alami H, Gagnon M, Fortin J. Some multidimensional unintended consequences of telehealth utilization: a multi-project evaluation synthesis. Int J Health Policy Manag 2019 Jun 1;8(6):337-352 [FREE Full text] [doi: 10.15171/ijhpm.2019.12] [Medline: 31256566]
39. van Eijk RP, Nikolakopoulos S, Roes KC, Middelkoop BM, Ferguson TA, Shaw PJ, et al. Critical design considerations for time-to-event endpoints in amyotrophic lateral sclerosis clinical trials. J Neurol Neurosurg Psychiatry 2019 Dec;90(12):1331-1337 [FREE Full text] [doi: 10.1136/jnnp-2019-320998] [Medline: 31292200]
40. Pinto A, Almeida JP, Pinto S, Pereira J, Oliveira AG, de Carvalho M. Home telemonitoring of non-invasive ventilation decreases healthcare utilisation in a prospective controlled trial of patients with amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry 2010 Nov;81(11):1238-1242. [doi: 10.1136/jnnp.2010.206680] [Medline: 20826878]
41. de Almeida JP, Pinto A, Pinto S, Ohana B, de Carvalho M. Economic cost of home-telemonitoring care for BiPAP-assisted ALS individuals. Amyotroph Lateral Scler 2012 Oct;13(6):533-537. [doi: 10.3109/17482968.2012.703675] [Medline: 22873565]
42. Levenson MS. Regulatory-grade clinical trial design using real-world data. Clin Trials 2020 Aug;17(4):377-382. [doi: 10.1177/1740774520905576] [Medline: 32063037]
43. Rogers JR, Lee J, Zhou Z, Cheung YK, Hripcsak G, Weng C. Contemporary use of real-world data for clinical trial conduct in the United States: a scoping review. J Am Med Inform Assoc 2021 Jan 15;28(1):144-154. [doi: 10.1093/jamia/ocaa224] [Medline: 33164065]

Abbreviations

ALS: amyotrophic lateral sclerosis
ALSFRS-R: Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised
MND: motor neuron disease
CE: Conformité Européenne

Edited by R Kukafka; submitted 14.03.21; peer-reviewed by M De Carvalho, Y Zhang; comments to author 10.05.21; revised version received 23.05.21; accepted 24.05.21; published 22.09.21.

Please cite as:
van Eijk RPA, Beelen A, Kruitwagen ET, Murray D, Radakovic R, Hobson E, Knox L, Helleman J, Burke T, Rubio Pérez MÁ, Reviers E, Genge A, Steyn FJ, Ngo S, Eaglesham J, Roes KCB, van den Berg LH, Hardiman O, McDermott CJ
A Road Map for Remote Digital Health Technology for Motor Neuron Disease
J Med Internet Res 2021;23(9):e28766
URL: https://www.jmir.org/2021/9/e28766
doi:10.2196/28766
PMID:34550089

Original Paper

Developing an Asthma Self-management Intervention Through a Web-Based Design Workshop for People With Limited Health Literacy: User-Centered Design Approach

Hani Salim[1,2*], MMed; Ping Yein Lee[3*], MMed; Sazlina Sharif-Ghazali[2,4*], PhD; Ai Theng Cheong[2*], PhD; Jasmine Wong[2*], MBBS; Ingrid Young[5*], PhD; Hilary Pinnock[1*], MD; RESPIRE Collaboration[1]

[1]NIHR Global Health Research Unit on Respiratory Health (RESPIRE), Usher Institute, The University of Edinburgh, Edinburgh, United Kingdom

[2]Department of Family Medicine, Medical Faculty and Health Sciences, Universiti Putra Malaysia, Serdang, Malaysia

[3]UM eHealth Unit, Faculty of Medicine, University Malaya, Petaling Jaya, Malaysia

[4]Malaysian Research Institute on Ageing, Universiti Putra Malaysia, Serdang, Malaysia

[5]Centre for Biomedicine, Self and Society, Usher Institute, University of Edinburgh, Edinburgh, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Hilary Pinnock, MD
NIHR Global Health Research Unit on Respiratory Health (RESPIRE)
Usher Institute
The University of Edinburgh
Old Medical School
Doorway 3, Teviot Place
Edinburgh, EH8 9AG
United Kingdom
Phone: 44 0131 650 94
Email: hilary.pinnock@ed.ac.uk

Abstract

Background: Technology, including mobile apps, has the potential to support self-management of long-term conditions and can be tailored to enhance adoption. We developed an app to support asthma self-management among people with limited health literacy in a web-based workshop (to ensure physical distancing during the COVID-19 pandemic).

Objective: The aim of this study is to develop and test a prototype asthma self-management mobile app tailored to the needs of people with limited health literacy through a web-based workshop.

Methods: We recruited participants from a primary care center in Malaysia. We adapted a design sprint methodology to a web-based workshop in five stages over 1 week. Patients with asthma and limited health literacy provided insights into real-life self-management issues in stage 1, which informed mobile app development in stages 2-4. We recruited additional patients to test the prototype in stage 5 using a qualitative research design. Participants gave feedback through a concurrent thinking-aloud process moderated by a researcher. Each interview lasted approximately 1 hour. Screen recordings of app browsing activities were performed. Interviews were audio-recorded and analyzed using a thematic approach to identify utility and usability issues.

Results: The stakeholder discussion identified four themes: individual, family, friends, and society and system levels. Five patients tested the prototype. Participants described 4 ways in which the app influenced or supported self-management (utility): offering information, providing access to an asthma action plan, motivating control of asthma through support for medication adherence, and supporting behavior change through a reward system. Specific usability issues addressed navigation, comprehension, and layout.

Conclusions: This study proved that it was possible to adapt the design sprint workshop to a web-based format with the added advantage that it allowed the development and the testing process to be done efficiently through various programs. The resultant app incorporated advice from stakeholders, including sources for information about asthma, medication and appointment reminders, accessible asthma action plans, and sources for social support. The app is now ready to move to feasibility testing.

(J Med Internet Res 2021;23(9):e26434) doi:10.2196/26434

KEYWORDS

asthma; self-management; design sprint; health literacy; mobile phone

Introduction

Background

Supported self-management for asthma (written action plans and regular review) is highly effective at improving control and reducing acute attacks [1-3]; however, globally, it is challenging to implement for 334 million people living with asthma [4-7]. One of the challenges is the need to tailor support for people with limited health literacy. Health literacy is defined as the degree to which individuals can obtain, process, and understand the necessary health information needed to make appropriate health decisions [8]. Studies have associated limited health literacy with erroneous health beliefs and poor adherence to self-management activities [9,10]. Malaysia has a high burden of limited health literacy in the general population [11], and asthma control is challenged by a lack of patient education, overreliance on unscheduled visits, and lack of action plan ownership [12-14].

The use of digital technologies for internet-based information is more common in the younger age group than in the middle and older age groups [15]. Malaysia's multigenerational household culture and strong family orientation have helped younger family members assist older generations in using digital technology to stay connected and find information [16]. The pandemic has further seen widespread adoption of digital technologies by a broader age group of users in diagnosis, prevention, and surveillance [17]. Three-fourth of Malaysians are now smartphone users, with most (60.9%) of them in the lowest income group [18]. Although health-related information-seeking behavior on the internet is greater in those with good health literacy [19], our previous qualitative work among people with limited health literacy suggests that a mobile app is a preferred medium to deliver supported self-management, including a pictogram-based asthma action plan and signposting to reliable asthma information sources.

Goals

For the aforementioned reasons, developing asthma self-management tailored to limited health literacy needs is an important context for the web-based design sprint workshop, as various studies have shown that the extensive use of pictograms, images, and prompts was appealing to participants and may improve the understanding of information in mobile apps [20-23]. However, it is essential to involve users early in the design stage, as some of the unique features that people want can be time-consuming and costly to build, and a balance may need to be found between desired features [24] and those with evidence-based recommendations [25,26].

Using a design sprint methodology, we seek to optimize user experience in app development by integrating patients into the 5-stage process of mapping, sketching, designing, developing, and testing [24,27]. We used the health literacy framework [8] to underpin the overall structure of the interventional work. The COVID-19 pandemic and physical distancing requirements meant that we had to conduct our workshop in a web-based format. In this study, we report the outcomes of the workshop deliberations and our experience of conducting a week-long remote 5-stage program attended by patients with asthma and health care professionals (HCPs).

Methods

Ethical Consideration

The workshop received ethical approval from the Medical Research and Ethics Committee of the Ministry of Health, Malaysia (ID: NMRR-19-3609-52292) and sponsorship approval by the Academic and Clinical Central Office for Research & Development at the University of Edinburgh (ID: AC20011). Informed consent was obtained from all the participants before the workshop.

Study Design

We conducted a 5-stage design sprint workshop using a web-based and qualitative research approach. We used the 5-stage design sprint process as a roadmap to develop the intervention. We adapted the methodology and constructed the workshop into five stages: (1) understanding and mapping problems, (2) sketching of solutions, (3) deciding on solutions to problems, (4) developing a prototype, and (5) testing a low-fidelity prototype [27,28]. The process was originally designed in the technology sector by the Google Ventures team for business start-up teams [27]. The involvement of the target population and early testing enhance intervention effectiveness and increase the likelihood of adoption at the implementation stage [24]. Owing to the rapid development and testing stages, this is an ideal concept for a low-resource setting, that is, time and cost [24].

Stakeholder (patient and HCP) discussions provided insights into self-management issues in stage 1, which informed mobile app development in stages 2-4. We recruited patients with asthma and limited health literacy to test the low-fidelity prototype in stage 5 and provided feedback through qualitative interviews. A low-fidelity prototype is a modeled prototype with limited technical functionality [29] that is quick to create and can be easily improved in the light of feedback in the testing stages. The details of each stage are listed in Table 1.

Table 1. Process, outcomes, and web-based adaptations of the workshop.

Stages and objectives	Process	Adaptations for web-based delivery of the workshop
Stage 1: understand and mapping problem		
• To identify the objectives of the prototype and the workshop • To map out problems from health care professionals and patients' perspectives which technology can help to solve	• As a team, we first discussed and agreed on the workshop's long-term goals for the workshop's prototype and aims through a structured discussion between the patients and health care professionals. • We listed a list of problems relating to self-managing asthma from stakeholders (patients and clinicians) point of views. • We constructed an end-to-end process of how patients cared for their asthma, and we targeted the problems we could potentially provide solutions for using the mobile app.	• Through a web conference site (Microsoft Teams), we brainstormed the long-term goal for the app and the workshop's aim. • In a separate browser, using a web-based board (Miro), we gathered the problems, potential solutions, and mapping of the target where the solutions can occur. Interviews were audio recorded during the workshop. • We considered scientific literature and previous study we have conducted as expert input.
Stage 2: sketch solution		
• To understand a broad range of problems and solutions concerning asthma self-management	• Focusing on the problems, each researcher reviewed existing ideas which we could potentially use and improve for the prototype. • The individual researcher then presented their findings and the reasons why the ideas being chosen. • Using this information, we then drew crude scenes with our contents which we believed would be suitable for the app. • After presenting the scenes and critical discussions, we voted on the best scene and content for the prototype.	• Reviewing and compiling sketches of ideas were done on Miro synchronously by all researchers. • We presented these sketches of ideas to the whole team on Microsoft Teams. • Individually, using colorful sticky notes and marker pens, we drew the crude scenes. We took photographs of these scenes and uploaded these on Miro. • Each researcher was given three blue dots for the voting, and they placed a dot on the best ideas.
Stage 3: decide the solution for the problem		
• To decide on solutions that answered our long-term objectives	• The winning scenes and content comprised topics on asthma education, asthma symptoms monitoring, and supporting people living with asthma. • We took the winning scenes from our sketches, and we constructed an end-to-end process (storyboard) on how these scenes and content would appear on an App. • The storyboard was first constructed in text form before we transformed it visually.	• The most voted ideas were put together, and we had another round of voting where each researcher was given a pink dot, and the team leader was given three purple dots on the web-based board (Miro). • The text and visual version of the storyboard was constructed as a group and through discussions on the web-based board (Miro) and Microsoft Teams.
Stage 4: prototype development		
• To build ideas for a low-fidelity prototype	• Each of the research members and the App developer was assigned roles to ensure the successful development of the low-fidelity prototype for the final day testing process.	• On Figma, a prototype development site, the low-fidelity prototype was developed. • Every researcher and the App developer completed their tasks (ie, content and language check) within Figma. • Brainstorming of prompts for the testing day was conducted on Google Sheets among the researchers. • Google sheet also was where a virtual scoreboard was set up for every researcher to capture the patient's evaluation of the prototype.
Stage 5: test low-fidelity prototype		

Stages and objectives	Process	Adaptations for web-based delivery of the workshop
• To validate the solutions for the patients through a qualitative method	• We tested the solutions to 5 patients using a concurrent think-aloud process. • We gathered verbal and visual feedback about the low-fidelity prototype from the patients, which we will use to build a high-fidelity prototype.	• Interviews were conducted through Microsoft Teams by a moderator, HS, with a patient observed by other researchers (whose video and audio function was turned off). • HS and each patient could see each other for an ice-breaking session at the start of the testing session. This session was essential to create rapport and to ensure the patient's readiness, mentally and technically. • The patient was then given a link to the prototype where they browsed through the prototype, gave comments and answered prompted questions. The patient's screen was shared within the MS Team. • The observers synchronously collected the patients' replies on the utility and usability prompts about the prototype during the interview on the virtual scoreboard and the observer's field notes. • The moderator had accessed to the scoreboard and would be able to pick up any point which needed further clarification. • The interviews were audio recorded, and browsing activities were video recorded.

The web-based workshop discussions were conducted on a web conference platform (Microsoft Teams), and brainstorming of the idea was conducted on a web-based board (Miro). While conducting the exercise on the web-based board, the workshop participants remained connected to the web conference site to allow ongoing discussion. The web-based board was superseded by a prototype development (Figma) site in stages 4 and 5, whereas discussions remained on Microsoft Teams. The app developer supported information technology activities. Two weeks before the workshop, one of the researchers contacted each of the participants (patients and HCPs) to assess technical skills such as the ability to log on, use a meeting platform, and logistic issues such as the quality of the internet connection.

Setting

The workshop, which took place between June 22 and 26, 2020, was conducted through a secure web-based meeting platform using a virtual whiteboard to facilitate information sharing between the researchers and app developers. Our original plan for a face-to-face workshop was changed to a web-based format to overcome the restriction of the order of movement because of the COVID-19 pandemic; an additional advantage was that it allowed participants from different locations and time zones to participate.

The patients were from 2 urban public primary care clinics in central Malaysia. Asthma is managed in primary care clinics, although chronic and acute care management through the provision of an asthma action plan to support asthma self-management is uncommon. Malaysia has a dual health system, public and private, where the public health system provides the leading service for the population with copayment of Ringgit Malaysia, RM 1 (US $0.23) per visit.

Samples and Recruitment

Stakeholder Discussion (Stage 1)

A total of 3 patients and 2 HCPs who cared for asthma in the primary care settings, 2 app developers, 4 researchers from Universiti Putra Malaysia, and 1 from the University of Edinburgh were involved in the stakeholder discussion in stage 1 (Table 2).

Table 2. Summary of the stakeholders involved in each stage.

Stakeholder	Stage				
	1	2	3	4	5
1. Researchers	✓[a]	✓	✓	✓	✓
2. Patients	✓				✓
3. Health care professionals	✓				
4. App developers	✓	✓	✓	✓	✓

[a]Stakeholder present.

Testing of the Prototype (Stage 5)

Five patients, recruited from the Klang Asthma Cohort, participated in testing the prototype at stage 5 (Table 2). The Klang Asthma Cohort database is one of the research outputs of RESPIRE (National Institute for Health Research Global Health Research Unit on Respiratory Health) in Malaysia. The database contains 1280 people with asthma recruited from

primary health care clinics in the Klang district. They provided consent to be called with invitations to participate in asthma-related research. Inclusion criteria for patients invited for this study were physician-diagnosed asthma, aged >18 years, smartphone user, limited health literacy, and were assessed by screening using the *Bahasa Malaysia* language version of the Health Literacy-Q47 scale [30]. The initial Health Literacy-Q47 scale [31] was translated and validated in Malaysia with a Cronbach α=.96 [30].

Data Collection

Demographic information was collected from the database of patients who agreed to participate. Stakeholders' discussions in stage 1 focused on challenges by patients and HCPs around (1) asthma education, (2) asthma self-management, (3) monitoring of symptoms, (4) emotional support or lifestyle advice, (5) social support, and (6) clinic set up (Multimedia Appendix 1).

In stage 5, we tested the prototype with 5 patients to assess its utility and usability using a set of semistructured questions in a concurrent think-aloud manner (Multimedia Appendix 2). During this session, HS, the main interviewer, guided the process. Four other researchers, PYL, SSG, ATC, and JW, observed the interview while the 2 app developers, Aidil Goh and Muhammad Marzuqi, managed the technical aspects of the sessions. The interviews took 1 hour and were conducted in *Bahasa Malaysia*, the patients' preferred language.

Qualitative interviews were audio recorded, other web-based discussions and browsing activities were video recorded, and web-based board exercises were captured and archived as described in Table 1. All interviews were transcribed verbatim.

Data Analysis

For this qualitative study, we used thematic analysis to obtain rich data from the stakeholders' discussions in stage 1 and the interviews in stage 5. The texts were analyzed iteratively using a deductive thematic analysis approach, as outlined by Braun et al [32]. The deductive thematic analysis seeks to answer the researcher's theory or analytical interest within the topic [32]. Phases in the thematic analysis included [32] (1) familiarization with the data by reading and rereading and noting down initial ideas (memoing), (2) duplicate coding (HS and JW) of one interview and comparing decisions to agree on standardizing the coding framework before coding all the transcripts, (3) discussing emerging themes with the research team, (4) reviewing themes with the wider research team and generating a map of the analysis (HP, SSG, PYL, ATC, or IY), (5) defining themes iteratively, and (6) presenting the deductive analysis with a selection of extracts. The data were organized using NVivo 11 (QSR International) qualitative data analysis software (HP and IY).

Results

Participants' Characteristics

Five patients (including 3 who attended stage 1) attended stage 5 (prototype testing). Table 3 summarizes the demographics of the patients involved in stage 5.

Table 3. Patients' demographic involved in stage 5.

ID	Age (years)	Gender	Education level	Health literacy score[a]	Use of pictorial asthma action plan at 6 months	Access to a digital device
P1	44	Female	Secondary	30	Yes	Smartphone and PC
P2	36	Male	Tertiary	32	No	Smartphone
P3	40	Female	Tertiary	17	Yes	Smartphone and PC
P4	38	Male	Tertiary	21	No	Smartphone
P5	19	Male	Secondary	31	No	Smartphone and PC

[a]Score less than 33 is considered as limited health literacy.

The Outcomes of Each Stage

Stage 1: Understanding and Mapping Problems

Overall, 3 patients and 2 HCPs (a family physician and a medical officer from Klang district) contributed to the stakeholder's discussion. The stakeholder discussion themes were categorized as relating to individuals, family and friends, society, and systems (Multimedia Appendix 3). The key problem used to inform the app design was education sources for asthma and support in the community, enabling self-management using pictorial action plans, reminders for medications, and asthma reviews or appointments. The log of history of asthma control, preventer intake, and information on expected best peak expiratory flow rate were features that could support patients during asthma review or appointment to discuss with their HCPs.

Stage 2-4: Sketching, Designing, and Developing the Prototype

Informed by the stakeholder discussion and findings of our previous qualitative study, we worked through the stages of sketching solutions and designed and developed the prototype (Multimedia Appendix 4). Through a round of voting in stage 3, solutions supporting self-management based on evidence-based practice were selected [3]. All winning solutions were clustered around four aspects of care: (1) education, (2) supporting self-management, (3) supporting behavior change, and (4) social support (Table 4). We created a storyboard for the prototype, and we developed the prototype app based on the storyboard, which was then tested in stage 5.

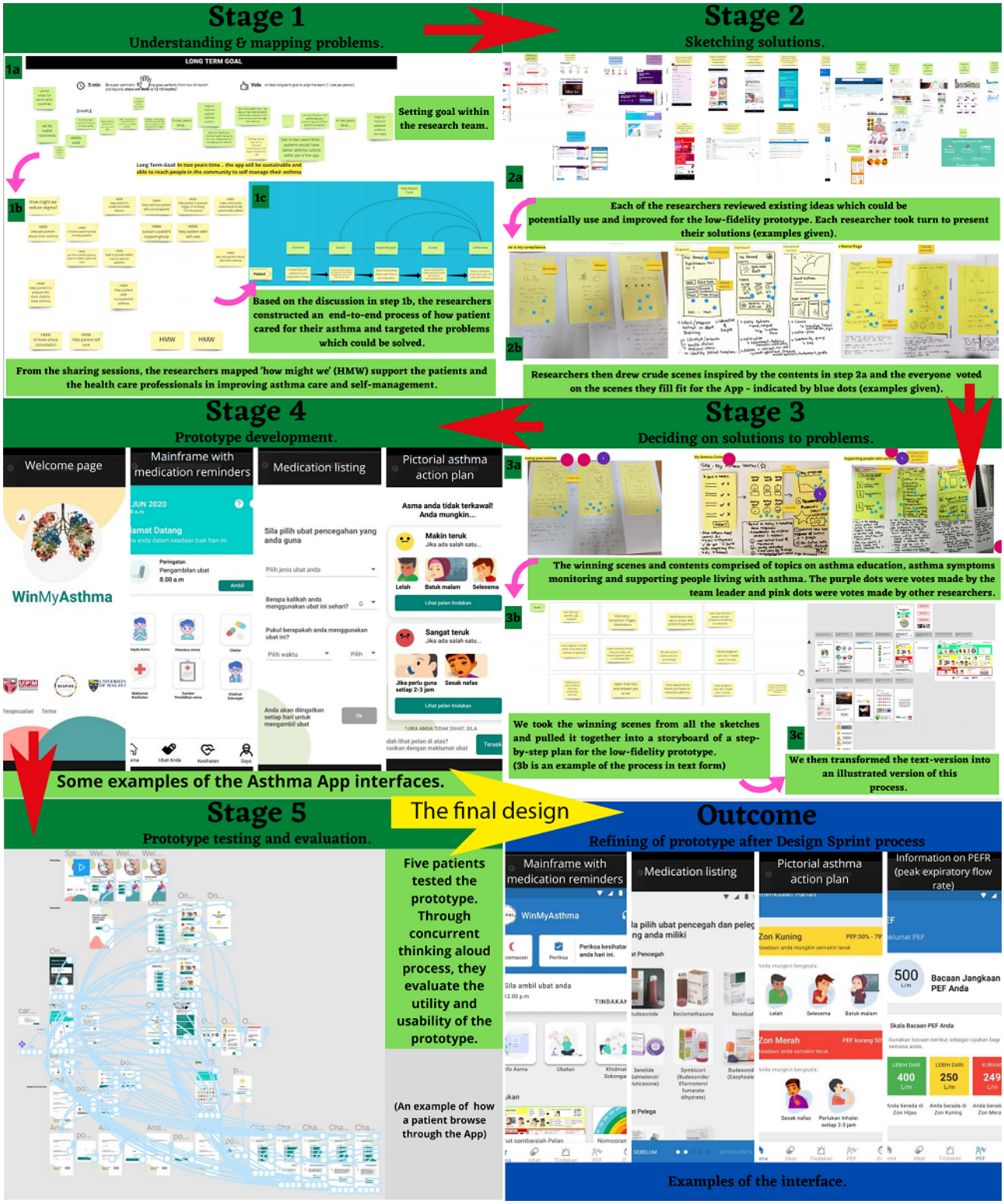
Table 4. Asthma App content and design features.

Main theme and section	Content	Features
Education		
About asthma, its symptoms and diagnosis; asthma medications	• Information in text and videos about asthma, symptoms, triggers, how the diagnosis is made, exacerbations and myths around asthma attack. Other information includes types of medications used to treat asthma, its function, and potential side-effects of the medications. There will be video-based instructions on the inhaler technique.	• The links to the Ministry of Health portal on asthma was provided under specific headings to facilitate the search for reliable information.
Supporting self-management		
Self-monitoring of symptoms	• Patients indicate any experience of asthma symptoms in the last 24 hours, which will translate into control and prompts to check the action plan.	• Tick-box list of potential asthma symptoms; ticking any one symptom will prompt a pop-up on advice to look at an action plan with a click button.
Asthma action plan	• A pictorial asthma action plan was used. Illustrations and wordings were validated in a series of discussions with stakeholders.	• List of zones are displayed, and patient choose which zone are appropriate for them.
Supporting behavior change		
Asthma medication and appointment reminder	• Patients provide information about medications and appointments which will trigger a reminder system at the timing of choice.	• Matrix of images of medications used and drop-down menu for frequency and timing
Asthma diary	• Asthma control and medication uptake will be recorded in the diary, including best PEFR[a].	• Monthly calendar, which displays asthma control and adherence
Reward system	• Achieving good asthma control and medication adherence will be translated into points.	• Display of scale of points achieved for good asthma control and adherence
Others		
Social support	• Information regarding support groups for asthma in Malaysia	• The links to various support groups available in Malaysia

[a]PEFR: peak expiratory flow rate.

The entire process is illustrated in Figure 1, and the decisions on the prototype content are summarized in Table 4. The app was written in the Malay language.

Figure 1. The Design Sprint process was undertaken on the web-based discussion board and meeting platform.



Stage 5: Testing Low-Fidelity Prototype

Patients who attended stage 1, along with 2 other patients, attended stage 5. Results are presented under the 2 main themes of utility (ie, app influence or support self-management) and usability (ease of use).

Utility

Patients commented that the app influenced the decision to self-manage in four ways: offering information, providing an accessible asthma action plan, motivating and supporting improved medication adherence, and promoting behavior change through a reward system.

Offering Information

Patients considered that the app provided essential information regarding asthma and how to manage it. P4 explained, “the information about asthma in the App is interesting and informative.” Although many were comfortable reading

text-based information, some preferred audio-visual formats such as videos. P5 explained, "the information about asthma, maybe it can be in the video, it's more interesting than just text."

Providing Accessible Asthma Action Plans

Patients felt that having an action plan on the phone made the plan accessible when needed. As P1 described, "when you need the plan, you just open the App in your phone and click on the plan [action], to see it." In this format, patients considered it easier to access and use the app-based plan than the paper-based action plan, which they may not carry unless they attend medical appointments.

Motivating and Supporting Improved Medication Adherence

The medication reminder function of the app was viewed as a good support for achieving good adherence to daily preventers. P5 stated, "It will be difficult not to remember taking the medications because of the reminder, and because I use the phone frequently, it is hard to ignore the reminder (chuckled)."

Promoting Behavior Change Through a Reward System

The app was designed to encourage behavior change through a positive reward system, where good asthma control and adherence to twice daily preventive inhalers would be awarded points, and the cumulative points were visualized clearly. Patients liked this approach. P1 was incredibly excited to see the reward points on the app: "wow, there is a reward points, this is great!" This excitement was shared with other patients who preferred to see the tangible results of their actions. P3 elaborated on how the reward system could influence behavior:

> *That's nice when I get points for taking the medications. I do want to see that I accumulate points and the scale moved further. And I can only do this if my control is good and if I take my medications.*

Usability

The patients could easily comprehend the information and instructions in the app. P2 elaborated on how he achieved this: "the instructions are accompanied with illustrations. It makes it easier to understand it." Although patients liked the use of illustrations and fewer words, the small font size used was challenging for some. P1 said:

> *[The] writing is small. I tried to put on my glasses, I still can't see it (chuckled).*

We used cartoon-based illustrations of the medications and in the steps of the asthma action plan. Some studies have suggested that the pictorial asthma action plan may be useful for patients with asthma other than adults. P3 suggested:

> *I think the illustrated plan can be appealing and useful for children and their carer. It's very easy to understand.*

Patients pointed out that a lack of navigational symbols meant that it was not always clear how to move from one interface to another. P4 stated his confusion:

> *There is no sign or indication on what to do next. I was a bit lost on what should I do now. Perhaps an arrow would help to tell that I can move forward.*

Otherwise, patients were mostly satisfied with the simple layout. P1 gave an example of this:

> *In terms of the layout, it's quite easy to navigate around the App. It's ok for me.*

The time spent using the app varied between 10 and 45 minutes. Some (younger) patients seemed comfortable navigating from one interface to the other and were keen to click buttons to explore the app functions. In contrast, others were dependent on symbols or prompted to navigate, which the app lacked at the testing stage.

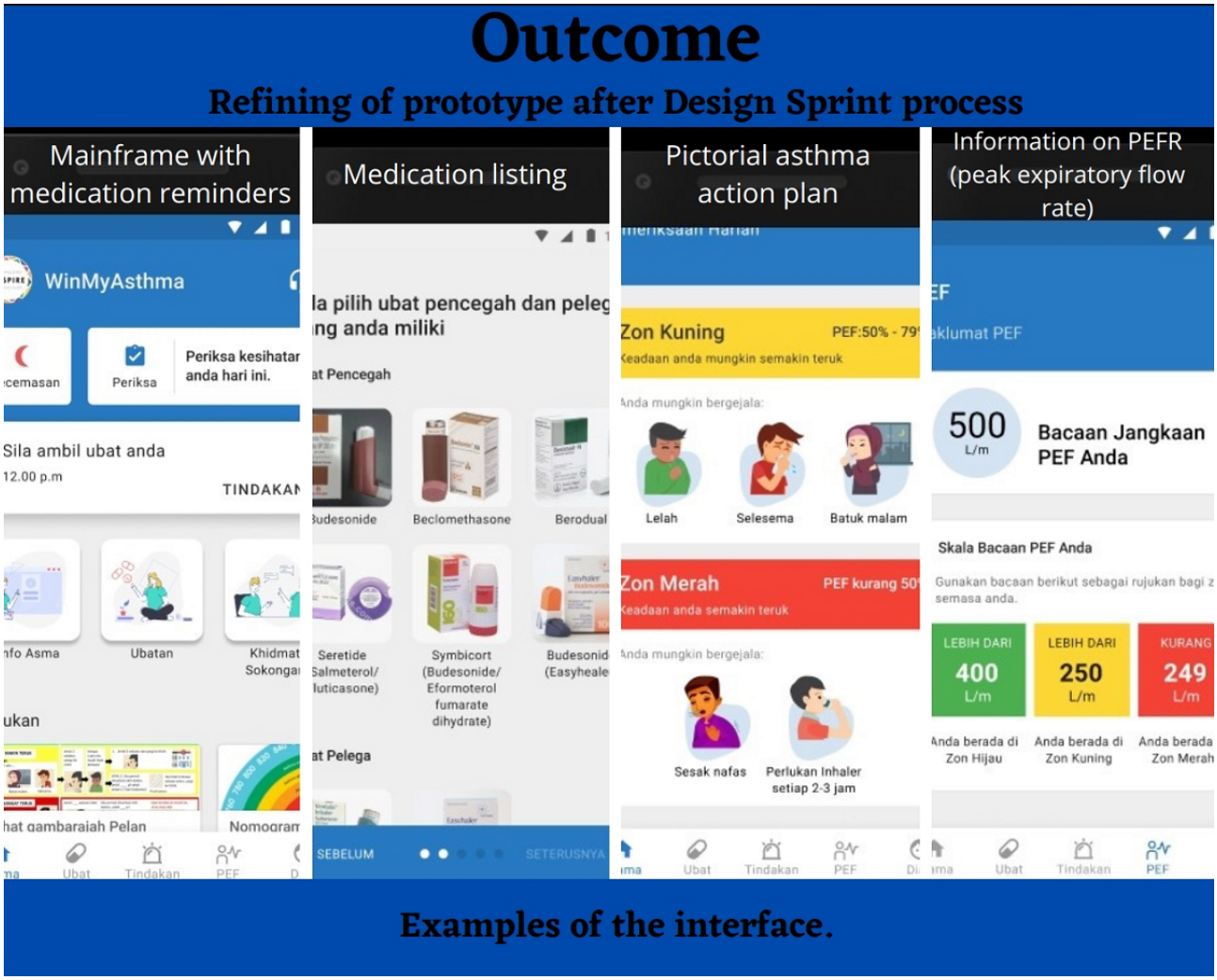
The language used in the app was generally satisfactory, although there was a linguistic misunderstanding of *breathlessness* and *wheeze* in the *Bahasa Malaysia* language. P4 described his confusion on the *Bahasa Malaysia* words for breathlessness and wheeze:

> *Mengalami sesak nafas (breathlessness) and lelah (wheezing), are different? I thought it's the same thing?*

After the Workshop

The design was refined and finalized after testing in stage 5 (Figure 2).

Figure 2. The finalized app design.



Discussion

Principal Findings

We conducted a design sprint workshop and employed a web-based format to ensure the safety of researchers and patients during the global pandemic. In this 5-stage workshop, we developed a low-fidelity prototype based on theoretical frameworks and refined it based on patient feedback during the design and testing stages. Patients described the resultant app to influence their ability to self-manage in four ways: offering information, providing accessible asthma action plans, motivating and supporting improved medication adherence, and promoting behavior change through a reward system. Specific usability issues were related to navigation, comprehension, and layout.

Strengths and Limitations

The involvement of stakeholders and the multidisciplinary approach at the development stage in the design sprint process are among the strengths of this study, which may increase the chances of the intervention meeting the needs of the target population. The 5-stage design sprint structure allowed the development and testing process to be performed quickly and efficiently at a low cost, which would likely be favorable in low-resource settings. Constructing a low-fidelity prototype offers many advantages in the initial stages of prototype development. It allows a quick gathering of requirements, ideas, and concepts and can be built rapidly [29]. The disadvantage was that the low-fidelity prototype lacked some core functionality (such as navigation features), so that the patients on day 5 gave feedback on a limited version of the app. Nevertheless, the feedback was beneficial and enabled the app to be refined after the workshop to produce a high-fidelity prototype.

The web-based approach connected people in different geographical locations and ensured safety during the global pandemic. We recognized that we do not have any participants aged >50 years, perhaps because this approach may be more appealing to younger age groups, although with limited health literacy, and it may not reflect the feasibility of using web-based methods for older age groups. We overcame limited internet access by providing an internet data voucher; however, we had to exclude those with no access to any digital devices. We provided training on the various platforms to be used in the workshop, thus overcoming the lack of digital skills. These strategies may assure researchers or intervention developers working in countries or settings with high levels of limited health literacy that this web-based methodology is of value.

The pandemic context may have explained the relatively small number of participants recruited during the testing stage. Some participants found it challenging to commit uninterrupted time to a web-based workshop while being *locked down* at home with their domestic or caring or *home-schooling* responsibilities. For comparison, face-to-face intervention design workshops have been reported with 14 participants in 5-day workshops [33] and 38 participants over 6 weeks [34].

Adaptation of the Design Sprint Workshop to a Web-Based Format

Patients with asthma and HCPs were recruited for a workshop in March 2020, but this could not occur because of the compulsory lockdown imposed by the local authorities in response to the COVID-19 pandemic. Therefore, we adapted the workshop to the challenges of a web-based format and its potential impact on participants' research experience [35]. A high-speed internet connection and technical skills in conducting this workshop were vital. To overcome some of these practical issues, we supported all participants with mobile web-based access through an RM 10 (US $2.42) internet data voucher to ensure that the participants would not bear the cost of internet access. Participants' experience with technologies was around social media, that is Facebook, video call, that is, WhatsApp, and information searching platforms, that is, Google and YouTube. We found a lack of experience with videoconferencing platforms and the software we planned to use during the testing day. Thus, we conducted training sessions for all participants to avoid technical problems during the workshop.

In the context of interviews, the literature suggests that web-based data collection can produce data of similar quality to face-to-face interviews [35]. Although using a web-based platform to interview patients was a new experience for the researchers, we found that the web-based programs eased discussions, and interviews took place quickly and effectively. Compared with traditional qualitative interviews, one advantage was that other researchers could observe the interview sessions on the web-based platform, and they could make concurrent fieldnotes. From the patient's feedback, although they knew they were being observed by additional researchers, being at their own home helped them forget about being observed and anxious.

In web-based discussions and interviews, dictation software has been used to capture audio data in text format, thereby avoiding transcription errors [36,37]. However, this was not possible in our context because the medium of interaction was the *Bahasa Malaysia* language, and the extensive use of colloquial language made it impossible to use any dictation software. Conducting research on the web raises concerns about participants' confidentiality and data security. We ensured that entry to the workshop was password-protected to control access to maintain the participant's confidentiality securely. The recordings were stored in a secure manner. For example, files from the workshop were encrypted and stored in a secure research data storage facility.

Web-Based Design Workshop and the Context of Limited Health Literacy in a Low-to-Middle–Income Country

Our app focused on designing a pictorial asthma action plan as a core strand of tailoring supported self-management for people with limited health literacy. Other features were a simple language for symptom assessment, education and information resources, provision of visual and audio medication reminders, and practical behavior change strategies such as a reward system. A clear message from our previous qualitative work was that participants wanted an interactive approach to support an asthma action plan with few words and clear pictures. Our original plan was to provide a paper-based pictorial action plan, but as it is not interactive, we reconsidered potential formats and decided to deliver the pictorial asthma action plan using a mobile app. Using a mobile app was seen as promoting a sense of autonomy to feel empowered in managing asthma.

Although previous reviews have reported a lack of interest in action plans [21], others have reported on the keenness of people to use action plans in mobile apps [38], although none have explored plans tailored for people with limited health literacy or the innovative pictorial representation of actions in a mobile app. In the United States, a study found that action plans were written at the literacy level of sixth-grade (11- to 12-year-olds), which will be a challenge to those without formal education or only receiving primary school education [39]. The same study also found that more graphics within an action plan may be needed to increase the ease of use [39]. In our web-based design workshop, the extensive use of images, icons, and the use of simple language were among the strategies used to overcome the challenge of understanding a written asthma action plan.

Recommendations for Practice, Policy, and Research

We outlined recommendations for practice, policy, and research based on this study in Textbox 1. On a practical level, to help researchers' concentration throughout the 5 days of the 7-hour workshops, we included frequent breaks and provided high-energy snacks. Each session was either 1 hour long with a 15-minute break or 45 minutes long with a 10-minute break. Committing to time in a workshop alone in front of the monitor can be challenging and mentally draining, so a week before the workshop, each researcher received a supply of high-energy snacks through the post. We also provided colored sticky notepads and permanent markers with similar tip sizes to ensure that all scanned sketches and writings were clear when uploaded on the web-based discussion board.

Textbox 1. Recommendations for practice, policy, and research.

Practice

- Although some participants only join the workshop for short periods, researchers and technical colleagues have to concentrate on the web for long periods. Adequate breaks, attention to nutrition, and general comfort are essential.
- The web-based platforms may be unfamiliar to many participants; training before the workshop gives confidence and helps reduce technical problems on the day.
- Owing to the relatively low cost and a short time spent from development to testing, the web-based design sprint methodology may be suitable for low-resource settings.
- Remote conduct ensured that high-risk stakeholders were shielded during a pandemic and overcame geographical barriers.

Policy

- The process is *a sprint* so that the end product can be developed to a short timescale to meet pressing deadlines.

Research

- The feasibility study of conducting a more extensive scale web-based intervention design program is necessary to ensure its practicality.

Conclusions

Working with people with limited health literacy enabled the development of an app that could support them in self-managing their asthma. Specific components included sources of information on asthma, pictorial asthma action plan, simple language, audio-visual prompts, and rewards for supporting adherence to daily therapy and scheduled reviews. Despite practical challenges, a 5-day web-based design workshop proved to be manageable, enabling meaningful engagement from patients and HCPs so that a prototype is now ready for feasibility testing.

Acknowledgments

This research was commissioned by the UK RESPIRE using UK Aid from the UK Government. The views expressed in this publication are those of the authors and not necessarily those of the National Institute for Health Research or the UK Department of Health and Social Care. The RESPIRE collaboration comprises UK grant holders, partners, and research teams as listed on the RESPIRE website. The authors are grateful to the study subjects for participating in this study. The authors would like to thank the experts from Pocket Pixel Sdn Bhd, Mohd Khalid IA, Goh A, and Zainal MM for their support in the conduct of this study.

Authors' Contributions

HS, HP, PYL, SSG, and IY were involved in formulating the ideas for the study. HS and JW planned and conducted training for the participants. HS, PYL, SSG, ATC, and JW conducted the workshop and collected the data. HS conducted data analysis, and all authors were involved in data interpretation. HS drafted the manuscript. All authors critically revised the manuscript for intellectual content. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Prompts for stakeholder discussion.
[DOCX File , 22 KB - jmir_v23i9e26434_app1.docx]

Multimedia Appendix 2
Utility and usability prompts.
[DOCX File , 21 KB - jmir_v23i9e26434_app2.docx]

Multimedia Appendix 3
Themes from the stakeholders' discussion.
[DOCX File , 25 KB - jmir_v23i9e26434_app3.docx]

Multimedia Appendix 4

Stages 2-4.
[DOCX File , 86 KB - jmir_v23i9e26434_app4.docx]

References

1. Pinnock H, Parke HL, Panagioti M, Daines L, Pearce G, Epiphaniou E, PRISMS and RECURSIVE groups. Systematic meta-review of supported self-management for asthma: a healthcare perspective. BMC Med 2017 Mar 17;15(1):64 [FREE Full text] [doi: 10.1186/s12916-017-0823-7] [Medline: 28302126]
2. Pinnock H, Epiphaniou E, Pearce G, Parke H, Greenhalgh T, Sheikh A, et al. Implementing supported self-management for asthma: a systematic review and suggested hierarchy of evidence of implementation studies. BMC Med 2015 Jun 01;13(1):127 [FREE Full text] [doi: 10.1186/s12916-015-0361-0] [Medline: 26032941]
3. Taylor S, Pinnock H, Epiphaniou E, Pearce G, Parke H, Schwappach A. A rapid synthesis of the evidence on interventions supporting self-management for people with long-term conditions: PRISMS–Practical systematic review of self-management support for long-term conditions. Health Serv Del Res 2014:02530. [doi: 10.3310/hsdr02530] [Medline: 25642548]
4. Sulaiman N, Aroni R, Thien F, Schattner R, Simpson P, Del Colle E, et al. Written Asthma Action Plans (WAAPs) in Melbourne general practices: a sequential mixed methods study. Prim Care Respir J 2011 Jun 18;20(2):161-169 [FREE Full text] [doi: 10.4104/pcrj.2011.00010] [Medline: 21336464]
5. Ställberg B, Lisspers K, Hasselgren M, Janson C, Johansson G, Svärdsudd K. Asthma control in primary care in Sweden: a comparison between 2001 and 2005. Prim Care Respir J 2009 Dec;18(4):279-286 [FREE Full text] [doi: 10.4104/pcrj.2009.00024] [Medline: 19455269]
6. Wiener-Ogilvie S, Pinnock H, Huby G, Sheikh A, Partridge MR, Gillies J. Do practices comply with key recommendations of the British Asthma Guideline? If not, why not? Prim Care Respir J 2007 Dec 21;16(6):369-377 [FREE Full text] [doi: 10.3132/pcrj.2007.00074] [Medline: 18026674]
7. Levy M, Andrews R, Buckingham R, Evans H, Francis C, Houston R. Why asthma still kills: the National Review of Asthma Deaths (NRAD). Royal College of Physcians. 2015. URL: https://www.rcplondon.ac.uk/projects/outputs/why-asthma-still-kills [accessed 2021-08-05]
8. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12(1):80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
9. Eckman MH, Wise R, Leonard AC, Dixon E, Burrows C, Khan F, et al. Impact of health literacy on outcomes and effectiveness of an educational intervention in patients with chronic diseases. Patient Educ Couns 2012 May;87(2):143-151. [doi: 10.1016/j.pec.2011.07.020] [Medline: 21925823]
10. Federman AD, Wolf MS, Sofianou A, O'Conor R, Martynenko M, Halm EA, et al. Asthma outcomes are poor among older adults with low health literacy. J Asthma 2014 Mar 14;51(2):162-167. [doi: 10.3109/02770903.2013.852202] [Medline: 24102556]
11. National Health and Morbidity Survey 2015. 2015. URL: http://www.iku.gov.my/nhms/ [accessed 2021-06-10]
12. Lee P, Khoo E. How well were asthmatic patients educated about their asthma? A study at the emergency department. Asia Pac J Public Health 2004 Jun 30;16(1):45-49. [doi: 10.1177/101053950401600108] [Medline: 18839867]
13. Zainudin BM, Lai CK, Soriano JB, Jia-Horng W, De Guia TS, Asthma Insights Reality in Asia-Pacific (AIRIAP) Steering Committee. Asthma control in adults in Asia-Pacific. Respirology 2005 Nov;10(5):579-586. [doi: 10.1111/j.1440-1843.2005.00753.x] [Medline: 16268910]
14. Salim H, Ghazali SS, Lee PY, Cheong A, Harrun NH, Isa SM, et al. Health literacy levels and its determinants among people with asthma in Malaysian primary healthcare settings: a cross-sectional study. BMC Public Health 2021 Jun 22;21(1):1186 [FREE Full text] [doi: 10.1186/s12889-021-11194-w] [Medline: 34158013]
15. Gordon NP, Crouch E. Digital information technology use and patient preferences for internet-based health education modalities: cross-sectional survey study of middle-aged and older adults with chronic health conditions. JMIR Aging 2019 Apr 04;2(1):e12243 [FREE Full text] [doi: 10.2196/12243] [Medline: 31518291]
16. Hopkins J. Networked individualism and networked families in Malaysia. In: Cabañes JV, Uy-Tioco CS, editors. Mobile Media and Social Intimacies in Asia. Dordrecht: Springer; 2020:31-45.
17. Golinelli D, Boetto E, Carullo G, Nuzzolese AG, Landini MP, Fantini MP. Adoption of digital technologies in health care during the COVID-19 pandemic: systematic review of early scientific literature. J Med Internet Res 2020 Nov 06;22(11):e22280 [FREE Full text] [doi: 10.2196/22280] [Medline: 33079693]
18. Malaysian Communications and Multimedia Commission Annual Report 2017. Ministry of Communications and Multimedia Malaysia, Putrajaya. 2020. URL: https://www.mcmc.gov.my/en/about-us/annual-reports/annual-reports# [accessed 2021-08-12]
19. Lee HY, Jin SW, Henning-Smith C, Lee J, Lee J. Role of health literacy in health-related information-seeking behavior online: cross-sectional study. J Med Internet Res 2021 Jan 27;23(1):e14088 [FREE Full text] [doi: 10.2196/14088] [Medline: 33502332]
20. Wu AC. The promise of improving asthma control using mobile health. J Allergy Clin Immunol Pract 2016 Jul;4(4):738-739. [doi: 10.1016/j.jaip.2016.04.003] [Medline: 27393782]

21. Hui C, Walton R, McKinstry B, Vasileiou E, Pinnock H. What do people with asthma want to see in an asthma self-management app? A review of views expressed in online social discussion forums. In: Proeedings of the Health Informatics Conference. 2017 Presented at: Health Informatics Conference 2017; Oct 3-4, 2017; Edinburgh p. 1-7 URL: https://www.research.ed.ac.uk/en/publications/what-do-people-with-asthma-want-to-see-in-an-asthma-self-manageme
22. Hui CY, Walton R, McKinstry B, Pinnock H. Time to change the paradigm? A mixed method study of the preferred and potential features of an asthma self-management app. Health Informatics J 2020 Jun 13;26(2):862-879 [FREE Full text] [doi: 10.1177/1460458219853381] [Medline: 31195877]
23. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. JMIR Mhealth Uhealth 2015 May 13;3(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4283] [Medline: 25972309]
24. Jake-Schoffman D, McVay M. Using the Design Sprint process to enhance and accelerate behavioral medicine progress: a case study and guidance. Transl Behav Med 2021 May 25;11(5):1099-1106. [doi: 10.1093/tbm/ibaa100] [Medline: 33057685]
25. Global Initiative for Asthma – GINA. 2020. URL: https://ginasthma.org/pocket-guide-for-asthma-management-and-prevention/ [accessed 2021-05-21]
26. British Thoracic Society. 2016. URL: https://www.brit-thoracic.org.uk/quality-improvement/guidelines/asthma/ [accessed 2021-05-03]
27. Knapp J, Zeratsky J, Kowitz B. Sprint: How to Solve Big Problems and Test New Ideas in Just Five Days. New York: Simon and Schuster; 2016:1-288.
28. Banfield R, Lombardo C, Wax T. Design Sprint: A Practical Guidebook for Building Great Digital Products. Sebastopol, California: O'Reilly Media; 2015:1-272.
29. Virzi R, Sokolov J, Karis D. Usability problem identification using both low- and high-fidelity prototypes. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1996 Presented at: CHI96: CHI '96 ACM Conference on Human Factors in Computing Systems; April 13 - 18, 1996; Vancouver British Columbia Canada p. 236-243. [doi: 10.1145/238386.238516]
30. Duong TV, Aringazina A, Baisunova G, Nurjanah, Pham TV, Pham KM, et al. Measuring health literacy in Asia: Validation of the HLS-EU-Q47 survey tool in six Asian countries. J Epidemiol 2017 Feb;27(2):80-86 [FREE Full text] [doi: 10.1016/j.je.2016.09.005] [Medline: 28142016]
31. Sørensen K, Van den Broucke S, Pelikan JM, Fullam J, Doyle G, Slonska Z, HLS-EU Consortium. Measuring health literacy in populations: illuminating the design and development process of the European Health Literacy Survey Questionnaire (HLS-EU-Q). BMC Public Health 2013 Oct 10;13(1):948 [FREE Full text] [doi: 10.1186/1471-2458-13-948] [Medline: 24112855]
32. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019 Jun 13;11(4):589-597. [doi: 10.1080/2159676X.2019.1628806]
33. Martinez W, Threatt AL, Rosenbloom ST, Wallston KA, Hickson GB, Elasy TA. A patient-facing diabetes dashboard embedded in a patient web portal: design sprint and usability testing. JMIR Hum Factors 2018 Sep 24;5(3):e26 [FREE Full text] [doi: 10.2196/humanfactors.9569] [Medline: 30249579]
34. Pfisterer KJ, Boger J, Wong A. Prototyping the automated food imaging and nutrient intake tracking system: modified participatory iterative design sprint. JMIR Hum Factors 2019 May 09;6(2):e13017 [FREE Full text] [doi: 10.2196/13017] [Medline: 31094336]
35. Seymour WS. In the flesh or online? Exploring qualitative research methodologies. Qual Res 2016 Aug 17;1(2):147-168. [doi: 10.1177/146879410100100203]
36. Adler CL, Zarchin YR. The "virtual focus group": using the Internet to reach pregnant women on home bed rest. J Obstet Gynecol Neonatal Nurs 2002 May;31(4):418-427. [doi: 10.1111/j.1552-6909.2002.tb00064.x] [Medline: 12146931]
37. Tuttas CA. Lessons learned using Web conference technology for online focus group interviews. Qual Health Res 2015 Jan 05;25(1):122-133. [doi: 10.1177/1049732314549602] [Medline: 25192765]
38. Tinschert P, Jakob R, Barata F, Kramer J, Kowatsch T. The potential of mobile apps for improving asthma self-management: a review of publicly available and well-adopted asthma apps. JMIR Mhealth Uhealth 2017 Aug 02;5(8):e113 [FREE Full text] [doi: 10.2196/mhealth.7177] [Medline: 28768606]
39. Yin HS, Gupta RS, Tomopoulos S, Wolf MS, Mendelsohn AL, Antler L, et al. Readability, suitability, and characteristics of asthma action plans: examination of factors that may impair understanding. Pediatrics 2013 Jan 03;131(1):116-126. [doi: 10.1542/peds.2012-0612] [Medline: 23209106]

Abbreviations

HCP: health care professional
RESPIRE: National Institute for Health Research Global Health Research Unit on Respiratory Health

Edited by R Kukafka; submitted 14.12.20; peer-reviewed by L Kayser, E Arden-Close; comments to author 14.02.21; revised version received 21.03.21; accepted 06.05.21; published 09.09.21.

Please cite as:
Salim H, Lee PY, Sharif-Ghazali S, Cheong AT, Wong J, Young I, Pinnock H, RESPIRE Collaboration
Developing an Asthma Self-management Intervention Through a Web-Based Design Workshop for People With Limited Health Literacy: User-Centered Design Approach
J Med Internet Res 2021;23(9):e26434
URL: https://www.jmir.org/2021/9/e26434
doi:10.2196/26434
PMID:34499039

Original Paper

Effectiveness of a Dyadic Buddy App for Smoking Cessation: Randomized Controlled Trial

Philipp Schwaninger[1], MSc; Corina Berli[1], PhD; Urte Scholz[1,2], PhD; Janina Lüscher[1], PhD

[1]Applied Social and Health Psychology, Department of Psychology, University of Zurich, Zurich, Switzerland

[2]University Research Priority Programme "Dynamic of Healthy Aging", Department of Psychology, University of Zurich, Zurich, Switzerland

Corresponding Author:
Philipp Schwaninger, MSc
Applied Social and Health Psychology
Department of Psychology
University of Zurich
Binzmühlestrasse 14/14
Zurich, 8050
Switzerland
Phone: 41 44 635 72 25
Email: philipp.schwaninger@psychologie.uzh.ch

Abstract

Background: Tobacco smoking is one of the biggest public health threats. Smartphone apps offer new promising opportunities for supporting smoking cessation in real time. This randomized controlled trial investigated the effectiveness of an app that encourages individuals to quit smoking with the help of a social network member (*buddy*) in daily life.

Objective: The objective of this study is to test the effectiveness of the SmokeFree buddy app compared with a control group with self-reported smoking abstinence and carbon monoxide (CO)–verified smoking abstinence as primary outcomes and self-reports of smoked cigarettes per day (CPD) as a secondary outcome.

Methods: A total of 162 adults who smoked participated in this single-blind, two-arm, parallel-group, intensive longitudinal randomized controlled trial. Around a self-set quit date (ie, 7 days before the self-set quit date and 20 days after) and 6 months later, participants of the intervention and control groups reported on daily smoking abstinence and CPD in end-of-day diaries. Daily smoking abstinence was verified via daily exhaled CO assessments. This assessment was administered via an app displaying results of exhaled CO, thus addressing self-monitoring in both groups. In addition, participants in the intervention group used the SmokeFree buddy app, a multicomponent app that facilitates social support from a buddy of choice.

Results: A significant reduction in CPD from baseline to the 6-month follow-up was observed among participants in both groups. Multilevel analyses revealed no significant intervention effect on self-reported and CO-verified daily smoking abstinence at the quit date and 3 weeks later. However, CPD was lower at the quit date and 3 weeks later in the intervention group than in the control group. No significant differences between groups were found for any outcome measures 6 months after the quit date. Overall, low app engagement and low perceived usefulness were observed.

Conclusions: Despite some encouraging short-term findings on the amount of smoking, the SmokeFree buddy app did not have beneficial effects on smoking abstinence over and above the self-monitoring control condition. Future studies should examine whether and what support processes can be effectively stimulated and how app use can be improved to better achieve this goal.

Trial Registration: ISRCTN Registry 11154315; https://www.isrctn.com/ISRCTN11154315

International Registered Report Identifier (IRRID): RR2-10.1186/s12889-019-7723-z

(J Med Internet Res 2021;23(9):e27162) doi:10.2196/27162

KEYWORDS

mHealth; smartphone app; smoking cessation; buddy; social support; mobile phone

Introduction

Background

Smoking kills more than 8 million people each year and remains one of the leading preventable causes of premature death worldwide [1,2]. Cigarette smoking is a major risk factor for life-threatening diseases such as lung cancer, coronary heart disease, and stroke and other noncommunicable diseases [2]. Despite the evidence that quitting smoking is associated with rapid improvements in health and reduced risk for noncommunicable diseases, in 2017, every fourth person aged 15 years and older smoked in Switzerland [3]. The high relapse rates indicate that quitting is difficult [4]. Various socioenvironmental stimuli associated with smoking may lead to high-risk situations for relapse, even long after quitting [5]. Strengthening the capability to manage such high-risk situations (eg, cue-induced cravings) is crucial for smoking cessation interventions [6,7]. The use of mobile health technologies (eg, smartphone apps) offers promising opportunities for smoking cessation interventions in everyday life, with the potential to deliver support in situations when it is most needed [8,9].

Smartphone Apps to Promote Smoking Cessation

In Switzerland, 92% of adults owned a smartphone in 2018 [10]. Considering this high prevalence of smartphone use, delivering smoking cessation interventions using smartphone apps is a promising approach. With the widespread reach and proximity to the user, interventions are accessible in everyday life [11]. The location and time independence may reduce perceived barriers to treatment (eg, transportation, time, and cost) and make potential engagement convenient [11,12]. Furthermore, apps can provide interactive and tailored intervention features as well as an interactive platform to allow instant sharing of experiences and resources (eg, social support) [11,12]. All these factors contribute to the cost-effectiveness of smoking cessation app interventions [11].

Considering the potential of app interventions, existing apps predominantly provide simple tools (eg, calculators, calendars, trackers, or distractors) [13,14]. The majority of available apps had very little evidence-based content to support quit attempts, such that in 55% of the apps, no behavior change techniques (BCTs) [15] were present [16]. In addition to the content of the smoking cessation app, their effectiveness is less clear. In the context of text messaging interventions for smoking cessation, research has found that such interventions are effective in improving smoking cessation rates and reducing health service costs [17-19]. In contrast, despite the wide availability of smoking cessation apps and the rapid release of new apps [13], only a few studies have tested their effectiveness [17]. For instance, a smartphone decision aid app including behavioral support was compared with a simpler information-only app in a randomized controlled trial (RCT) across different countries [20]. The decision aid app significantly increased continuous smoking abstinence at 1, 3, and 6 months after quitting [20]. A recent review of five studies that tested the effectiveness of smoking cessation apps showed no clear evidence for increasing quit rates [17]. There is a lack of scientific evidence on the effectiveness of smoking cessation apps. We echo the conclusion of the authors that more RCTs with long-term follow-ups (at least 6 months; Russell Standard [21]) are needed to provide evidence on the effectiveness of smoking cessation apps [17]. Therefore, this RCT tests the effectiveness of the SmokeFree buddy app, which was part of the SmokeFree campaign of the Federal Office of Public Health in Switzerland. It is a multicomponent app that, in particular, aims to facilitate social support from a buddy of one's own social network.

App features that connect individuals who smoke with health care workers, smokers and ex-smokers, or persons from their own social network to foster social support resources are promising [5]. Social support is conceptualized as coping assistance, that is, the supporters' engagement in coping efforts through the provision of emotional help (eg, demonstration of understanding and valuing) and practical assistance (eg, advice and informational and instrumental support) [22]. Social support may help individuals cope with high-risk situations for smoking (eg, cue-induced cravings) or buffering stress while quitting [9,23]. Research shows that face-to-face individual counseling, group therapy, and telephone counseling that convey social support (eg, practical assistance by identifying high-risk situations or emotional help by encouragement) were found to be effective for smoking cessation [24-26]. Partner or peer support interventions, however, did not show clear effectiveness in smoking cessation [9,27]. These intervention approaches are very diverse, and one explanation is that these partner or peer interventions were not successful in increasing social support in the first place [27]. Furthermore, research shows that support receipt might have null or negative effects on recipients depending on the timing, content, and way how support is provided [28]. More evidence regarding partner or peer support interventions is needed [9,27,29]. Apps provide promising potential to enhance partner or peer support resources and availability during actual experiences in people's everyday lives [30]. Content analyses of smoking cessation apps, however, have found that apps rarely reference users outside of the app to a quit helpline or provide opportunities to reach out for social support [13]. Only 2.7% (6/225) of rated Android smoking cessation apps had content regarding social support [14]. This RCT tested the SmokeFree buddy app that fosters support resources from a buddy of choice. This app offers theory-based instructions for the buddy on how to support individuals who smoke during a quit attempt. The instructions may positively influence the quality of support, and the accessibility of the app is important for the ideal timing of support exchanges (ie, when it is most needed). In addition, individuals who smoke rely on persons from their own personal social network rather than on unacquainted buddies. This may influence the tailoring of support exchanges because a buddy of choice should know the target person and their needs better than a stranger [31].

Aims of This Study

Given the research gaps outlined above, this study aims to test the effectiveness of the SmokeFree buddy app in daily life. The SmokeFree buddy app was developed by the Federal Office of Public Health in Switzerland in collaboration with the Institute of Global Health of the University of Geneva, Switzerland. The app is designed to be used simultaneously by an individual who smokes and a personal buddy of choice. In brief, a chat function

is the direct communication channel for all text messages and connects the individual who smokes and the personal buddy. The individual who smokes can, for example, indicate the current mood state and intensity and can use notification buttons such as sending an emergency message, communicating craving to smoke, or communicating a lapse. The app then informs the buddy about text messages and notifications and the current mood or craving and provides them with the option of reacting immediately, for example, with a supportive message. The app further provides evidence-informed background information (eg, how to cope with craving) as well as preset supportive messages for the buddy, which can be customized or supplemented with a personal comment. A performance statistic shows the achievements to date (eg, number of smoke-free days), and a knowledge base contains information on smoking and quitting, which is also visible and accessible to the buddy. For more details about the SmokeFree buddy app, please refer to the study protocol [32].

This RCT examined whether the SmokeFree buddy app promotes daily abstinence rates and reduces the number of cigarettes smoked per day (CPD) in adult participants who smoke at a self-set quit date, 3 weeks later (end of intervention), and 6 months later in comparison with a control group (CG) that did not use the SmokeFree buddy app [32]. We hypothesized that adult participants in the intervention group (IG) will show higher daily abstinence rates and lower CPD at the self-set quit date, 3 weeks later (end of intervention), and 6 months later than those in the CG (hypothesis 1). Furthermore, our RCT uses intensive longitudinal data to test the app in participants' everyday lives, which additionally allows the analysis of the potential time effects of the intervention. Therefore, we additionally assumed a higher increase in daily abstinence rates and a higher decrease in CPD over time in the IG than in the CG (hypothesis 2). The hypotheses were prospectively registered (ISRCTN 11154315). The primary outcome measures were daily self-reported smoking abstinence and daily smoking abstinence using exhaled carbon monoxide (CO) [32]. This RCT assessed smoking abstinence objectively on a daily basis with the iCO Smokerlyzer (Bedfont Scientific Ltd [33]), a CO-monitoring device. The secondary outcome measure was self-reports of smoked CPD [32].

Methods

Overview

We conducted a single-blind, two-arm, parallel-group, intensive longitudinal RCT *smoking cessation with smartphone apps* (for a detailed description, please see the study protocol [32]). The trial consisted of a *baseline diary* (3 consecutive days), a background assessment, a *challenge diary* from 7 days before the self-set quit date until 20 days after the self-set quit date (28 consecutive days), and a *follow-up diary* 6 months after the self-set quit date (3 consecutive days). Participants were randomly allocated to the intervention (SmokeFree buddy app) or control condition (no SmokeFree buddy app). This trial was approved by the Ethics Committee of the Faculty of Arts and Social Sciences of the University of Zurich (reference number: 17.12.13) and was prospectively registered (ISRCTN11154315).

Participants and Procedures

Participants were adults who smoked at least one cigarette per day, intended to quit smoking during the study, and owned a smartphone with access to mobile internet. The exclusion criteria were insufficient knowledge of the German language, working in 24-hour shifts, use of a smoking cessation app, or participation in a professional smoking cessation program. Participants were recruited via advertisements in newspapers, web-based platforms, webpages, and flyers at the university, medical facilities, and local companies from April 2018 to August 2019 in Switzerland. The eligibility criteria were assessed using a web-based screening questionnaire. Participants started with a web-based end-of-day diary questionnaire for 3 consecutive days (*baseline diary*). After completion, participants were invited to the lab for a background assessment. The interviewer conducting the background assessment randomly assigned participants to one of the two groups (IG and CG), according to a computer-generated allocation sequence that was concealed in a set of sealed, numbered envelopes. Participants were blinded to the group assignment (ie, single-blind RCT; for the detailed randomization procedure, refer to the study protocol [32]). At the background assessment, all participants provided written informed consent, completed a questionnaire, and were asked to self-set a quit date (BCT *goal setting of behavior* [15]) within the next 6 weeks. The self-set quit date determined the start of the 28-day diary period (*challenge diary*), starting 7 days before their self-set quit date. Furthermore, all participants received a personal iCO Smokerlyzer and instructions on how to use it with the corresponding app to measure the exhaled CO daily. The first measurement of CO with the iCO Smokerlyzer was conducted at the background assessment. All participants were instructed to fill out an end-of-day diary (daily text message with link to questionnaire) and to measure the exhaled CO daily using the personal iCO Smokerlyzer for 28 consecutive days during the *challenge diary* phase (7 days before the self-set quit date, on the self-set quit date until 20 days after the self-set quit date). Six months after the self-set quit date, participants in the CG and IG were asked to participate in the *follow-up diary* phase, with end-of-day diaries and daily measures of exhaled CO using the personal iCO Smokerlyzer for 3 consecutive days. As an incentive for study participation, all adults who smoke received a personal iCO Smokerlyzer with a value of CHF 60 (US $63). Furthermore, with completion of the 6-month follow-up, participants were given entry into a lottery with a main prize of CHF 200 (US $209) and 40 prizes with a value of CHF 50 (US $52) shopping vouchers. Buddy participants were reimbursed with CHF 50 (US $52) at the completion of the study. On the basis of a power of 0.80 and a two-tailed type 1 error probability of P=.05, an a priori power analysis yielded a sample size of 128 to detect a 24% difference in smoking abstinence for the IG compared with the CG, drawing on meta-analyses of mobile phone interventions on short- and long-term smoking abstinence (rate ratios [RRs] between 1.7 and 2.1 [34,35]). Assuming an attrition rate of 25%, 6 months after the quit date resulted in a total sample size of 160 participants (80 participants per group). For more details, please refer to the study protocol [32].

Intervention

As described above, participants in the CG and IG set a self-set quit date and were instructed to fill out the end-of-day diaries and to measure exhaled CO with the iCO Smokerlyzer during the *challenge diary* phase. The CO Smokerlyzer app displayed participants' CO results and progress with smoking cessation and thus addressed the BCTs *feedback on behavior* and *self-monitoring of behavior* [15]. Only the IG participants were additionally introduced to the SmokeFree buddy app. Participants in the IG were instructed to identify a personal buddy of choice and to use the SmokeFree buddy app during the *challenge diary* phase (starting 7 days before the self-set quit date). The SmokeFree buddy app is available for free in the App Store and Google Play Store (version 4.0). The app comprises the following BCTs: *social support (emotional), social support (practical), social reward, feedback on behavior, self-monitoring of behavior, others monitoring with awareness, discrepancy between current behavior and goal standard, nonspecific reward,* and *information about health consequences* [15]. The buddy had to be a nonsmoker for at least 6 months, owning a smartphone with access to mobile internet. The exclusion criteria for the buddy were insufficient knowledge of the German language and working in 24-hour shifts. The buddy also provided informed consent and received a background questionnaire with instructions regarding the SmokeFree buddy app. The buddy participants were also instructed to fill out an end-of-day diary and to use the SmokeFree buddy app during the *challenge diary* phase (for 28 consecutive days) to interact with the participant who smoked. No recommendations regarding timing, frequency, and intensity of use were provided to participants who smoked and their buddies.

Measures

Primary Outcome: Daily Smoking Abstinence (Self-reported and CO Verified)

Participants indicated in a daily evening diary questionnaire whether they smoked that day or not by answering the question, "Did you smoke today (including only one puff)?" (coded as 0=yes and 1=no) [36,37]. To objectively assess smoking abstinence, participants were instructed each evening to assess exhaled CO with their personal iCO Smokerlyzer. The iCO Smokerlyzer is a CO-monitoring device that is used with a corresponding Smokerlyzer app [33]. At the end of the end-of-day diary questionnaire, participants were instructed to take a CO measurement with their personal iCO Smokerlyzer and the corresponding app. The Smokerlyzer app displayed the test values in parts per million (ppm) and automatically generated a result report. Participants were instructed to send their result report directly from the app via email to the study team and to report the ppm value in the end-of-day diary. The iCO Smokerlyzer test values were categorized as 1 for abstinent (0-6 ppm) and as 0 for not abstinent (>6 ppm) [33,38]. A recent study points to comparable results between the iCO Smokerlyzer and the more studied piCO$^+$ Smokerlyzer, which has been shown to validly distinguish between smokers and nonsmokers [38,39].

Secondary Outcome: CPD

If participants answered the smoking abstinence measure with yes (=today smoked), they were asked to report the number of cigarettes they had smoked that day. Otherwise, the daily number of cigarettes smoked was coded as 0 [37].

Covariates

The daily use of nicotine replacement products (NRPs) and a variable denoting whether days were weekdays or weekends were included as covariates in the sensitivity analyses. Research shows that using NRPs increases the probability of smoking cessation success [40], and smoking behavior has been shown to differ during weekends [41], which might be related to trigger exposure. NRPs were not provided to participants; however, the SmokeFree buddy app (ie, the knowledge base feature) incorporated encouragement to use NRPs for smoking cessation. Participants indicated in the daily diary whether they used NRPs or not by answering the question "Did you use nicotine replacement products today (eg, nicotine gum, nicotine patch, medications as for example Champix, etc)?" (0=no and 1=yes). Furthermore, with each timestamp of the daily diary, a variable was calculated, whether it was a weekend or not (0=no weekend and 1=weekend).

Data Analysis

Intention to Treat

In line with the outcome criteria recommended for smoking cessation trials (Russell Standard [21]), our final models included all randomized subjects (intention to treat). Participants who were lost to the *challenge* or *follow-up diary* phase were treated as smoking for the analyses, with the primary outcome variables of daily abstinence (0=not abstinent and 1=abstinent).

Analysis of the Intervention Effect at the Quit Date and End of the Intervention

To account for the nested structure of repeated measures within individuals, multilevel modeling was used [42]. To examine the intervention effect on the two dichotomous primary outcome measures, we ran logistic regressions using generalized linear mixed models (GLMMs). The continuous outcome measure, CPD, was a count variable that was highly skewed with a large number of zeroes. Therefore, we applied a GLMM for count outcomes using a negative binomial distribution with a logarithmic link function [43]. Owing to the logarithmic link function, regression coefficients are on a log scale and interpreted as RRs. The distance to an RR of 1 is interpreted as the percentage increase (above one) or decrease (below one) in the outcome for a one-unit increase in the predictor [44]. The specified multilevel models use all available diary entries using maximum likelihood estimations [42]. All analyses were conducted using SPSS 26.

To test our research question and corresponding hypotheses, all outcome measures were modeled as a function of *time*, *group* (coded as 0=CG and 1=IG; hypothesis 1), and *group by time* (hypothesis 2). For this purpose, a linear *time* variable for the 28 consecutive days of the *challenge diary* was created (in 1-day units). To test group differences at the quit date and 3 weeks later, the *time* variable was centered accordingly (centered on

the quit date: day 8=0 or centered on the last day: day 28=0). An initial graphic display of the data showed a discrete change in the outcomes at the quit date, indicating two qualitatively distinct phases before and after. To model this nonlinear trajectory over time (shift), a dummy-coded variable *quit* was computed (coded as 0=quit day and days after the quit date [days 8-28] and 1=days before the quit date [days 1-7]). To test group differences in this shift, we included an interaction term of the *quit* and the *group* variable (*group by quit*). Finally, to test for differential time effects before and after the quit date, we generated and added a second time variable reflecting an interaction between *quit* and *time* (ie, *quit by time* with the following coding: day 1=−7, day 2=−6, ..., day 7=−1, day 8=0, day 9=0, day 10=0, etc). No differential time effects of the two groups before the quit date emerged; therefore, this interaction was not included in the final model. In all analyses, we routinely included the grand-mean-centered CPD during the *baseline diary* phase as an indicator of nicotine dependence as a covariate. All models specified a maximal random effects structure, which, in the case of nonconvergence, was successively reduced until convergence was achieved [45].

Analysis of the Intervention Effects at Follow-up

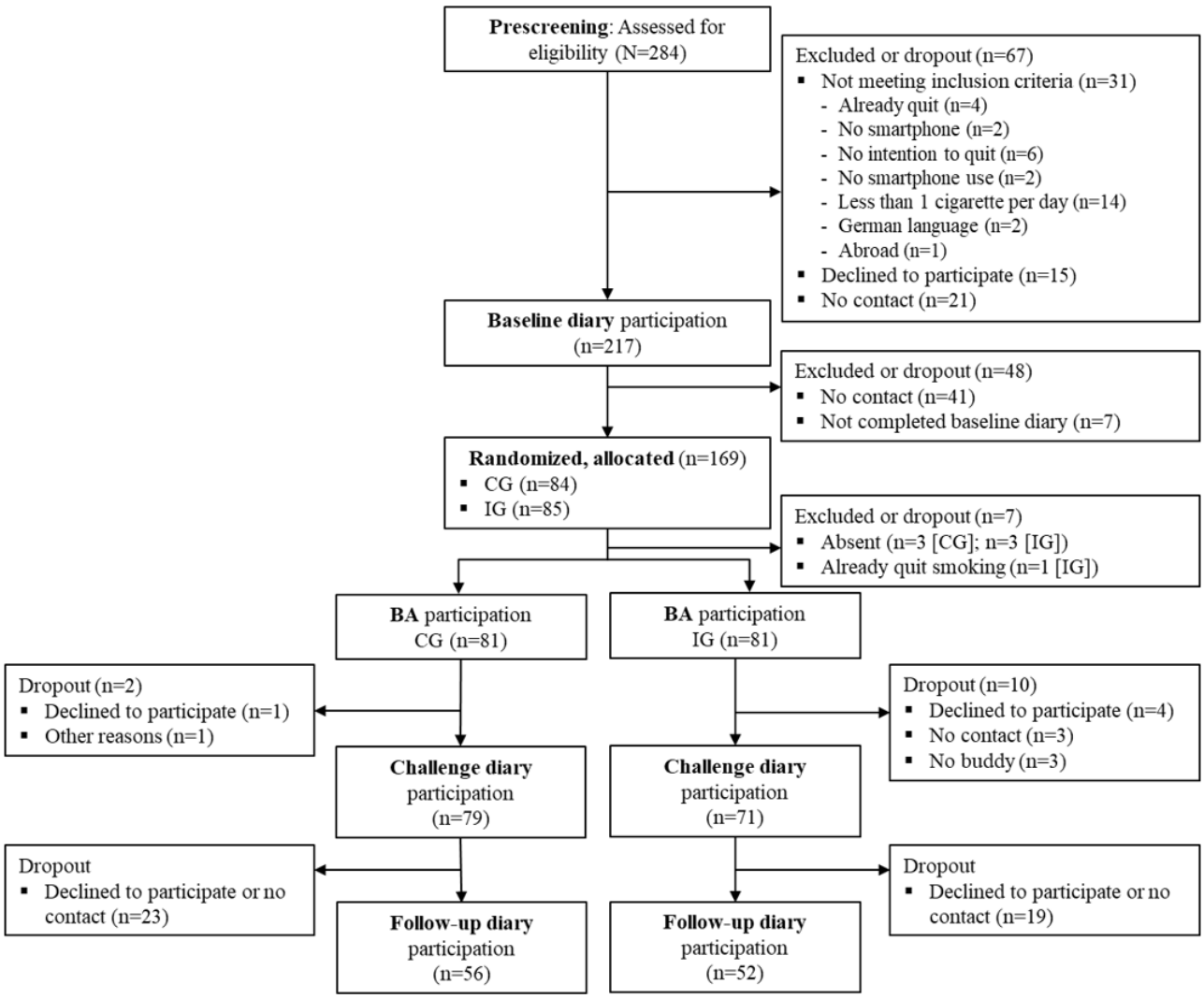
To investigate the intervention effect 6 months after the quit date, group differences regarding the primary and secondary outcome measures were examined by conducting independent samples two-tailed *t* tests (averaged days 1-3 of the *follow-up diary* phase per person). If variances were unequal between the groups, Welch *t* test was used. We used a Bonferroni-adjusted significance level to account for multiple comparisons.

Results

Sample

A total of 284 participants were screened, and 169 (59.5%) participants who completed the *baseline diary* were randomized to the CG (n=84 in total) and IG (n=85 in total; refer to Figure 1 for the participant flow).

Figure 1. Participant flow. BA: background assessment; CG: control group; IG: intervention group.



Overall, 162 randomized participants attended the background assessment, 81 in the CG and 81 in the IG. Table 1 displays the background characteristics of the participants in the CG and IG. Participants in the CG and IG did not differ on these variables at the background assessment (Table 1), proving successful randomization. The 162 participants had a mean age of 31.3

(SD 10.9) years and smoked 12.8 CPD (SD 7.1) before study participation. The majority of the participants were male (89/162, 54.9%). There were no significant background differences between groups regarding the intention and the desire to stop smoking reported at the background assessment as well as the self-regulation variables such as self-efficacy, action control, and social support receipt reported during the *baseline diary* (Table 1).

Of the 162 participants, 12 (7.4%) did not start with the *challenge diary*. By treating the dropout as smoking (intention to treat), the analyses of the primary outcomes had the maximum number of available data points (N=162; 3339/4536, 73.61% possible data points). Due to technical problems with the iCO Smokerlyzer, there were additional missing values for the CO-verified abstinence measure. CO measures were completely missing for 8 participants (103 missing data points). Additional missing CO values occurred due to various reasons, such as iCO Smokerlyzer did not work (57 days), iCO Smokerlyzer or smartphone forgotten (36 days), no time (5 days), or other unknown reasons (34 days). Therefore, the analysis of the CO-verified abstinence measure had 235 additional missing data points (n=154; 3104/4536, 68.43% possible data points). Secondary outcome data, CPD, were available for 97% (79/81) of the participants in the CG and 88% (71/81) of the participants in the IG (total n=150; 3003/4536, 66.2% possible data points).

Table 1. Baseline characteristics of the target persons in the intervention group and control group (N=162).

Variables	Intervention group (n=81)	Control group (n=81)	Intervention group vs control group[a]	*P* value
Gender, n (%)			0.2 (1)[b]	.64
Female	35 (43)	38 (47)		
Male	46 (57)	43 (53)		
Unmarried, n (%)	65 (80)	64 (79)	3.7 (2)[b]	.16
Higher education, n (%)	32 (40)	38 (47)	2.2 (3)[b]	.53
Employed, n (%)	35 (43)	41 (51)	3.1 (4)[b]	.54
Age (years), mean (SD)	31.32 (11.50)	31.32 (10.27)	0.00 (160)[c]	.99
Daily number of cigarettes (background assessment), mean (SD)	13.07 (7.33)	12.44 (6.83)	−0.57 (160)[c]	.57
Nicotine dependence, mean (SD)	3.33 (2.08)	3.56 (2.17)	0.67 (160)[c]	.51
Exhaled carbon monoxide (ppm[d]), mean (SD)	9.81 (6.34)	10.65 (9.05)	0.67 (139)[c]	.50
Intention to stop smoking, mean (SD)	5.40 (0.71)	5.33 (0.77)	−0.58 (160)[c]	.56
Desire to stop smoking, mean (SD)	5.00 (0.70)	4.95 (0.67)	−0.46 (160)[c]	.65
Baseline self-efficacy[e], mean (SD)	2.83 (1.34)	2.72 (1.24)	−0.55 (160)[c]	.59
Baseline action control, mean (SD)	2.48 (1.06)	2.75 (1.11)	1.54 (160)[c]	.13
Baseline social support receipt, mean (SD)	1.22 (1.87)	1.56 (1.85)	1.18 (160)[c]	.24

[a]Groups were compared for baseline characteristics using independent *t* tests for continuous data and chi-square tests for categorical data.

[b]Chi-square (*df*) values for categorical data.

[c]*t* test (*df*) values for continuous data.

[d]ppm: parts per million.

[e]Baseline variables were assessed during the *baseline diary* phase.

In total, 108 participants completed the *follow-up diary* phase (Figure 1). There were no significant differences in terms of background characteristics between participants who dropped out of the study and participants with follow-up data (Table S1 in Multimedia Appendix 1). By treating the dropout as smoking (intention to treat), the analyses regarding the primary outcomes included the full sample (N=162). Due to missing or not working iCO Smokerlyzer devices, the CO-verified abstinence measure was available for 86.4% (140/162) participants. Secondary outcome data were available for 66.7% (108/162) participants.

Descriptive Results

Table 2 shows the average CPD during the *baseline diary* phase and *challenge diary* phase and the 3 weeks and 6 months continuous abstinence rates by groups. The IG had a higher 3 weeks and 6 months continuous abstinence rate than the CG; however, these differences did not reach significance. For the CG, dependent *t* tests indicated a significant decrease in CPD from the baseline diary to the challenge diary *(postquit;* t_{72}=11.29; *P*<.001; *d*=2.66) and from the baseline diary to the follow-up diary (t_{55}=6.14; *P*<.001; *d*=1.66). The IG also showed a significant decrease in CPD from the baseline diary to the

challenge diary *(postquit;* t_{70}=13.25; *P*<.001; *d*=3.17) and from the baseline diary to the follow-up diary (t_{51}=6.51; *P*<.001; *d*=1.82). Figure 2 displays the means of the three outcome measures over the 28 days of the *challenge diary* by groups.

Table 2. Average cigarettes per day during the three diary phases and 3 weeks and 6 months continuous abstinence rates for the intervention group and control group (N=162).

Variables	Intervention group (n=81)	Control group (n=81)	Intervention group vs control group[a]	*P* value
CPD[b] baseline diary (N=162), mean (SD)	12.41 (7.50)	11.48 (7.08)	−0.81 (160)[c]	.42
CPD challenge diary prequit (n=149), mean (SD)	12.31 (7.29)	10.48 (6.15)	−1.67 (147)[c]	.10
CPD challenge diary postquit (n=144), mean (SD)	1.99 (4.03)	2.51 (3.14)	0.86 (142)[c]	.39
CPD follow-up diary (n=108), mean (SD)	6.13 (8.17)	4.99 (5.74)	−0.84 (91)[c]	.40
3 weeks continuous abstinence (ITT[d]) (N=162), n (%)	25 (30.9)	19 (23.5)	1.1 (1)[e]	.19
6 months continuous abstinence[f] (ITT) (N=162), n (%)	18 (22.2)	11 (13.6)	2.1 (1)[e]	.11

[a]Groups were compared for number of cigarettes per day and continuous abstinence rates using independent *t* tests for continuous data and chi-square tests for categorical data.

[b]CPD: cigarettes per day; the available data for the variable cigarettes per day varied due to dropout.

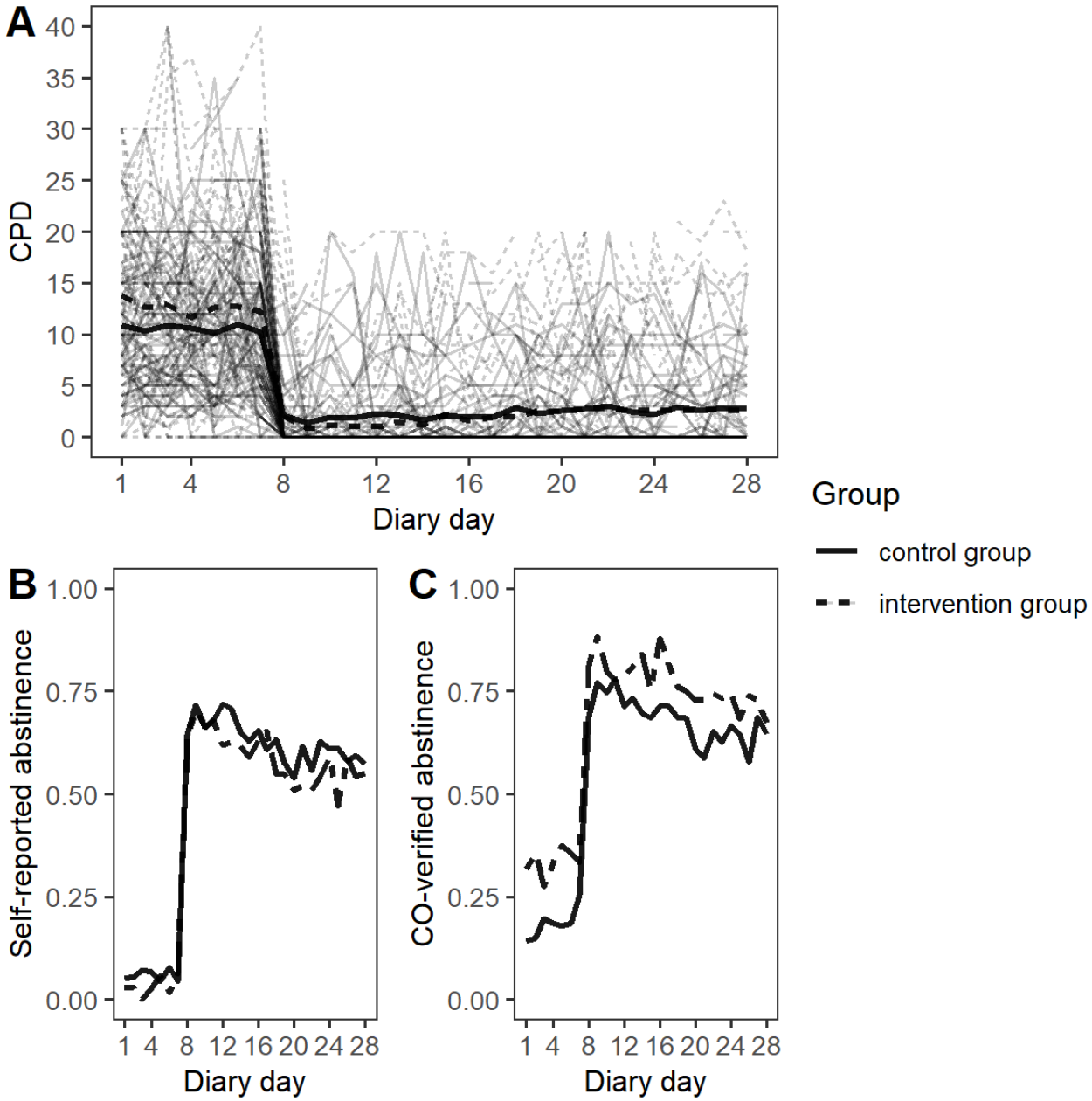
[c]*t* test (*df*) values.

[d]ITT: intention to treat; lost participants were treated as smoking (full sample).

[e]Chi-square (*df*) values.

[f]Six months continuous abstinence is defined as ≤5 cigarettes since quit date and no smoking during the follow-up diary phase (carbon monoxide verified).

Figure 2. Means of cigarettes per day (A), self-reported abstinence (B) and carbon monoxide–verified abstinence (C) over 28 days of the challenge diary phase by groups. CPD: cigarettes per day; CO: carbon monoxide.



Intervention Effects at the Quit Date and End of Intervention

The results of the three GLMMs are presented in Tables 3 and 4. In terms of daily self-reported smoking abstinence (GLMM 1), the intercept describes the probability of smoking abstinence for the CG at the quit date (77.4%, 95% CI 59.4%-88.9%). No significant *group* effect emerged: the IG (85.1%, 95% CI 42.4%-97.8%) did not report higher smoking abstinence than the CG. No significant *time* trend and no significant *group-by-time* effect were observed. This indicates that the daily self-reported abstinence was stable for both groups from the quit date on (disconfirming hypothesis 2), and the nonsignificant group difference on the quit date remained until the end of the intervention (disconfirming hypothesis 1; probability of abstinence on day 28: CG: 50.2%, 95% CI 17.2%-83.1% and IG: 74.2%, 95% CI 6.2%-99.2%). These results are not in support of our hypotheses. However, for both groups, a significant increase in the probability of daily self-reported abstinence from the phase before the quit date to the phase after the quit date occurred. Before the quit date, no difference was observed in the probability of daily self-reported abstinence between the CG (0.47%, 95% CI 0.05%-4.3%) and the IG (0.66%, 95% CI 0.1%-27.5%; to test group differences before the quit date, we recentered the time variable accordingly [centered on the first day: day 1=0]). The level-2 random effects indicated considerable variation between participants in the intercepts and trajectories over time (Table S2 in Multimedia Appendix 1).

Table 3. Fixed effects of generalized linear mixed models testing the effect of the intervention group versus the control group on daily self-reported abstinence and carbon monoxide abstinence (generalized linear mixed models 1 and 2).

Fixed effects	GLMM[a] 1[b] (daily self-reported abstinence) (N=162)			GLMM 2[c] (daily carbon monoxide abstinence) (n=154)		
	b[d] (SE)	*P* value	OR[e] (95% CI)	b (SE)	*P* value	OR (95% CI)
Intercept[f]	1.23 (0.43)	.005	3.43 (1.47-8.01)	3.37 (0.50)	<.001	29.21 (10.88-78.46)
CPD[g] (baseline diary)[h]	−0.10 (0.03)	.002	0.91 (0.85-0.97)	−0.14 (0.03)	<.001	0.87 (0.82-0.93)
Group (0=control group; 1=intervention group)	0.51 (0.61)	.40	1.67 (0.50-5.55)	−1.01 (0.69)	.14	0.37 (0.10-1.40)
Time[i]	−0.06 (0.04)	.13	0.94 (0.87-1.02)	−0.06 (0.04)	.08	0.94 (0.88-1.01)
Group×time	0.03 (0.06)	.63	1.03 (0.92-1.15)	0.03 (0.05)	.48	1.03 (0.94-1.14)
Quit (0=after quit date; 1=before quit date)	−6.60 (0.72)	<.001	0.001 (0.00-0.01)	−4.36 (0.63)	<.001	0.01 (0.00-0.04)
Group×quit	0.35 (0.91)	.70	1.42 (0.24-8.42)	−0.08 (0.80)	.92	0.92 (0.19-4.43)
Quit×time	−0.003 (0.08)	.97	0.98 (0.86-1.16)	0.15 (0.08)	.06	1.16 (0.997-1.34)

[a]GLMM: generalized linear mixed model. Random effects are reported in Table S2 in Multimedia Appendix 1.

[b]Generalized linear mixed model 1 (logistic regression): N=162 persons with a maximum of 28 days; n=3339 out of 4536 possible diary entries.

[c]Generalized linear mixed model 2 (logistic regression): n=154 persons with a maximum of 28 days; n=3104 out of 4536 possible diary entries.

[d]b: unstandardized regression coefficient.

[e]OR: odds ratio.

[f]Intercept indicates the level of the outcome for the control group at quit date (day 8).

[g]CPD: cigarettes per day.

[h]The grand-mean-centered cigarettes per day during the *baseline diary* phase as covariate.

[i]Linear time trend centered on the quit date (day 8=0).

Table 4. Fixed effects of generalized linear mixed model testing the effect of the intervention group versus the control group on smoked cigarettes per day (generalized linear mixed model 3).

Fixed effects	GLMM[a] 3[b]: CPD[c] (n=150)		
	b[d] (SE)	*P* value	Rate ratio (95% CI)
Intercept[e]	−0.12 (0.16)	.44	0.88 (0.65-1.21)
CPD (baseline diary)[f]	0.09 (0.01)	<.001	1.09 (1.06-1.12)
Group (0=control group; 1=intervention group)	−0.71 (0.23)	.002	0.49 (0.31-0.78)
Time[g]	0.03 (0.01)	.01	1.03 (1.01-1.05)
Group×time	−0.001 (0.01)	.95	0.99 (0.97-1.03)
Quit (0=after quit date; 1=before quit date)	2.52 (0.17)	<.001	12.40 (8.89-17.30)
Group×quit	0.64 (0.23)	.006	1.89 (1.20-2.96)
Quit×time	−0.04 (0.03)	.13	0.96 (0.91-1.01)

[a]GLMM: generalized linear mixed model. Random effects are reported in Table S3 in Multimedia Appendix 1.

[b]Generalized linear mixed model 3 (negative binomial): n=150 persons with a maximum of 28 days; n=3003 out of 4536 possible diary entries.

[c]CPD: cigarettes per day.

[d]b: unstandardized regression coefficient.

[e]Intercept indicates level of the outcome for the control group at quit date (day 8).

[f]The grand-mean-centered cigarettes per day (cigarettes per day) during the *baseline diary* phase as covariate.

[g]Linear time trend centered on the quit date (day 8=0).

The GLMM 2 results with daily CO abstinence as an outcome confirmed the result pattern stated above. The intercept as the probability of CO abstinence for the CG at the quit date was 96.7% (95% CI 91.7%-98.7%). No significant *group* effect emerged: the IG (92.4%, 95% CI 50.1%-99.1%) did not report higher CO abstinence. No significant *time* trend from the quit date on and no significant *group-by-time* effect were observed. This indicates that the daily CO abstinence was stable for both

groups from the quit date on (disconfirming hypothesis 2), and the nonsignificant group difference on the quit date remained until the end of the intervention (disconfirming hypothesis 1; probability of CO abstinence on day 28: CG: 89.6%, 95% CI 69.2%-97% and IG: 87.6%, 95% CI 20.7%-99.3%). These results are not in support of our hypotheses. However, a significant increase in the probability of daily CO abstinence from the phase before the quit date to the phase after the quit date occurred for both groups. No significant group differences were observed in the probability of daily CO abstinence before the quit date between the CG (27.2%, 95% CI 3.9%-77.4%) and the IG (25.6%, 95% CI 0.2%-93.8%). The level-2 random effects indicated considerable variation between participants in the intercepts and trajectories over time (Table S2 in Multimedia Appendix 1).

In terms of CPD as an outcome (GLMM 3), the intercept describes the estimated CPD on the quit date in the CG. The average number of cigarettes smoked at the quit date in the CG was low, with about 1 cigarette (0.88) compared with about 12 CPD before the quit date. A significant difference emerged between the IG and CG (b=−0.71, SE 0.23; RR 0.49; *P*=.002). This result indicates that participants in the IG smoked approximately half of the number of cigarettes at the quit date compared with participants in the CG. A significant positive *time* trend (*time*; days 8-28), but no significant *group-by-time* effect emerged. This indicates that for both groups, CPD increased across days from the quit date on. The significant difference between groups remained until the end of intervention, with the CG reporting smoking on average 1.5 cigarettes and the IG on average half the amount (b=−0.73, SE 0.25; RR 0.48; *P*=.003). These results are in support of our hypothesis 1 but not in support of our hypothesis 2, assuming different time trends across groups. A significant decrease in CPD from the phase before the quit date to the phase after the quit date occurred. This nonlinear change in CPD was significantly greater for the IG. The level-2 random effects indicated considerable variation among the participants in the intercepts (Table S3 in Multimedia Appendix 1).

Sensitivity Analyses

We conducted sensitivity analyses of the intervention effect at the quit date and the end of the intervention to test whether results differed when controlling for the use of NRPs and a variable denoting whether days were weekdays or weekends (see the corresponding results in Tables S4 and S5 in Multimedia Appendix 1). Due to the additional covariates, we could only include participants with *challenge diary* entries (n=150). The result pattern of the models with daily CO abstinence (GLMM A2; Table S4 in Multimedia Appendix 1) and CPD (GLMM A3; Table S5 in Multimedia Appendix 1) as outcome remained the same. In the model with daily self-reported smoking abstinence (GLMM A1; Table S4 in Multimedia Appendix 1), an intervention effect emerged, such that the probability of self-reported abstinence on the quit date was significantly higher for the IG (92.6%, 95% CI 67.2%-98.7%) than for the CG (76.6%, 95% CI 61%-87.2%). This significant difference between groups remained until the end of the intervention (no difference in time trend between groups). The additional covariates were not the reason for the significant group effect on self-reported smoking abstinence, as the effect remained when running a model without the additional covariates (n=150; Table S6 in Multimedia Appendix 1; GLMM A4 and A5). These analyses excluded the dropout participants, that is, the 12 participants without *challenge diary* data (no intention to treat), and therefore estimated group differences less conservatively.

Intervention Effects at Follow-up

At the 6 months follow-up, regarding the outcome daily self-reported abstinence, we analyzed the full sample (intention to treat) and found no difference between the CG (mean 0.26, SD 0.42; n=81) and IG (mean 0.32, SD 0.46; n=81; t_{160}=−0.93; *P*=.36; *d*=−0.15). We found the same result pattern for the daily CO abstinence. No group differences emerged between the CG (mean 0.36, SD 0.47; n=67) and IG (mean 0.40, SD 0.48; n=73; t_{138}=−0.48; *P*=.63; *d*=−0.08). The same result patterns were found in the sensitivity analyses when excluding dropouts from the analysis (no intention to treat). For daily self-reported abstinence, no difference emerged between the CG (mean 0.37, SD 0.46; n=56) and IG (mean 0.50, SD 0.49; n=52; t_{106}=−1.40; *P*=.16; *d*=−0.27). In addition, for daily CO abstinence, no difference emerged between the CG (mean 0.55, SD 0.49; n=44) and IG (mean 0.67, SD 0.46; n=44; t_{86}=−1.13; *P*=.26; *d*=−0.24). On average, participants in the CG smoked fewer CPD during the *follow-up diary* phase than participants in the IG (Table 2). This difference was not significant (t_{91}=−0.84; *P*=.40; *d*=−0.18).

Intervention Fidelity

Of the 81 participants in the IG, 10 did not start the *challenge diary* (Figure 1). Of the remaining 71 participants, 1 did not find a buddy but started the *challenge diary* and used the SmokeFree buddy app alone. Excluding this participant did not change the result pattern. To measure app engagement, participants in the IG were asked if they used the SmokeFree buddy app on a daily basis. In addition, the daily perceived usefulness of the app was assessed. Furthermore, we had access to the Flurry Analytics account of the app developer. Flurry Analytics is a platform for tracking and analyzing users' app engagement (app sessions). Of 47 participants we obtained informed consent to analyze objective data of their SmokeFree buddy app use (Flurry Analytic). Overall, we observed low app engagement and low reported perceived usefulness (Table S5 in Multimedia Appendix 1). A total of 3 of the 71 participants indicated that they had never used the app through the *challenge diary* phase (no objective use data were available). Excluding these 3 participants did not change the result pattern.

Discussion

Principal Findings

The aim of this study is to examine the effectiveness of the SmokeFree buddy app, using an RCT with intensive longitudinal data and daily objective measures of smoking abstinence via exhaled CO in participants' daily lives. The SmokeFree buddy app group was compared with a CG that did not use the SmokeFree buddy app. Both groups set a quit date, filled in daily diaries, and measured exhaled CO with an app targeting self-monitoring and feedback on behavior. First, we

hypothesized that the IG will show higher daily abstinence and lower CPD at the quit date, 3 weeks later, and 6 months later than the CG (hypothesis 1). Second, we hypothesized a higher increase or decrease in the outcome measures over time from the quit date to 3 weeks later in the IG than in the CG (hypothesis 2).

Both groups significantly reduced their CPD from the baseline diary phase to the follow-up diary phase. The observed 6 months continuous abstinence rates (CO verified; Table 2) in the IG (18/81, 22%) and CG (11/81, 14%) were promising, as other recent RCTs of smoking cessation apps show continuous abstinence rates (6 months) from 7.8% [46], 10.2% [20] to 16.1% [47]. The CG was found to comparably reduce smoking and achieve similar abstinence rates as the IG. Therefore, our findings do not confirm hypothesis 1 with regard to daily abstinence rates. Regarding the secondary outcome, we found a significantly higher decrease from before to after the quit date in CPD for the IG and significantly lower CPD at the quit date and 3 weeks later, but not 6 months later, compared with the CG. Disconfirming hypothesis 2, we did not find differences in the increase or decrease in the outcome measures over time between the IG and CG. In both groups, the abstinence measures were low prequit, instantaneously shifted to a high level at the quit date, and remained stable during the 3 weeks postquit. In both groups, CPD was high prequit, instantaneously shifted to a low level at the quit date, and then slightly increased during the 3 weeks postquit (Figure 2). At the 6-month follow-up, CPD levels were approximately half of the CPD levels prequit (Table 2).

To our knowledge, this is the first RCT of an app for smoking cessation that focuses particularly on facilitating social support from a personally chosen buddy. Enhancing social support resources should theoretically help individuals to quit smoking [5,9]. There is evidence from intensive longitudinal studies that higher daily support receipt from one's partner was related to less daily smoking, and effects were more pronounced after a self-set quit date when support is theoretically most needed [23,37]. Previous intervention studies regarding partner or peer social support to improve smoking cessation, however, could not clearly demonstrate the effectiveness of smoking cessation compared with a CG [9,27]. An explanation is that these partner or peer support interventions were not capable of increasing social support in the first place [27]. Our findings match the results of these social support interventions, with no consistent evidence of its effectiveness [27]. The higher decrease and lower levels of CPD at the quit date and 3 weeks later might be explained by the presence of a buddy of choice which increased support and self-regulatory efforts (eg, self-monitoring and self-efficacy) to cope with single smoking events in the short term. However, the analysis of the secondary outcomes did not include all randomized participants. Excluding the 12 participants who did not start the challenge diary phase led to significant intervention effects on self-reported daily smoking abstinence at the quit date and the end of the intervention (Table S6 in Multimedia Appendix 1).

An explanation for the null effects of the intervention could be that the SmokeFree buddy app was compared with a CG that reported on their smoking and assessed CO on a daily basis, which is a strong self-monitoring intervention itself. It might be that the effect of the SmokeFree buddy app over and above the self-monitoring component was smaller than assumed. Another potential explanation for our inconsistent findings might be the low observed app engagement (Table S7 in Multimedia Appendix 1). At the end of the intervention, participants reported their reasons for potentially low app engagement. One reason often mentioned was the preference for established text messaging channels (WhatsApp and SMS text messages), phone calls, or face-to-face communication, which may have made the additional app medium redundant. Several participants also stated that they did not experience any benefits from using the app and did not feel the need to use the app more often. This is in line with reports of the low perceived usefulness of the app (Table S7 in Multimedia Appendix 1) and may have attenuated the evidence-informed support provision of buddy participants. Moreover, some participants also indicated that they generally use few apps, forgot to use the app, or stopped using the app due to relapse. Furthermore, other social exchange processes may have co-occurred, which may have been perceived as aversive or unwanted (eg, social negativity [48]) and so have no beneficial effects on smoking cessation.

Research has shown a general lack of engagement with digital interventions [49]. Low engagement with app content may limit the potential intervention effects [50]. An explanation for the low app engagement and inconsistent findings may be that the main content of the SmokeFree buddy app to enhance social support resources may not be attractive or fruitful for everyone, especially for individuals who do not rely on social relationships during health behavior change (eg, smoking cessation as *solo struggle*) [31,51]. In line with this, the intensive longitudinal data of this study indicate individual heterogeneity regarding the outcome measures at the quit date and in their time trajectories. Future studies should investigate the effects of the SmokeFree buddy app on hypothesized mechanisms to gain more insight into whether and what processes have been stimulated [32]. More research on the engagement with digital interventions and on factors most related to intervention effectiveness is needed (refer to the conceptual framework of Perski et al [50]). For instance, reminders have been found to be a simple strategy to increase user engagement with a digital smoking cessation intervention, such as tailored text messaging or email prompts [52,53]. In addition, human guidance (eg, by the study coordinator or health care worker) may also improve engagement with digital behavior change interventions [54] because human monitoring may lead to a certain accountability to one's actions or inactions (eg, supportive accountability) [55]. Finally, in the context of social support interventions, we have to deepen our understanding of how, for whom, and under what conditions social support beneficially influences smoking behavior [31].

Strengths and Limitations

This study had several strengths. First, the study provides the first experimental evidence of the SmokeFree buddy app, an app with a new approach of buddy support in everyday life compared with unacquainted buddies in other intervention studies. Second, we verified our primary outcome measure of

self-reported daily smoking abstinence with daily assessments of exhaled CO with the iCO Smokerlyzer. Third, the intensive longitudinal data are a major strength of this study. The daily assessment gave us a more precise display of smoking behavior and verification of continuous abstinence after a quit date compared with macrotime follow-ups only [56]. Furthermore, the intensive longitudinal study design allowed us to investigate the temporal developments of intervention effects and assess the heterogeneity of trajectories between individuals. We assessed self-reported app engagement during the study and measured it in real time. The daily diary design with daily CO measures increases the ecological validity of our data and is the first of its kind.

This RCT also had some limitations that need to be acknowledged. First, this RCT investigated a multicomponent app and thus is not able to disentangle the effects of individual intervention components. Future app intervention development should be guided by the multiphase optimization strategy [57] to identify effective social support components for smoking cessation first before running an RCT that evaluates a multicomponent intervention package [58,59]. Second, our data indicated low app engagement, which may have limited potential intervention effects.

Conclusions and Implications

Although the abstinence rates of this study are comparable with those of other studies of smoking cessation apps [17] and both conditions (IG and CG) led to significantly decreased CPD, we found that the SmokeFree buddy app was not superior to a CG with primarily self-monitoring components. The app provides evidence for a greater short-term reduction of CPD. There is a need for more research on the effectiveness of and the engagement with smoking cessation apps, considering the wide availability and the rapid release of new smoking cessation apps. App components aimed at enhancing beneficial social interactions with health care providers, partners, or peers to support smoking cessation require further investigation [60]. This study contributes to the scientific evidence on the effectiveness of smoking cessation apps in everyday life.

Acknowledgments

JL (CR12I1_166348/1) was funded by the Swiss National Science Foundation while writing this manuscript. The assessment of the Flurry Analytic data was supported by the Federal Office of Public Health. The funder played no role in the design of the study, the analysis and interpretation of data, or in writing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables regarding the comparison of baseline characteristics between drop-out participants and participants with follow-up data (Table S1); random effects results of the generalized linear models 1-3 (Tables S2 and S3); additional generalized linear models testing effects of the intervention group versus the control group on all three outcomes with additional covariates (Tables S4 and S5); additional generalized linear models without the drop-out participants coded smoking (Table S6); and self-reported daily app use, daily objective app use, and daily reported perceived usefulness of the SmokeFree buddy app (Table S7).
[DOCX File , 41 KB - jmir_v23i9e27162_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2102 KB - jmir_v23i9e27162_app2.pdf]

References

1. Tobacco. World Health Organization. 2020. URL: https://www.who.int/news-room/fact-sheets/detail/tobacco [accessed 2020-11-02]
2. American Cancer Society. 2020. URL: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2020/cancer-facts-and-figures-2020.pdf [accessed 2020-07-06]
3. Schweizerische Gesundheitsbefragung 2017. Federal Statistical Office. 2018. URL: https://www.bfs.admin.ch/bfsstatic/dam/assets/6426300/master [accessed 2020-07-06]
4. Hughes JR, Keely J, Naud S. Shape of the relapse curve and long-term abstinence among untreated smokers. Addiction 2004 Jan;99(1):29-38. [doi: 10.1111/j.1360-0443.2004.00540.x] [Medline: 14678060]
5. Scharf D, Ferguson S, Tindle H, Shiffman S. Smoking cessation. In: Norcross JC, VandenBos GR, Freedheim DK, Pole N, editors. Psychopathology and Health. First Edition. Washington, DC: American Psychological Association; 2016:535-549.
6. Ferguson SG, Shiffman S. The relevance and treatment of cue-induced cravings in tobacco dependence. J Subst Abuse Treat 2009 Apr;36(3):235-243. [doi: 10.1016/j.jsat.2008.06.005] [Medline: 18715743]

7. Hendershot CS, Witkiewitz K, George WH, Marlatt GA. Relapse prevention for addictive behaviors. Subst Abuse Treat Prev Policy 2011;6(17):1-17 [FREE Full text] [doi: 10.1186/1747-597X-6-17] [Medline: 21771314]
8. Naughton F. Delivering "Just-In-Time" smoking cessation support via mobile phones: current knowledge and future directions. Nicotine Tob Res 2017 Mar;19(3):379-383. [doi: 10.1093/ntr/ntw143] [Medline: 27235703]
9. Westmaas JL, Bontemps-Jones J, Bauer JE. Social support in smoking cessation: reconciling theory and evidence. Nicotine Tob Res 2010 Jul 30;12(7):695-707. [doi: 10.1093/ntr/ntq077] [Medline: 20513695]
10. Smartphone-marktentwicklung des smartphone-marktes. Federal Communications Commission. URL: https://www.comcom.admin.ch/comcom/en/Homepage/documentation/facts-and-figures/mobile-telephony/smartphone-markt.html [accessed 2020-11-18]
11. Vilardaga R, Casellas-Pujol E, McClernon JF, Garrison KA. Mobile applications for the treatment of tobacco use and dependence. Curr Addict Rep 2019 Jun 9;6(2):86-97 [FREE Full text] [doi: 10.1007/s40429-019-00248-0] [Medline: 32010548]
12. McClure JB, Hartzler AL, Catz SL. Design considerations for smoking cessation apps: feedback from nicotine dependence treatment providers and smokers. JMIR Mhealth Uhealth 2016;4(1):e17 [FREE Full text] [doi: 10.2196/mhealth.5181] [Medline: 26872940]
13. Abroms LC, Lee WJ, Bontemps-Jones J, Ramani R, Mellerson J. A content analysis of popular smartphone apps for smoking cessation. Am J Prev Med 2013 Dec;45(6):732-736 [FREE Full text] [doi: 10.1016/j.amepre.2013.07.008] [Medline: 24237915]
14. Hoeppner BB, Hoeppner SS, Seaboyer L, Schick MR, Wu GW, Bergman BG, et al. How smart are smartphone apps for smoking cessation? A content analysis. Nicotine Tob Res 2016 May;18(5):1025-1031. [doi: 10.1093/ntr/ntv117] [Medline: 26045249]
15. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
16. Ubhi HK, Kotz D, Michie S, van Schayck OC, Sheard D, Selladurai A, et al. Comparative analysis of smoking cessation smartphone applications available in 2012 versus 2014. Addict Behav 2016 Jul;58:175-181 [FREE Full text] [doi: 10.1016/j.addbeh.2016.02.026] [Medline: 26950256]
17. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev 2019 Oct 22;10:CD006611. [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
18. Guerriero C, Cairns J, Roberts I, Rodgers A, Whittaker R, Free C. The cost-effectiveness of smoking cessation support delivered by mobile phone text messaging: Txt2stop. Eur J Health Econ 2013 Oct;14(5):789-797 [FREE Full text] [doi: 10.1007/s10198-012-0424-5] [Medline: 22961230]
19. Scott-Sheldon LA, Lantini R, Jennings EG, Thind H, Rosen RK, Salmoirago-Blotcher E, et al. Text messaging-based interventions for smoking cessation: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016;4(2):e49 [FREE Full text] [doi: 10.2196/mhealth.5436] [Medline: 27207211]
20. BinDhim NF, McGeechan K, Trevena L. Smartphone Smoking Cessation Application (SSC App) trial: a multicountry double-blind automated randomised controlled trial of a smoking cessation decision-aid 'app'. BMJ Open 2018 Dec 21;8(1):e017105 [FREE Full text] [doi: 10.1136/bmjopen-2017-017105] [Medline: 29358418]
21. West R, Hajek P, Stead L, Stapleton J. Outcome criteria in smoking cessation trials: proposal for a common standard. Addiction 2005 Mar;100(3):299-303. [doi: 10.1111/j.1360-0443.2004.00995.x] [Medline: 15733243]
22. Thoits PA. Mechanisms linking social ties and support to physical and mental health. J Health Soc Behav 2011 Jun;52(2):145-161. [doi: 10.1177/0022146510395592] [Medline: 21673143]
23. Scholz U, Stadler G, Ochsner S, Rackow P, Hornung R, Knoll N. Examining the relationship between daily changes in support and smoking around a self-set quit date. Health Psychol 2016 May;35(5):514-517. [doi: 10.1037/hea0000286] [Medline: 26462057]
24. Lancaster T, Stead LF. Individual behavioural counselling for smoking cessation. Cochrane Database Syst Rev 2017 Dec 31;3:CD001292. [doi: 10.1002/14651858.CD001292.pub3] [Medline: 28361496]
25. Stead LF, Carroll AJ, Lancaster T. Group behaviour therapy programmes for smoking cessation. Cochrane Database Syst Rev 2017 Dec 31;3:CD001007. [doi: 10.1002/14651858.CD001007.pub3] [Medline: 28361497]
26. Matkin W, Ordóñez-Mena JM, Hartmann-Boyce J. Telephone counselling for smoking cessation. Cochrane Database Syst Rev 2019 May 02;5:CD002850 [FREE Full text] [doi: 10.1002/14651858.CD002850.pub4] [Medline: 31045250]
27. Faseru B, Richter KP, Scheuermann TS, Park EW. Enhancing partner support to improve smoking cessation. Cochrane Database Syst Rev 2018 Aug 13;8:CD002928 [FREE Full text] [doi: 10.1002/14651858.CD002928.pub4] [Medline: 30101972]
28. Rafaeli E, Gleason M. Skilled support within intimate relationships. J Fam Theory Rev 2009;1(1):20-37. [doi: 10.1111/j.1756-2589.2009.00003.x]
29. May S, West R. Do social support interventions ("buddy systems") aid smoking cessation? A review. Tob Control 2000 Dec;9(4):415-422 [FREE Full text] [doi: 10.1136/tc.9.4.415] [Medline: 11106712]

30. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
31. Graham AL, Papandonatos GD, Zhao K. The failure to increase social support: it just might be time to stop intervening (and start rigorously observing). Transl Behav Med 2017 Dec 9;7(4):816-820 [FREE Full text] [doi: 10.1007/s13142-016-0458-9] [Medline: 28070778]
32. Lüscher J, Berli C, Schwaninger P, Scholz U. Smoking cessation with smartphone applications (SWAPP): study protocol for a randomized controlled trial. BMC Public Health 2019 Oct 29;19(1):1400 [FREE Full text] [doi: 10.1186/s12889-019-7723-z] [Medline: 31664959]
33. iCO Smokerlyzer®: User manual. Bedfont Scientific Ltd. URL: https://www.bedfont.com/documents/iCO-Smokerlyzer-manual.pdf [accessed 2020-09-14]
34. Ybarra ML, Jiang Y, Free C, Abroms LC, Whittaker R. Participant-level meta-analysis of mobile phone-based interventions for smoking cessation across different countries. Prev Med 2016 Aug;89:90-97 [FREE Full text] [doi: 10.1016/j.ypmed.2016.05.002] [Medline: 27154349]
35. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016;4:CD006611. [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
36. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. Br J Addict 1991 Sep;86(9):1119-1127. [doi: 10.1111/j.1360-0443.1991.tb01879.x] [Medline: 1932883]
37. Lüscher J, Stadler G, Scholz U. A daily diary study of joint quit attempts by dual-smoker couples: the role of received and provided social support. Nicotine Tob Res 2017 Dec 13;20(1):100-107. [doi: 10.1093/ntr/ntx079] [Medline: 28387852]
38. Wong HY, Subramaniyan M, Bullen C, Siddiq AN, Danaee M, Yee A. The mobile-phone-based iCO Smokerlyzer: comparison with the piCO Smokerlyzer among smokers undergoing methadone-maintained therapy. Tob Induc Dis 2019 Sep 9;17(September):65 [FREE Full text] [doi: 10.18332/tid/111355] [Medline: 31582954]
39. Erb P, Raiff BR, Meredith SE, Dallery J. The accuracy of a lower-cost breath carbon monoxide meter in distinguishing smokers from non-smokers. J Smok Cessat 2014 Jan 16;10(1):59-64. [doi: 10.1017/jsc.2013.37]
40. Hartmann-Boyce J, Chepkin SC, Ye W, Bullen C, Lancaster T. Nicotine replacement therapy versus control for smoking cessation. Cochrane Database Syst Rev 2018 Dec 31;5:CD000146. [doi: 10.1002/14651858.CD000146.pub5] [Medline: 29852054]
41. Bailey SR, Jeffery CJ, Hammer SA, Bryson SW, Killen DT, Ammerman S, et al. Assessing teen smoking patterns: the weekend phenomenon. Drug Alcohol Depend 2012 Jan 01;120(1-3):242-245 [FREE Full text] [doi: 10.1016/j.drugalcdep.2011.07.014] [Medline: 21885211]
42. Bolger N, Laurenceau JP. Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research. New York: The Guilford Press; 2013.
43. Xie H, Tao J, McHugo GJ, Drake RE. Comparing statistical methods for analyzing skewed longitudinal count data with many zeros: an example of smoking cessation. J Subst Abuse Treat 2013 Jul;45(1):99-108. [doi: 10.1016/j.jsat.2013.01.005] [Medline: 23453482]
44. Atkins DC, Baldwin SA, Zheng C, Gallop RJ, Neighbors C. A tutorial on count regression and zero-altered count models for longitudinal substance use data. Psychol Addict Behav 2013 Mar;27(1):166-177 [FREE Full text] [doi: 10.1037/a0029508] [Medline: 22905895]
45. Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: keep it maximal. J Mem Lang 2013 Apr;68(3):255-278 [FREE Full text] [doi: 10.1016/j.jml.2012.11.001] [Medline: 24403724]
46. Baskerville NB, Struik LL, Guindon GE, Norman CD, Whittaker R, Burns C, et al. Effect of a mobile phone intervention on quitting smoking in a young adult population of smokers: randomized controlled trial. JMIR Mhealth Uhealth 2018 Oct 23;6(10):e10893. [doi: 10.2196/10893] [Medline: 30355563]
47. Garrison KA, Pal P, O'Malley SS, Pittman BP, Gueorguieva R, Rojiani R, et al. Craving to Quit: A randomized controlled trial of smartphone app-based mindfulness training for smoking cessation. Nicotine Tob Res 2020 Mar 16;22(3):324-331. [doi: 10.1093/ntr/nty126] [Medline: 29917096]
48. Brooks KP, Schetter CD. Social negativity and health: conceptual and measurement issues. Social and Personality Psychology Compass 2011;5(11):904-918. [doi: 10.1111/j.1751-9004.2011.00395.x]
49. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
50. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
51. Paay J, Kjeldskov J, Skov M, Lichon L, Rasmussen S. Understanding individual differences for tailored smoking cessation apps. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015 Presented at:

CHI '15: CHI Conference on Human Factors in Computing Systems; April 18-23, 2015; Seoul Republic of Korea p. 1699-1708. [doi: 10.1145/2702123.2702321]
52. Graham AL, Papandonatos GD, Jacobs MA, Amato MS, Cha S, Cohn AM, et al. Optimizing text messages to promote engagement with internet smoking cessation treatment: results from a factorial screening experiment. J Med Internet Res 2020 Apr 02;22(4):e17734 [FREE Full text] [doi: 10.2196/17734] [Medline: 32238338]
53. McClure JB, Shortreed SM, Bogart A, Derry H, Riggs K, St JJ, et al. The effect of program design on engagement with an internet-based smoking intervention: randomized factorial trial. J Med Internet Res 2013;15(3):e69 [FREE Full text] [doi: 10.2196/jmir.2508] [Medline: 23529377]
54. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
55. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
56. Shiffman S. How many cigarettes did you smoke? Assessing cigarette consumption by global report, Time-Line Follow-Back, and ecological momentary assessment. Health Psychol 2009 Sep;28(5):519-526 [FREE Full text] [doi: 10.1037/a0015197] [Medline: 19751076]
57. Collins LM. Optimization of Behavioral, Biobehavioral, and Biomedical Interventions : The Multiphase Optimization Strategy (MOST). Switzerland: Springer; 2018.
58. Piper ME, Fiore MC, Smith SS, Fraser D, Bolt DM, Collins LM, et al. Identifying effective intervention components for smoking cessation: a factorial screening experiment. Addiction 2016 Jan;111(1):129-141 [FREE Full text] [doi: 10.1111/add.13162] [Medline: 26582269]
59. Bernstein SL, Dziura J, Weiss J, Miller T, Vickerman KA, Grau LE, et al. Tobacco dependence treatment in the emergency department: a randomized trial using the Multiphase Optimization Strategy. Contemp Clin Trials 2018 Mar;66:1-8 [FREE Full text] [doi: 10.1016/j.cct.2017.12.016] [Medline: 29287665]
60. Chib A, Lin SH. Theoretical advancements in mHealth: a systematic review of mobile apps. J Health Commun 2018;23(10-11):909-955 [FREE Full text] [doi: 10.1080/10810730.2018.1544676] [Medline: 30449261]

Abbreviations

BCT: behavior change technique
CG: control group
CO: carbon monoxide
CPD: cigarettes per day
GLMM: generalized linear mixed model
IG: intervention group
NRP: nicotine replacement product
ppm: parts per million
RCT: randomized controlled trial
RR: rate ratio

Edited by R Kukafka; submitted 14.01.21; peer-reviewed by J Brewer, C Spears; comments to author 02.03.21; revised version received 26.04.21; accepted 24.05.21; published 09.09.21.

Please cite as:
Schwaninger P, Berli C, Scholz U, Lüscher J
Effectiveness of a Dyadic Buddy App for Smoking Cessation: Randomized Controlled Trial
J Med Internet Res 2021;23(9):e27162
URL: https://www.jmir.org/2021/9/e27162
doi:10.2196/27162
PMID:34499045

Original Paper

Acceptability of a Brief Web-Based Theory-Based Intervention to Prevent and Reduce Self-harm: Mixed Methods Evaluation

Chris Keyworth[1], PhD; Rory O'Connor[2], PhD; Leah Quinlivan[3], PhD; Christopher J Armitage[3,4,5], PhD

[1]School of Psychology, University of Leeds, Leeds, United Kingdom

[2]Suicidal Behaviour Research Laboratory, Institute of Health and Wellbeing, University of Glasgow, Glasgow, United Kingdom

[3]NIHR Greater Manchester Patient Safety Translational Research Centre, The University of Manchester, Manchester, United Kingdom

[4]Manchester Centre for Health Psychology, The University of Manchester, Manchester, United Kingdom

[5]Manchester University NHS Foundation Trust, The University of Manchester, Manchester, United Kingdom

Corresponding Author:
Chris Keyworth, PhD
School of Psychology
University of Leeds
Woodhouse
Leeds, LS2 9JT
United Kingdom
Phone: 44 01133433738
Email: c.keyworth@leeds.ac.uk

Abstract

Background: The volitional help sheet (VHS) for self-harm equips people with a means of responding automatically to triggers for self-harm with coping strategies. Although there is some evidence of its efficacy, improving acceptability and making the intervention available in a web-based format may be crucial to increasing effectiveness and reach.

Objective: This study aims to use the Theoretical Framework of Acceptability (TFA) to explore the acceptability of the VHS, examine for whom and under what circumstances this intervention is more or less acceptable, and develop a series of recommendations for how the VHS can be used to support people in reducing repeat self-harm.

Methods: We explored acceptability in two phases. First, our patient and public involvement partners evaluated the original VHS from a lived experience perspective, which was subsequently translated into a web-based format. Second, a representative sample of adults in the United Kingdom who had previously self-harmed were recruited via a YouGov survey (N=514) and were asked to rate the acceptability of the VHS based on the seven constructs of the TFA, namely, *affective attitude*, *burden*, *perceived effectiveness*, *ethicality*, *intervention coherence*, *opportunity costs*, and *self-efficacy*. Data were analyzed using descriptive statistics, one-tailed *t* tests, and binary logistic regression. A directed content analysis approach was used to analyze qualitative data.

Results: Participants in the web-based survey rated the VHS as positive (*affective attitude*; t_{457}=4.72; *P*<.001); were confident using it (self-efficacy; t_{457}=9.54; *P*<.001); felt they did not have to give up any benefits, profits, or values when using it (*opportunity costs*; t_{439}=−15.51; *P*<.001); understood it and how it worked (*intervention coherence*; t_{464}=11.90; *P*<.001); and were confident that it would achieve its purpose (*perceived effectiveness*; t_{466}=2.04; *P*=.04). The TFA domain *burden* appeared to be an important indicator of acceptability. Lower levels of perceived burden when using the VHS tool were more prevalent among younger adults aged 18-24 years (OR 3.63, 95% CI 1.50-8.78), people of White ethnic background (OR 3.02, 95% CI 1.06-8.613), and people without a long-term health condition (OR 1.53, 95% CI 1.01-2.30). Perceived modifications to further improve acceptability included improved formatting (*burden*), the feature to add new situations and responses or amend existing ones (*ethicality*), and clearer instructions and further detail about the purpose of the VHS (*intervention coherence*).

Conclusions: Our findings show high levels of acceptability among some people who have previously self-harmed, particularly among younger adults, people of White ethnic backgrounds, and people without long-term health conditions. Future research should aim to improve acceptability among older adults, people from minority ethnic groups, and people with long-term health conditions.

(J Med Internet Res 2021;23(9):e28349) doi:10.2196/28349

KEYWORDS

self-harm; implementation intentions; acceptability; online; volitional help sheet; digital health; mental health

Introduction

Background

Self-harm is a major public health concern that has a major impact on health care services [1,2] and is growing in prevalence in the United Kingdom [3]. Consequently, the management of self-harm is a widely recognized challenge, and developing preventative strategies is vital [4,5]. Self-harm may include self-harm with suicidal intent (suicidal attempts), self-harm without suicidal intent (nonsuicidal self-harm [NSSH]), or suicidal thoughts [6].

There are many reasons why people engage in self-harm [7] but common among them are triggers, such as feelings of defeat or entrapment [8] that increase the urge to self-harm. Providing people with a means of responding to such critical situations may lessen the likelihood of self-harm. Implementation intentions [9] may be valuable in this regard, as they automatize coping responses to trigger critical situations.

Implementation intentions are *if-then* plans that help people to link a critical situation (ie, *if*) with an appropriate response (ie, *then*). There is a wide body of research illustrating the effectiveness of implementation intentions [10], yet little research has examined the effectiveness of implementing intention-based interventions for reducing repeat self-harm. To help people reduce repeated self-harm, implementation intentions can help people recognize when they feel the urge to self-harm and provide alternative coping strategies. The if-then plans work by making automatic links [10] in memory between a critical situation ("If I am tempted to self-harm when I want to get some attention...") and an appropriate response ("...then I will do something else instead of self-harming"). Implementation intentions have been shown to be effective in reducing self-harm in people recently admitted to the hospital for self-harm [11]. In this study, participants were provided with a tool, the volitional help sheet (VHS), designed to assist with the formation of implementation intentions to reduce self-harm. The tool provides people with a list of critical situations where the urge to self-harm may be heightened, and a list of responses designed to increase the likelihood of not self-harming [12]. The development of the VHS for self-harm has been described elsewhere [11]. Briefly, the VHS provides a theoretically driven framework for participants to construct their own implementation intentions, drawing on theories of suicidal behavior [13], self-harm motivation literature [14], and the transtheoretical model of change [15]. However, previous research is limited by high rates of attrition at follow-up [11]. Other studies report that the effectiveness of implementation intentions for reducing repeat self-harm may vary as a function of self-harm history [16]. Given the likelihood of successful implementation and effectiveness of interventions may be dependent upon perceptions of acceptability [17,18], it is necessary to comprehensively examine the acceptability of the VHS for self-harm in further detail.

Intervention acceptability is an important consideration in the design, implementation, and evaluation of health care interventions [19,20]. The likelihood of successful implementation and effectiveness may depend on perceptions of acceptability [17,20]. For example, interventions perceived as acceptable by those delivering or receiving them are more likely to result in favorable outcomes, including adherence to treatment programs [18], support for public health policy [17], or acceptance of behavior change interventions [21]. The Theoretical Framework of Acceptability (TFA) [20] is an established guide for assessing the acceptability of interventions. It defines acceptability as "a multifaceted construct that reflects the extent to which people delivering or receiving a health care intervention consider it to be appropriate, based on anticipated or experienced cognitive and emotional responses to the intervention" [20]. The TFA comprises seven domains: (1) affective attitude (how individuals feel about taking part in an intervention), (2) burden (the amount of effort required to engage with an intervention), (3) perceived effectiveness (whether individuals perceive an intervention as likely to achieve its purpose), (4) ethicality (the extent to which an intervention fits with individuals' personal values), (5) intervention coherence (whether individuals understand an intervention and how it works), (6) opportunity costs (what is given up, such as time, to take part in an intervention), and (7) self-efficacy (how confident individuals are performing the intervention). The advantage of using the TFA, as opposed to more general approaches to investigating acceptability, is that the TFA allows a more systematic assessment of intervention acceptability, and this approach allows researchers to target specific TFA domains in future iterations of interventions (eg, addressing perceived burden of interventions) [22].

Objective

This study has three specific objectives: first, to evaluate the acceptability of the VHS from a lived experience perspective (patient and public involvement [PPI]) and redevelop the tool according to feedback; second, to examine for whom and under what circumstances the VHS is more or less acceptable; and finally, to develop a series of practice recommendations for how the VHS can be used to support people in reducing repeat self-harm.

Methods

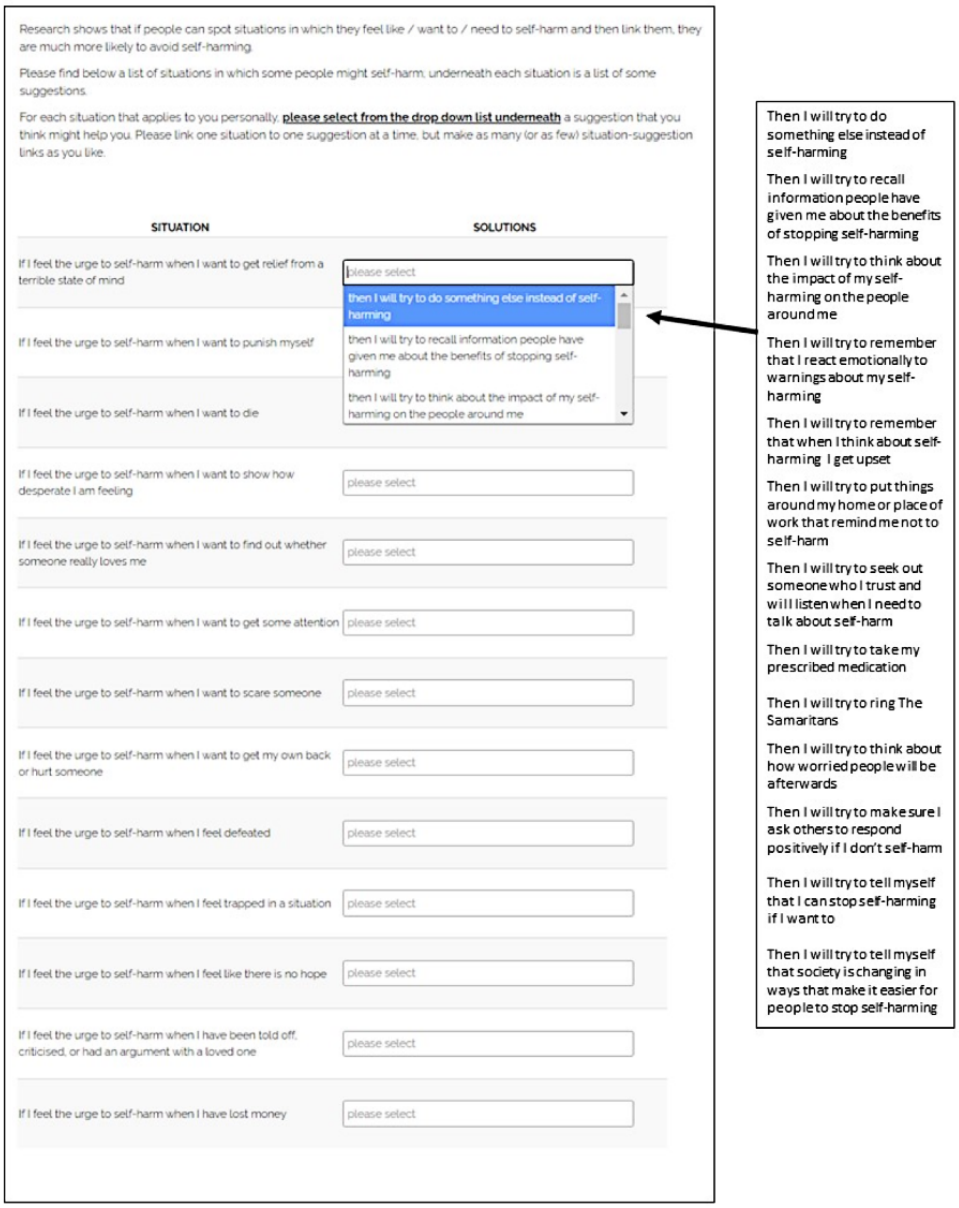
Overview

Ethical approval was obtained from the University Research Ethics Committee (ref: 2020-8446-15312), and informed consent was obtained from all participants. The VHS provides people with a list of critical situations where the urge to self-harm may be heightened, and a list of coping responses designed to decrease the likelihood of self-harming [12]. The development of the VHS for self-harm has been described elsewhere [11]. Briefly, the VHS provides a theoretically driven framework for participants to construct their own implementation intentions, drawing on theories of suicidal behavior [13], self-harm

motivation literature [14], and the transtheoretical model of change [15].

Acceptability was explored in two phases. In phase 1, the original VHS, used as part of a previous study [11], was distributed among a PPI group for initial feedback and translated to a web-based format by the research team. This was done by creating on a single webpage, as part of a web-based questionnaire, a list of situations alongside which participants could choose an appropriate response from a drop-down menu for each critical situation (a screenshot of the VHS is provided for illustrative purposes in Figure 1). As the intervention was presented on a single screen within the questionnaire, participants were able to print the VHS after participation. Participants were also given a physical copy of the VHS as part of the study. In phase 2, a national sample of people in the United Kingdom who had previously self-harmed was recruited via a survey panel company (YouGov), as part of a larger 6-month follow-up study examining the effectiveness of the VHS for reducing self-harm (ClinicalTrials.gov identifier: NCT04420546).

Figure 1. The volitional help sheet for self-harm.



Phase 1: PPI and Tool Development

The VHS has been shown to be effective for reducing self-harm in people recently admitted to the hospital following an episode of self-harm [11]. The researchers initially made minor changes to the VHS for translation into a web-based format, in line with a previous VHS delivered on the web [23]. To ensure that the intervention was suitable for as broad a population as possible, the VHS was evaluated from a lived experience perspective by members from a PPI group (n=10). The group was specifically trained to provide feedback on research methods and materials to be used as part of intervention delivery, and all members of the group had a history of self-harm, suicidal behavior, or receiving mental health services. The feedback from our PPI contributors was used to ensure that the VHS was understandable and sensitive to people's thoughts and emotions about self-harm (the majority of the PPI group had a history of self-harm). Participants provided (on hard copies of the VHS)

feedback and suggestions for improvements both on the wording of the instructions of the VHS, and the *critical situations* and *appropriate responses*. Previous deployments of the VHS have predominantly used paper and pencil, whereby participants are asked to physically draw a line linking any situations that applied to them, to one solution at a time (participants could make as many situation-solution links as they would like). On the basis of the feedback, amendments were made to the wording of 11 of the situations and nine of the solutions. Two solutions were removed because the group felt the wording was inappropriate ("Then I will tell myself that I can stop self-harming if I want to" and "Then I will make sure I am rewarded by others if I don't self-harm"). Two new situations ("If I feel the urge to self-harm when I have been told off, criticised, or had an argument with a loved one" and "If I feel the urge to self-harm when I have lost money"), and four new solutions ("Then I will try to ring The Samaritans," "Then I will try to think about how worried people will be afterwards," "Then I will try to make sure I ask others to respond positively if I don't self-harm," and "Then I will try to tell myself that I can stop self-harming if I want to") were added to the VHS. The final VHS, which included the two new situations and four new solutions, contained 13 situations and 13 solutions.

Phase 2: YouGov Survey

A national sample of adults in the United Kingdom who had previously self-harmed were recruited via a survey panel company (YouGov), as part of a larger study (ClinicalTrials.gov Identifier: NCT04420546) were invited to take part in a web-based questionnaire and were incentivized in accordance with the YouGov's point system (respondents accumulated points for taking part in surveys, which can be exchanged for cash or entry into a prize draw). To ensure that the final sample contained people with a prior history of self-harm, we asked a screen question: "Have you ever intentionally hurt yourself/self-harmed?" Response options were "yes, I have," "no, I have not," or "prefer not to say." The final sample was based on respondents answering, "yes, I have."

Following completion of the questionnaire (described below), participants were asked to complete the amended VHS to reduce self-harm (based on the feedback obtained from phase 1). Participants formed implementation intentions by linking critical situations with appropriate responses by choosing an appropriate response from a drop-down menu for each critical situation. Participants were free to make as many situation-response links as desired. Participants were then asked questions about their acceptability (n=514).

Measures

Sociodemographic Variables

Demographic variables, including age, gender, ethnicity, and social grade, were taken using the standard UK Office for National Statistics [24] measures.

History of NSSH, Suicidal Ideation, and Suicide Attempts

Three items drawn from the British Psychiatric Morbidity Survey [25]: "Have you ever seriously thought of taking your life, but not actually attempted to do so?" (suicidal ideation); "Have you ever made an attempt to take your life, by taking an overdose of tablets or in some other way?" (suicidal attempt); and "Have you ever deliberately harmed yourself in any way, but not with the intention of killing yourself? (ie, self-harm)" (NSSH). Response options for all questions were *Yes*, *No*, or *prefer not to say*. If respondents answered yes to any of the three questions, the timing of the last episode and frequency was asked.

Exposure to Death and Suicide

Participants were asked seven items [26,27] to establish whether any of their close friends or family had died, whether they had friends of family who had self-harmed, or who attempted or died by suicide (eg, "Has anyone among your family attempted suicide?").

Acceptability Measures

Only one study has deployed the TFA as a quantitative measure; Renko et al [28] assessed the acceptability of a training program to enhance teachers' physical activity promotion. However, the items were not sufficiently generic to adapt to the present purposes. Likert scale responses, with additional open-ended text questions to capture qualitative data with respect to each domain, were developed in line with the seven TFA constructs used to assess acceptability. Seven items were developed, for example, "On a scale of 0-10, how much effort was required to use the volitional help sheet?" (burden; *no effort*: 1 to *lots of effort*: 10). Item wordings were developed to closely resemble the definitions provided for each domain of the TFA [20]. The items used to measure each TFA domain are shown in Table 1.

For each of the seven TFA items, participants were invited to provide open-ended comments describing their experiences using the VHS with respect to each of the seven constructs of the TFA. Seven items were developed, corresponding to each of the TFA domains, for example, "Do you have any specific comments about how good or bad you felt when using the volitional help sheet?" (affective attitude). Participants' comments were combined, and content analysis was performed.

Table 1. The Theoretical Framework of Acceptability domains.

Domain	Question	Participant (n=514), n (%)	Values, mean (SD)	Comments (n=360), n (%)
Affective attitude	On a scale of 0-10, how good or bad did you feel when using the volitional help sheet?	476 (92.6)	5.50 (2.32)	33 (9.2)
Burden	On a scale of 0-10, how much effort was required to use the volitional help sheet?	471 (91.6)	5.09 (2.84)	78 (21.7)
Ethicality	On a scale of 0-10, how much was using the volitional help sheet a good fit with your personal values?	465 (90.5)	4.80 (2.68)	69 (19.2)
Self-efficacy	On a scale of 0-10, how confident were you about using the volitional help sheet?	468 (91.1)	6.16 (2.63)	37 (10.3)
Opportunity	On a scale of 0-10, to what extent did you give up any benefits, profits, or values when using the volitional help sheet?	440 (85.6)	2.95 (2.77)	38 (10.6)
Intervention coherence	On a scale of 0-10, how confident were you that you understood the volitional help sheet and how it works?	465 (90.5)	6.45 (2.63)	40 (11.1)
Perceived effectiveness	On a scale of 0-10, how confident were you that the volitional help sheet is likely to achieve its purpose?	467 (90.9)	5.24 (2.59)	65 (18.1)

Analyses

Quantitative data were analyzed using SPSS, version 26 (IBM Corporation). Descriptive statistics were used to summarize the sociodemographic variables. The chi-square test was used to compare our sample of people who reported a previous history of self-harm with general population data collected as part of the Adult Psychiatric Morbidity Survey [29]. In the absence of universally agreed criteria for determining acceptability with respect to deploying TFA quantitative measures, two approaches used in previous studies were adopted [30]. First, one-tailed *t* tests were used to assess how far above or below the neutral rating (5) people rated the seven TFA indicators. This measure was used to ensure that the neutral responses were accounted for. Second, the proportion of participants scoring at each point on the rating scale was assessed. These measures were used to assess the desirability and acceptability of the questionnaire according to each of the seven TFA domains. Binary logistic regression was used to explore the correlates of acceptability to identify which sociodemographic factors (gender, age, ethnicity, social grade, and recency of self-harm [past week or past year]; all dummy coded) were associated with higher or lower levels of perceived acceptability. Each of the main outcomes was recorded as a binary outcome (eg, high acceptability: 1 or low acceptability: 0). High acceptability was defined as scores above the neutral rating (ie, scores >5), and low acceptability was defined as scores ≤5, except for two domains (*burden* and *ethicality*) where high acceptability was defined as scores <5, and low acceptability was defined as scores ≥5.

Qualitative Analyses of Open‐Ended Comments

Participants were asked to rate the acceptability of the VHS based on the seven constructs of the TFA (described above): affective attitude, burden, perceived effectiveness, ethicality, intervention coherence, opportunity costs, and self-efficacy. Participants were invited to provide open-ended comments to each TFA question. A directed content analysis approach, which is suitable when the research uses an existing theory or framework to interpret the data, was used to identify and categorize instances of the TFA domains [31,32]. First, deductive coding was used to generate a content analysis framework in line with the TFA domains. Second, inductive coding was used to generate explanatory themes with respect to each TFA domain, whereby specific codes within each TFA domain were grouped into themes. Initial codes were generated and collated into potential themes by CK, who shared the coding framework and key illustrative quotes with CJA as the analysis progressed. Any areas of contention were discussed, and themes were refined accordingly to ensure the trustworthiness of the data. All the authors were involved in finalizing the main themes. NVivo, version 12 (QSR International) was used to organize the data. The codes focused on different aspects of acceptability with respect to using the VHS to reduce repeat self-harm, according to each TFA domain (eg, attitudes toward the intervention, and the perceived effort required to engage with the intervention). The themes were reviewed by coauthors, and there were no disagreements.

Results

Sample Characteristics

Overview

The total sample (n=514) comprised mostly women (331/514, 64.4%), 27.4% (141/514) were aged 18-34 years, 21.2% (109/514) were aged 35-44 years, 18.1% (93/514) were aged 45-54 years, and 33.3% (171/514) were aged ≥55 years. The majority of the sample was White (472/514, 91.8%), and 63.4% (326/514) were of higher social grade (nonmanual workers; Table 2). The characteristics of our sample closely resembled the characteristics of people who reported a history of self-harm according to the Adult Psychiatric Morbidity Survey of the general population [29] in terms of gender and age. However, our sample contained a lower proportion of people from a minority ethnic background compared with the national data.

Table 2. Sample demographics (N=514).

Variable	Population, n (%)	Value, mean (SD; range)	General population data (%)[a]	Chi-square difference between sample and population	*P* value
Gender	N/A	N/A[b]	N/A	N/A	N/A
Women	331 (64.4)		54.5	1.68	.19
Men	176 (34.2)		45.5	3.00	.08
Other or prefer not to say	7 (1.4)		N/A	N/A	N/A
Age (years)	N/A	45.80 (14.21; 18-77)	N/A	N/A	N/A
18-34	141 (27.4)		26.4	0.03	.87
35-44	109 (21.2)		17.8	0.28	.59
45-54	93 (18.1)		21.1	0.74	.39
≥55	171 (33.3)		34.6	0.09	.77
Ethnicity	N/A	N/A	N/A	N/A	N/A
White	472 (91.8)		87.1	1.33	.25
BAME[c]	16 (3.1)		12.9	6.79	.01
Prefer not to say	26 (5.1)		N/A	N/A	N/A
Social grade	N/A	N/A	N/A	N/A	N/A
Nonmanual worker	326 (63.4)		N/A	N/A	N/A
Manual or unemployed	188 (36.6)		N/A	N/A	N/A
Suicidal ideation (ever)	390 (75.9)	N/A	20.6	60.55	<.001
Past week	38 (7.4)		N/A	N/A	N/A
Past year	126 (24.5)		5.4	15.69	<.001
Longer ago	221 (43)		N/A	N/A	N/A
Would rather not say or did not answer	129 (25.1)		N/A	N/A	N/A
Suicidal attempt (ever)	212 (41.2)	N/A	6.7	31.69	<.001
Past week	N/A		N/A	N/A	N/A
Past year	N/A		0.7	1.85	.17
Longer ago	187 (36.4)		N/A	N/A	N/A
Would rather not say or did not answer	303 (59)		N/A	N/A	N/A
NSSH[d] (ever)	383 (74.5)	N/A	7.3	95.58	<.001
Past week	26 (5.1)		N/A	N/A	N/A
Past year	68 (13.2)		N/A	N/A	N/A
Longer ago	284 (55.3)		N/A	N/A	N/A
Would rather not say or did not answer	136 (26.5)		N/A	N/A	N/A
Exposure to suicide and death	N/A	N/A	N/A	N/A	N/A
Exposure to death (immediate family)	273 (53.1)		N/A	N/A	N/A
Exposure to death (close friend or relative)	390 (75.9)		N/A	N/A	N/A
Exposure to death by suicide (family or close friend)	159 (30.9)		N/A	N/A	N/A
Suicidal attempt (in the family)	175 (34)		N/A	N/A	N/A
Suicidal attempt (by close friends)	189 (36.8)		N/A	N/A	N/A
NSSH (in the family)	153 (29.8)		N/A	N/A	N/A
NSSH (by close friends)	214 (41.6)		N/A	N/A	N/A

[a]According to the Adult Psychiatric Morbidity Survey of the general population [29].

[b]N/A: not applicable.

[c]BAME: Black, Asian, and minority ethnic.

[d]NSSH: nonsuicidal self-harm.

Prevalence of Suicidal Ideation, Suicide Attempts, NSSH, and Exposure to Suicide and Death

Overall, 75.9% (390/514), 41.2% (212/514), and 74.5% (383/514) of the total sample reported suicidal ideation, suicide attempts, and NSSH, respectively (Table 1). Furthermore, 7.4% (38/514) of the total sample reported suicidal thoughts in the past week, and 24.5% (126/514) of the sample reported suicidal thoughts in the past year. Few people reported suicide attempts in the past week (4/514, 0.8%), and 3.9% (20/514) reported a suicide attempt in the past year. With respect to NSSH, 5.1% (26/514) reported NSSH in the past week, and 13.2% (68/514) reported NSSH in the past year.

Over half of the sample (273/514, 53.1%) reported experiencing the death of a family member, over three-fourths (390/514, 75.9%) of the sample reported experience of the death of a close friend or relative, and 30.9% (159/514) of the sample reported experience of death by suicide of a close friend or relative. Of the total sample, 34% (175/514) reported exposure to a family member making a suicide attempt, and 36.8% (189/514) reported exposure to a suicide attempt by a close friend. Exposure to NSSH by a family member was reported by 29.8% (153/514) of the sample, and NSSH by a close friend by 41.6% (214/514) of the sample.

With regard to lifetime history of self-harm, our sample reported a higher prevalence of suicidal ideation (390/514, 75.9% vs 20.6%), suicide attempts (212/514, 41.2% vs 6.7%), and NSSH (383/514, 74.5% vs 7.3%) compared with national data. With regard to self-harm in the previous year, our sample reported a higher prevalence of suicidal ideation (126/514, 24.5% vs 5.4%) compared with national data; the prevalence of suicide attempts in the previous year more closely resembled national data (0.7% vs 20/514, 3.9%).

Overall Acceptability of the VHS

One-sample *t* tests showed that participants rated the VHS favorably on five of the seven indicators by scoring above or below the respective midpoints. Participants rated the VHS positively (*affective attitude*; mean 5.50, SD 2.32; t_{457}=4.72; *P*<.001); were confident using the VHS (*self-efficacy*; mean 6.16, SD 2.63; t_{457}=9.54; *P*<.001); did not have to give up any benefits, profits, or values when using the VHS (*opportunity costs*; mean 2.95, SD 2.77; t_{439}=−15.51; *P*<.001); understood the VHS and how it worked (*intervention coherence*; mean 6.45, SD 2.63; t_{464}=11.90; *P*<.001); and were confident that it would achieve its purpose (*perceived effectiveness*; mean 5.24, SD 2.59; t_{466}=2.04; *P*=.04). No significant differences were found for the two TFA domains: *burden* (mean 5.09, SD 2.84) and *ethicality* (mean 4.80, SD 2.68). The mean ratings for each of the seven constructs are listed in Table 2.

The proportion of participants scoring at each point on the rating scale of each TFA item is presented in Multimedia Appendix 1. There were two key findings. First, there was a high proportion of responses at the upper end of the *self-efficacy* and *intervention coherence* items, compared with other items (97/468, 20.7% and 110/465, 23.6% of participants, respectively, rating 9 or 10 on the 0-10 scales). Second, there was a high proportion of responses at the lower end of the opportunity costs item compared with other items (186/440, 42.3% of participants rated 0 or 1 on the 0-10 scale).

Associations Between Sociodemographic Variables (Age, Gender, Ethnicity, and Current Health Status) and Acceptability

Table 3 shows the binary logistic regression results of perceived acceptability of the VHS according to sociodemographic variables.

No significant differences were found between men and women in perceived acceptability of any of the TFA variables. Lower levels of perceived burden were more prevalent among people aged 18-24 years, compared with those aged 25-34 years (OR 3.63, 95% CI 1.50-8.78), and 35-44 years (OR 2.55, 95% CI 1.06-6.15), and among people of White ethnic background, compared with people from minority ethnic groups (OR 3.02, 95% CI 1.06-8.61). Higher levels of perceived burden were more prevalent among people who reported having a long-term health condition (OR 1.53, 95% CI 1.01-2.30), compared with those who did not report having a long-term health condition.

Table 3. Associations between sociodemographic variables and acceptability of the volitional help sheet, according to Theoretical Framework of Acceptability variables.

Variables	β (95% CI)						
	Attitude	Burden	Ethicality	Self-efficacy	Opportunity costs	Intervention coherence	Perceived effectiveness
Gender (women)	1.10 (0.75-1.61)	1.17 (0.80-1.71)	1.26 (0.84-1.87)	1.28 (0.87-1.88)	0.91 (0.53-1.58)	1.17 (0.79-1.74)	1.26 (0.86-1.86)
Age (years; reference group: 18-24)							
25-34	0.81 (0.35-1.87)	3.63[a] *(1.50-8.78)*	0.83 (0.36-1.93)	1.51 (0.66-3.45)	0.64 (0.21-2.02)	1.55 (0.67-3.60)	0.82 (0.36-1.87)
35-44	0.76 (0.33-1.78)	2.56[b] *(1.06-6.15)*	0.80 (0.34-1.85)	1.72 (0.75-3.96)	0.64 (0.20-2.00)	1.70 (0.73-3.96)	0.78 (0.34-1.76)
45-54	1.11 (0.48-2.59)	1.70 (0.69-4.14)	0.73 (0.31-1.73)	1.58 (0.68-3.66)	1.04 (0.34-3.15)	1.31 (0.56-3.07)	0.71 (0.31-1.64)
55 or over	1.15 (0.51-2.57)	1.47 (0.63-3.43)	0.86 (0.38-1.94)	1.13 (0.51-2.50)	0.73 (0.25-2.16)	1.17 (0.53-2.61)	0.70 (0.32-1.53)
Ethnicity (minority ethnic groups)	1.86 (0.72-4.80)	3.02[b] *(1.06-8.61)*	1.56 (0.59-4.14)	0.77 (0.30-1.98)	2.88 (0.96-8.60)	3.05 (0.87-10.70)	1.51 (0.55-4.16)
Social grade (nonmanual)	0.93 (0.63-1.35)	1.01 (0.70-1.47)	0.92 (0.62-1.35)	1.42 (0.97-2.08)	1.05 (0.61-1.82)	1.21 (0.82-1.78)	1.08 (0.74-1.58)
Long-term chronic health condition	1.16 (0.77-1.74)	1.53[b] *(1.01-2.30)*	1.29 (0.84-1.98)	1.02 (0.68-1.53)	0.86 (0.49-1.53)	1.10 (0.73-1.67)	0.93 (0.62-1.39)
Marginalized group (yes)	0.75 (0.50-1.11)	1.20 (0.82-1.77)	0.93 (0.62-1.39)	0.90 (0.61-1.33)	1.26 (0.73-2.18)	1.36 (0.90-2.04)	1.16 (0.79-1.72)
Self-harm (past week; any measure)	0.66 (0.37-1.20)	1.55 (0.88-2.72)	1.34 (0.76-2.36)	1.22 (0.69-2.16)	0.72 (0.53-2.48)	0.91 (0.52-1.62)	0.84 (0.54-1.65)
Self-harm (past year; any measure)	0.71 (0.48-1.06)	1.64[b] *(1.11-2.41)*	0.82 (0.55-1.23)	0.73 (0.50-1.07)	0.78 (0.44-1.38)	1.08 (0.72-1.60)	0.86 (0.58-1.26)
Exposure to death (immediate family)	1.23 (0.85-1.77)	0.62[a] *(0.43-0.89)*	1.06 (0.73-1.55)	1.02 (0.71-1.48)	1.64 (0.95-2.84)	1.01 (0.70-1.48)	0.93 (0.64-1.33)
Exposure to death (close friend or relative)	1.15 (0.75-1.77)	0.86 (0.57-1.32)	1.11 (0.71-1.73)	0.91 (0.59-1.40)	0.94 (0.51-1.74)	1.25 (0.81-1.93)	0.86 (0.56-1.32)
Exposure to death by suicide (family or close friend)	0.80 (0.54-1.19)	0.97 (0.66-1.44)	1.07 (0.71-1.61)	0.73 (0.49-1.08)	1.06 (0.60-1.88)	0.97 (0.64-1.45)	1.11 (0.74-1.65)
Suicidal attempt (in the family)	*1.53*[b] *(1.04-2.25)*	0.95 (0.65-1.39)	1.02 (0.68-1.51)	1.11 (0.75-1.63)	1.32 (0.76-2.27)	1.23 (0.82-1.83)	1.11 (0.75-1.63)
Suicidal attempt (by close friends)	0.87 (0.60-1.28)	1.11 (0.76-1.62)	0.91 (0.61-1.35)	1.20 (0.81-1.76)	1.77[b] *(1.03-3.03)*	1.73 (0.95-2.14)	1.00 (0.68-1.47)
NSSH[c] (in the family)	1.13 (0.76-1.69)	1.00 (0.67-1.49)	1.53[b] *(1.02-2.30)*	1.51[b] *(1.00-2.27)*	1.04 (0.58-1.85)	2.10[a] *(1.35-3.27)*	1.33 (0.89-1.99)
NSSH (by close friends)	0.78 (0.53-1.14)	1.33 (0.92-1.94)	1.10 (0.74-1.61)	1.33 (0.91-1.95)	1.30 (0.75-2.24)	1.53[b] *(1.03-2.27)*	1.02 (0.70-1.48)

[a]*P*<.05.

[b]*P*<.01.

[c]NSSH: nonsuicidal self-harm.

Associations Among Prevalence of Suicidal Ideation, Suicide Attempts, NSSH, Exposure to Suicide and Death, and Acceptability

People who reported self-harm in the past year were more likely to report higher levels of burden (OR 1.64, 95% CI 1.11-2.41) than those who did not report self-harm in the past year. No significant differences were found for any of the TFA variables among people who had or had not self-harmed in the past week.

People who reported exposure to death (within their immediate family) were likely to perceive a lower burden (OR 0.62, 95% CI 0.43-0.89) compared with those who had not. People who reported exposure to a suicide attempt (within their immediate family) were more likely to report favorable attitudes toward the VHS (OR 1.53, 95% CI 1.04-2.25). People who reported exposure to a suicide attempt (by close friends) were more likely to report higher opportunity costs (the extent to which people gave up any benefits, profits, or values when using the VHS;

OR 1.77, 95% CI 1.03-3.03) compared with those who did not. People who reported NSSH (within their immediate family) were more likely to report higher levels of ethicality (fit with personal values; OR 1.53, 95% CI 1.02-2.30), self-efficacy (OR 1.51, 95% CI 1.00-2.27), and intervention coherence (confidence in understanding the intervention and how it works; OR 2.10, 95% CI 1.35-3.27), compared with those who did not. People who reported NSSH (by close friends) were more likely to report higher levels of intervention coherence (OR 1.53, 95% CI 1.03-2.27), compared with those who did not.

Qualitative Analyses of Open-Ended Comments

Overview

In all, 340 participants provided at least one open-text comment in response to the TFA items. After responses that were deemed invalid (eg, participants responding with *no comments*) were removed, all of the remaining responses (n=360 comments) were analyzed. Explanatory themes with illustrative quotes were presented within each TFA domain. With respect to opportunity costs, 38 comments were provided. The majority of comments were related to a lack of clarity with respect to this question. Consequently, the participants were unable to provide responses for this domain.

Affective Attitude

In total, 9.2% (33/360) of the comments were provided with respect to how good or bad participants felt about taking part in the intervention. Comments within this domain focused on *refinements to the intervention* and *the perceived relevance of situations and responses*. First, the suggested refinements were mainly with respect to general comments about the wording of the situations and responses. In particular, some participants suggested that some of the statements might induce negative thoughts about self-harm and act as a reminder of personal experiences of self-harm:

> *It brought back a lot of negative memories and feeling surrounding self-harm. Someone would need to be in a decent place in order to use it.* [P552, female]

Second, participants made reference to the importance of the intervention containing relevant situations to their own experiences where the urge to self-harm may be heightened and responses that are more appropriate. Some participants expressed a desire for more relevant options, as some participants had difficulty relating to either the situations or responses, and they did not link directly to their experiences:

> *I felt slightly frustrated. I don't know why. Maybe because I wanted more options.* [P348, male]

> *I felt that many of those situations were too specific and did not fit my experiences.* [P57, female]

Burden

In total, 21.6% (78/360) of the comments were provided with respect to the perceived effort required to use the VHS. Comments within this domain focused on *perceptions of usability* and *perceived technology-based challenges*. First, some participants reported specific features of the VHS that made it easy to use, including encouraging and promoting self-reflection as an important feature of the intervention. However, 13 participants reported the psychological effort required to engage with the VHS, such as difficulties in choosing an appropriate situation and response. Participants also described how situations or responses that were perceived to be not personally relevant added to the effort required to engage with the intervention: "Too much effort may result in people giving up completing it."

Second, participants described potential challenges of using a web-based platform that should be considered in future iterations of the intervention. First, due to the intervention being delivered on the web, 11 participants expressed technical or formatting difficulties that made it difficult to engage with the VHS. Technical issues included difficulties viewing on a mobile device or difficulties in selecting situations and responses from the drop-down menus. These were deemed minor and could be resolved in future iterations of the intervention:

> *I think it was just this interface made it difficult to scroll. I'm sure if it was on a different site it would be fine.* [P544, female]

> *I am doing it on the phone so is difficult to read sometimes.* [P339, male]

Ethicality

In total, 19.2% (69/360) of the comments were provided with respect to the extent to which using the VHS a good fit with an individual's personal values. Comments within this domain focused on *the perceived strengths of the intervention* and the *perceived importance of having modifiable statements*. First, participants reported that they recognized the situations and responses and indicated a good fit with their personal values. One participant reported that the practical approach to trying to reduce self-harm was a particular strength of the intervention. Another participant described how the intervention helped them focus on more positive thoughts when forming implementation intentions:

> *It identified to me other ways that I could keep thoughts positive.* [P571, female]

> *Practical without being patronising or guilt inducing.* [P348, male]

Second, there were a small number of participants, who responded negatively to the VHS in its current form; 15 participants perceived the situations and responses as a poor fit with their personal values (such as some of the solutions focusing on other people). In total, 6 respondents suggested acknowledging other situations and responses would strengthen the intervention, as would the feature to add new situations and responses or amend existing ones:

> *It makes the assumption that self harming is always bad.* [P437, female]

> *Like I said it makes me think I should be guilty for wanting to self harm.* [P101, female]

> *I don't really feel comfortable using other people as a solution I think examples of what to do instead (eg go for a walk) could be a way in the moment to realise thinks that might help.* [P709, female]

> *Self harming for attention/to hurt or worry others is attention seeking. There are other ways to seek attention. My own self harm was entirely hidden.* [P417, male]
>
> *I'd prefer something a bit more free-wheeling where I could add my own responses.* [P254, female]

Intervention Coherence

In total, 11.1% (40/360) of the comments were provided with respect to how confident respondents were that they understood the VHS and how it worked. Comments within this domain focused on *perceived clarity about the purpose of the intervention* and *perceived confidence in engaging with the intervention correctly.* First, 9 participants reported that they understood the intervention and were positive about its usefulness and helpfulness. However, the majority of the comments related to the need for improved clarity about the purpose of the VHS and how it works:

> *I believe I'm emotionally fairly intelligent and quite reflective so I believe I understood it fairly well.*
>
> *It was easy to understand but I imagine that some people would think they could only choose one option.* [P427, male]

Second, for some participants, a lack of clarity in how the VHS was described affected their confidence in using the intervention in the correct way. A need for clearer instructions for the VHS and, in particular, to help describe how the VHS is intended to work as a way of developing coping plans, was expressed:

> *The situation and the options were there but no idea how it is supposed to work to stop something that is not tailored towards a specific problem.* [P22, male]
>
> *I wasn't sure if I should have read all the options even the situations were not applicable to me. I apologise if I got that wrong.* [P387, male]

Perceived Effectiveness

In total, 18.1% (65/360) of the comments were provided with respect to how confident respondents were that the VHS is likely to achieve its purpose. Comments within this domain focused on *perceptions of when the intervention could be used* and *perceived receptiveness of people*. First, participants described how the VHS would be particularly helpful for people in specific contexts or under particular circumstances. Some participants believed that it may be helpful for people when the urge to self-harm was not at its height or as a preventative measure for self-harm: "I think in milder moments and before crisis point, it could help refocus the mind to a more positive and less destructive way of thinking."

Second, participants believed that the effectiveness of the VHS might depend on how receptive people are to the intervention, and they believed it would be effective for their own circumstances. Some participants expressed concerns that the purpose of the VHS was unclear and expressed a need for further guidance. Participants described how the intervention may induce negative emotions for some people, which may affect how people engage with the intervention:

> *I think for someone who self harms on instinct it would be useful to have solutions written down as a reminder.* [P249, female]
>
> *It requires a level of concentration which may be difficult to achieve if severely depressed.* [P422, male]

Self-efficacy

In total, 10.3% (37/360) of the comments were provided with respect to how confident respondents were about using the VHS. Comments within this domain focused on *perceptions of personal relevance* and *factors associated with engagement*. First, participants reported that the intervention was easy to use and navigate. However, others have reported that the likelihood of people using or needing a VHS may be dependent upon the relevance of the situations and solutions:

> *I felt like I learnt to use it. In some cases the situations and options did not fit "me."* [P549, male]

Perceptions of personal relevance were perceived to affect confidence in using the intervention. In some cases, participants reported that they had lower confidence in using the VHS because of some of the options not being relevant to their own circumstances; 3 participants expressed doubts that they were using the VHS correctly. Furthermore, one participant also described how motivation to engage with the intervention may differ depending on the situation presented:

> *Not sure on some of the questions, so put nearest guess to what I might do, most likely I would have selected a different option not available.* [P22, male]
>
> *Wasn't 100% sure about whether or not I could add the same motivation to different situations (I didn't).* [P254, female]

Discussion

Principal Findings

This paper describes the acceptability of a brief intervention based on implementation intentions to prevent and reduce self-harm. This is the first study to (1) apply the TFA to examine the acceptability of preventative strategies for self-harm, (2) examine for whom and under what circumstances the VHS is more or less acceptable for preventing and reducing self-harm, and (3) develop a series of recommendations for how the VHS can be used to support people to reduce repeated self-harm. This study has four key findings. First, the VHS was rated favorably on five of the seven TFA domains (*affective attitude*, *self-efficacy*, *opportunity costs*, *intervention coherence*, and *perceived effectiveness*). In particular, there was a high proportion of responses at the upper end of the *self-efficacy* and *intervention coherence* items, which suggests that participants understood the VHS and how it works, and they were confident in their ability to use it correctly. Furthermore, there was a high proportion of responses at the lower end of the opportunity cost item, which suggests that people anticipated no problems forming implementation intentions, nor did they have to give up something else, such as time, to engage with the intervention.

Second, with respect to sociodemographic variables, the TFA domain *burden* appeared to be an important indicator of

acceptability of the VHS. In particular, lower levels of perceived burden when using the VHS were more prevalent among younger adults (aged 18-24 years) than among older adults, and among people of White ethnic background, compared with participants from a minority ethnic background. This is consistent with national data showing higher rates of self-harm among younger adults [1,3] and among people of White ethnic backgrounds compared with other ethnic backgrounds [33]. Our intervention may benefit some vulnerable individuals. However, our results suggest that further modifications are required to increase acceptability among older adults and minority ethnic groups with respect to perceived burden. Moreover, higher levels of perceived burden were more prevalent among people who reported having a long-term health condition than among those who did not report having a long-term health condition. This may suggest that having the additional burden of a long-term condition may affect the perceptions of interventions designed to target mental health challenges. Given that 30% of the UK population reported living with a long-term condition and a mental health problem (approximately 4.6 million people) [34], this may consequently be an important target population for future studies.

Third, a history of self-harm was associated with perceived acceptability of the VHS. In our study, people who reported self-harm in the past year were more likely to report higher levels of perceived burden compared with those who did not report self-harm in the past year. The VHS has previously been found to be effective for people recently admitted to the hospital after an episode of self-harm [11]. Consequently, our findings suggest that the VHS could be improved further for people with a longer history of self-harm, to make the VHS less burdensome among people who have self-harmed relatively recently (ie, in the past year).

Fourth, our findings show that exposure to death and suicide is associated with perceived acceptability of the VHS. With respect to immediate family, people who reported exposure to death were more likely to report lower levels of perceived burden compared with those who did not. People who reported exposure to a suicide attempt were more likely to report favorable attitudes toward the VHS. Furthermore, people who reported exposure to NSSH were more likely to report higher levels of ethicality (fit with personal values), self-efficacy (confidence in using the VHS), and intervention coherence (confidence in understanding the intervention and how it works), compared with those who did not. With respect to close family and friends, people who reported exposure to NSSH were more likely to report higher levels of intervention coherence than those who did not. Wider research suggests that exposure to suicidal behavior may increase the risk of suicidal behavior through mechanisms such as imitation or social learning [26]. Our findings suggest that exposure to suicidal behavior may have positive effects on intervention acceptability and receptiveness to interventions, as they alleviate the perceived burden of interventions designed to support people in reducing repeat self-harm.

Overall, both our quantitative and qualitative findings suggest high levels of acceptability of the VHS. Consistent with the assumptions of the TFA that engagement in interventions may be dependent upon acceptability [20], our qualitative findings highlight: (1) factors that may affect engagement with the intervention and (2) suggestions to modify the intervention, which may increase acceptability.

With regard to the factors that may affect engagement with the intervention, there were a number of perceived challenges for participants when engaging with the intervention, which could be addressed in future iterations of the intervention, to further increase acceptability. The most commonly reported problems were related to (1) perceptions that statements might (for some people) induce negative thoughts about self-harm and act as a reminder of personal experiences of self-harm (*attitude*), (2) difficulties choosing an appropriate situation and response due to some being perceived as not personally relevant (*burden*), (3) technical issues including difficulties viewing on a mobile device or difficulties in selecting situations and responses from the drop-down menus (*burden*), and (4) perceptions of how receptive people are to the intervention (*perceived effectiveness*).

With regard to how to modify the intervention, in line with the respective TFA domains, there were a number of suggestions that may increase acceptability. The questions included (1) improved formatting of the intervention (*burden*), (2) the feature to add new situations and responses or amend existing ones (*ethicality*), and (3) clearer instructions and further details about the purpose of the VHS and how it works (*intervention coherence*).

Implications

Our findings suggest high levels of acceptability of the VHS. Participants reported that the VHS was easy to use, and encouraged and promoted self-reflection as a way of supporting people in avoiding future self-harm (*burden*). Minor technical issues were perceived to increase the intervention burden; however, these were judged by the research team to be minor issues. Interventions for self-care practices delivered web-based have been found to be acceptable and demonstrate some level of efficacy across a number of health conditions, including diabetes [35] and psoriasis [36]. This is important, as it suggests that technology-based interventions are perceived as an acceptable delivery mechanism for interventions targeting better management of long-term health conditions. Interventions for mental health problems delivered web-based are considered to be highly acceptable [37], and with further refinements, our intervention could be further developed and delivered as part of remote health care delivery [38,39]. However, further modifications and co-design are needed to make the VHS less burdensome for older adults and to explore specific requirements for people from minority ethnic groups and people with long-term health conditions. This could include clearer instructions about the purpose of the VHS and how it works, and improved formatting.

Our findings suggest that the VHS may be helpful for people in specific contexts or under particular circumstances. The timing of the last self-harm episode appears to be an important consideration when deploying the tool. The VHS was designed to be used as a support tool when someone has the urge to self-harm or immediately following an episode of self-harm [11]. Our findings suggest that further refinements are needed

for the tool to make it less burdensome (ie, requiring a lot of effort to engage with) for people who have self-harmed in the previous year. By making the tool easier to engage with, the tool may also be useful as part of long-term support strategies, such as when the urge to self-harm may not be at their height. It is encouraging that no gender differences in acceptability were found, and thus future research could explore the extent to which brief interventions for self-harm could be delivered by health care professionals, for example, research suggests that more emphasis should be placed on improving self-harm care for patients, with a focus on improving the implementation of self-harm management guidelines [40]. One possible route is to explore the use of the VHS alongside GP care, given that the recognition of primary care is an important place to potentially help people reduce repeat self-harm [41].

Strengths and Limitations

Previous studies examining the prevalence of self-harm have focused on general population samples [29] or adolescent samples [42]. To our knowledge, this is the first study to characterize a national community sample of adults who have previously self-harmed with respect to demographic variables, history of NSSH, suicidal ideation and suicide attempts, and exposure to death and suicide. This is important because knowing more about this community population allows more targeted preventative strategies for self-harm, with respect to specific subgroups who may benefit the most from interventions.

There was evidence of good acceptability of the VHS in line with the TFA domains. In the absence of a universally agreed criterion for TFA acceptability when deploying quantitative measures, we present an assessment criterion that fully operationalizes each of the seven TFA domains and can be used in future studies to assess perceived intervention acceptability. Further developments may be needed to increase perceived acceptability of the VHS for self-harm among older adults, people from minority ethnic groups, and people with long-term health conditions. This study had some limitations. Participants were identified from a pre-existing sample of the general public who reported a previous history of self-harm and were recruited and incentivized by YouGov to complete the questionnaire. Therefore, the sample may not be fully representative of a community with a history of self-harm. However, YouGov attempted to overcome this by seeking the widest possible variation in terms of demographic characteristics, according to people who reported a history of self-harm.

Due to a lack of available studies among community samples with a history of self-harm, we were unable to determine whether our sample is representative of this population. However, we were able to compare our sample with data from the Adult Psychiatric Morbidity Survey of the general population to compare demographic characteristics and self-harm outcomes among people who reported a history of self-harm. Our sample closely resembled the Adult Psychiatric Morbidity Survey data [29] in terms of gender and age. However, our sample contained a lower proportion of people from a minority ethnic background compared with the national data. Our sample also reported a higher prevalence of suicidal ideation (lifetime and past year), suicide attempts (lifetime), and NSSH (lifetime) compared with national data. We were unable to identify data on self-harm outcomes in the past week and NSSH outcomes in the past week or past year. Finally, due to the lack of an appropriate instrument to measure the seven TFA constructs, the research team developed a measure that could be deployed for the present purposes. It would be valuable to conduct follow-up work that examines the psychometric properties of the tool to determine whether it could be used in the context of other contexts.

Conclusions

A brief intervention based on implementation intentions has been shown to be effective in reducing self-harm in people recently admitted to the hospital after an episode of self-harm. Our findings show high levels of acceptability more generally in people who have previously self-harmed, particularly among younger adults, people of White ethnic backgrounds, and people without long-term health conditions. The intervention still has room for improvement with respect to further modifications around language and technical features, but with emphasis on other target populations. It is hoped that this intervention will provide a useful tool for both individuals to construct their own personalized implementation intentions and as part of long-term support for preventing self-harm delivered by health care professionals.

Acknowledgments

This work was funded by the National Institute for Health Research (NIHR) Greater Manchester Patient Safety Translational Research Centre. CJA is supported by the NIHR Manchester Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Proportion of responses at each point according to Theoretical Framework of Acceptability domains.
[DOCX File , 16 KB - jmir_v23i9e28349_app1.docx]

References

1. Hawton K, Bergen H, Casey D, Simkin S, Palmer B, Cooper J, et al. Self-harm in England: a tale of three cities. Multicentre study of self-harm. Soc Psychiatry Psychiatr Epidemiol 2007 Jul 21;42(7):513-521. [doi: 10.1007/s00127-007-0199-7] [Medline: 17516016]
2. Kapur N. Health services and suicide prevention. J Ment Health 2009 Jul 06;18(1):1-5. [doi: 10.1080/09638230802370704]
3. Carr MJ, Ashcroft DM, Kontopantelis E, Awenat Y, Cooper J, Chew-Graham C, et al. The epidemiology of self-harm in a UK-wide primary care patient cohort, 2001-2013. BMC Psychiatry 2016 Feb 29;16(1):53 [FREE Full text] [doi: 10.1186/s12888-016-0753-5] [Medline: 26923884]
4. Department of Health. Preventing suicide in England: Third progress report of the cross-government outcomes strategy to save lives. Department of Health. 2017. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/582117/Suicide_report_2016_A.pdf [accessed 2021-06-01]
5. World Health Organization. National suicide prevention strategies: progress, examples and indicators. World Health Organization. 2018. URL: https://apps.who.int/iris/handle/10665/279765 [accessed 2021-06-01]
6. O'Connor RC, Wetherall K, Cleare S, Eschle S, Drummond J, Ferguson E, et al. Suicide attempts and non-suicidal self-harm: national prevalence study of young adults. BJPsych Open 2018 May 09;4(3):142-148 [FREE Full text] [doi: 10.1192/bjo.2018.14] [Medline: 29922479]
7. Hjelmeland H, Hawton K, Nordvik H, Bille-Brahe U, De Leo D, Fekete S, et al. Why people engage in parasuicide: a cross-cultural study of intentions. Suicide Life Threat Behav 2002;32(4):380-393. [doi: 10.1521/suli.32.4.380.22336] [Medline: 12501963]
8. O'Connor RC, Smyth R, Ferguson E, Ryan C, Williams JM. Psychological processes and repeat suicidal behavior: a four-year prospective study. J Consult Clin Psychol 2013 Dec;81(6):1137-1143 [FREE Full text] [doi: 10.1037/a0033751] [Medline: 23855989]
9. Gollwitzer PM. Implementation intentions: strong effects of simple plans. Am Psychol 1999;54(7):493-503. [doi: 10.1037/0003-066x.54.7.493]
10. Gollwitzer P, Sheeran P. Implementation intentions and goal achievement: a meta‐analysis of effects and processes. Adv Experimet Soc Psychol 2006;38:69-119 [FREE Full text] [doi: 10.1016/S0065-2601(06)38002-1]
11. Armitage CJ, Rahim WA, Rowe R, O'Connor RC. An exploratory randomised trial of a simple, brief psychological intervention to reduce subsequent suicidal ideation and behaviour in patients admitted to hospital for self-harm. Br J Psychiatry 2016 May 02;208(5):470-476. [doi: 10.1192/bjp.bp.114.162495] [Medline: 26743808]
12. Armitage CJ. A volitional help sheet to encourage smoking cessation: a randomized exploratory trial. Health Psychol 2008 Sep;27(5):557-566. [doi: 10.1037/0278-6133.27.5.557] [Medline: 18823182]
13. O'Connor RC, Rasmussen S, Beautrais A. Recognition of suicide risk, crisis helplines, and psychosocial interventions: a selective review. In: O'Connor RC, Platt S, Gordon J, editors. International Handbook of Suicide Prevention: Research, Policy and Practice. London: Wiley Blackwell; 2011:435-456.
14. Madge N, Hewitt A, Hawton K, de Wilde EJ, Corcoran P, Fekete S, et al. J Child Psychol Psychiatry 2008 Jun;49(6):667-677. [doi: 10.1111/j.1469-7610.2008.01879.x] [Medline: 18341543]
15. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: Toward an integrative model of change. J Consul Clin Psychol 1983;51(3):390-395. [doi: 10.1037/0022-006x.51.3.390]
16. O'Connor RC, Ferguson E, Scott F, Smyth R, McDaid D, Park A, et al. A brief psychological intervention to reduce repetition of self-harm in patients admitted to hospital following a suicide attempt: a randomised controlled trial. Lancet Psychiatry 2017 Jun;4(6):451-460. [doi: 10.1016/s2215-0366(17)30129-3]
17. Diepeveen S, Ling T, Suhrcke M, Roland M, Marteau TM. Public acceptability of government intervention to change health-related behaviours: a systematic review and narrative synthesis. BMC Public Health 2013 Aug 15;13(1):756 [FREE Full text] [doi: 10.1186/1471-2458-13-756] [Medline: 23947336]
18. Hommel K, Hente E, Herzer M, Ingerski LM, Denson LA. Telehealth behavioral treatment for medication nonadherence: a pilot and feasibility study. Eur J Gastroenterol Hepatol 2013 Apr;25(4):469-473 [FREE Full text] [doi: 10.1097/MEG.0b013e32835c2a1b] [Medline: 23325274]
19. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. Br Med J 2015 Mar 19;350(mar19 6):h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
20. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
21. Stok FM, de Ridder DT, de Vet E, Nureeva L, Luszczynska A, Wardle J, et al. Hungry for an intervention? Adolescents' ratings of acceptability of eating-related intervention strategies. BMC Public Health 2016 Jan 05;16(1):5 [FREE Full text] [doi: 10.1186/s12889-015-2665-6] [Medline: 26729328]
22. Sekhon M, Cartwright M, Francis J. Application of a theoretical framework to assess intervention acceptability: a semi-structured interview study. Eur Health Psychol 2016;18:A.
23. Armitage CJ. Randomized test of a brief psychological intervention to reduce and prevent emotional eating in a community sample. J Public Health (Oxf) 2015 Sep 01;37(3):438-444. [doi: 10.1093/pubmed/fdv054] [Medline: 25935897]

24. UK Office for National Statistics. Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland. UK Office for National Statistics. 2021. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland [accessed 2021-06-01]
25. McManus S, Meltzer H, Brugha T, Bebbington PE, Jenkins R. Adult psychiatric morbidity in England: results of a household survey. Health and Social Care Information Centre. 2009. URL: https://digital.nhs.uk/data-and-information/publications/statistical/adult-psychiatric-morbidity-survey/adult-psychiatric-morbidity-in-england-2007-results-of-a-household-survey [accessed 2021-06-01]
26. Dhingra K, Boduszek D, O'Connor RC. Differentiating suicide attempters from suicide ideators using the Integrated Motivational-Volitional model of suicidal behaviour. J Affect Disord 2015 Nov 01;186:211-218. [doi: 10.1016/j.jad.2015.07.007] [Medline: 26247914]
27. O'Connor RC, Rasmussen S, Hawton K. Distinguishing adolescents who think about self-harm from those who engage in self-harm. Br J Psychiatry 2012 Apr 02;200(4):330-335. [doi: 10.1192/bjp.bp.111.097808] [Medline: 22403089]
28. Renko E, Knittle K, Palsola M, Lintunen T, Hankonen N. Acceptability, reach and implementation of a training to enhance teachers' skills in physical activity promotion. BMC Public Health 2020 Oct 16;20(1):1568 [FREE Full text] [doi: 10.1186/s12889-020-09653-x] [Medline: 33076881]
29. McManus S, Gunnell D, Cooper C, Bebbington PE, Howard LM, Brugha T, et al. Prevalence of non-suicidal self-harm and service contact in England, 2000-14: repeated cross-sectional surveys of the general population. Lancet Psychiatry 2019 Jul;6(7):573-581. [doi: 10.1016/s2215-0366(19)30188-9]
30. Keyworth C, Epton T, Goldthorpe J, Calam R, Armitage CJ. Acceptability, reliability, and validity of a brief measure of capabilities, opportunities, and motivations ("COM-B"). Br J Health Psychol 2020 Sep 20;25(3):474-501. [doi: 10.1111/bjhp.12417] [Medline: 32314500]
31. Atkins L, Francis J, Islam R, O'Connor D, Patey A, Ivers N, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implement Sci 2017 Jun 21;12(1):77 [FREE Full text] [doi: 10.1186/s13012-017-0605-9] [Medline: 28637486]
32. Murphy A, Gardner DM. Pharmacists' acceptability of a men's mental health promotion program using the Theoretical Framework of Acceptability. AIMS Public Health 2019;6(2):195-208 [FREE Full text] [doi: 10.3934/publichealth.2019.2.195] [Medline: 31297404]
33. McManus S, Bebbington P, Jenkins R, Brugha T. Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey 2014. NHS Digital. 2016. URL: https://files.digital.nhs.uk/pdf/q/3/mental_health_and_wellbeing_in_england_full_report.pdf [accessed 2021-06-01]
34. Naylor C, Parsonage M, McDaid D, Knapp M, Fossey M, Galea A. Long-term conditions and mental health: the cost of co-morbidities. The Kings Fund. URL: https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/long-term-conditions-mental-health-cost-comorbidities-naylor-feb12.pdf [accessed 2021-06-01]
35. Cotter AP, Durant N, Agne AA, Cherrington AL. Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence. J Diabetes Complications 2014 Mar;28(2):243-251 [FREE Full text] [doi: 10.1016/j.jdiacomp.2013.07.003] [Medline: 24332469]
36. Bundy C, Pinder B, Bucci S, Reeves D, Griffiths C, Tarrier N. A novel, web-based, psychological intervention for people with psoriasis: the electronic Targeted Intervention for Psoriasis (eTIPs) study. Br J Dermatol 2013 Aug 13;169(2):329-336. [doi: 10.1111/bjd.12350] [Medline: 23551271]
37. Berry N, Lobban F, Emsley R, Bucci S. Acceptability of interventions delivered online and through mobile phones for people who experience severe mental health problems: a systematic review. J Med Internet Res 2016 May 31;18(5):e121 [FREE Full text] [doi: 10.2196/jmir.5250] [Medline: 27245693]
38. Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, Hanson P, et al. Advantages and limitations of virtual online consultations in a NHS acute trust: the VOCAL mixed-methods study. Health Serv Deliv Res 2018;6(21):1-162 [FREE Full text]
39. Murphy M, Scott LJ, Salisbury C, Turner A, Scott A, Denholm R, et al. Implementation of remote consulting in UK primary care following the COVID-19 pandemic: a mixed-methods longitudinal study. Br J Gen Pract 2021 Jan 17;71(704):166-177. [doi: 10.3399/bjgp.2020.0948]
40. Leather JZ, O'Connor RC, Quinlivan L, Kapur N, Campbell S, Armitage CJ. Healthcare professionals' implementation of national guidelines with patients who self-harm. J Psychiatr Res 2020 Nov;130:405-411 [FREE Full text] [doi: 10.1016/j.jpsychires.2020.08.031] [Medline: 32891028]
41. Mughal F, Troya MI, Dikomitis L, Chew-Graham CA, Corp N, Babatunde OO. Role of the GP in the management of patients with self-harm behaviour: a systematic review. Br J Gen Pract 2020 Feb 10;70(694):364-373. [doi: 10.3399/bjgp20x708257]
42. Geulayov G, Casey D, McDonald KC, Foster P, Pritchard K, Wells C, et al. Incidence of suicide, hospital-presenting non-fatal self-harm, and community-occurring non-fatal self-harm in adolescents in England (the iceberg model of self-harm): a retrospective study. Lancet Psychiatry 2018 Feb;5(2):167-174. [doi: 10.1016/s2215-0366(17)30478-9]

Abbreviations

NIHR: National Institute for Health Research
NSSH: nonsuicidal self-harm
PPI: patient and public involvement
TFA: Theoretical Framework of Acceptability
VHS: volitional help sheet

Edited by R Kukafka; submitted 02.03.21; peer-reviewed by CY Lin, W van Ballegooijen, S Victor; comments to author 14.05.21; revised version received 14.06.21; accepted 27.07.21; published 14.09.21.

Please cite as:
Keyworth C, O'Connor R, Quinlivan L, Armitage CJ
Acceptability of a Brief Web-Based Theory-Based Intervention to Prevent and Reduce Self-harm: Mixed Methods Evaluation
J Med Internet Res 2021;23(9):e28349
URL: https://www.jmir.org/2021/9/e28349
doi:10.2196/28349
PMID:34518153

Original Paper

Web-Based Behavioral Intervention Utilizing Narrative Persuasion for HIV Prevention Among Chinese Men Who Have Sex With Men (HeHe Talks Project): Intervention Development

Meiqi Xin[1], PhD; Neil S Coulson[2], PhD; Crystal Li Jiang[3], PhD; Elizabeth Sillence[4], PhD; Andrew Chidgey[5], BSc; Norman Nok Man Kwan[6], BSS; Winnie W S Mak[7], PhD; William Goggins[1], SCD; Joseph Tak Fai Lau[1], PhD; Phoenix Kit Han Mo[1], PhD

[1]Centre for Health Behaviours Research, Jockey Club School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, Hong Kong

[2]School of Medicine, University of Nottingham, Nottingham, United Kingdom

[3]Department of Media and Communication, City University of Hong Kong, Hong Kong, Hong Kong

[4]Department of Psychology, Northumbria University, Newcastle, United Kingdom

[5]AIDS Concern, Hong Kong, Hong Kong

[6]Health and Care Service Department, Hong Kong Red Cross, Hong Kong, Hong Kong

[7]Department of Psychology, The Chinese University of Hong Kong, Hong Kong, Hong Kong

Corresponding Author:
Phoenix Kit Han Mo, PhD
Centre for Health Behaviours Research
Jockey Club School of Public Health and Primary Care
The Chinese University of Hong Kong
202D, School of Public Health, Prince of Wales Hospital
30-32 Ngan Shing Street, Shatin, New Territories
Hong Kong
Hong Kong
Phone: 852 22528765
Email: phoenix.mo@cuhk.edu.hk

Abstract

Background: In the era of potent antiretroviral therapy, a high level of condomless anal intercourse continues to drive increases in HIV incidence in recent years among men who have sex with men. Effective behavior change strategies for promoting HIV-preventive behaviors are warranted. Narrative persuasion is a novel health communication approach that has demonstrated its persuasive advantages in overcoming resistance to counterattitudinal messages. The efficacy of narrative persuasion in promoting health behavior changes has been well documented, but critical research gaps exist for its application to HIV prevention.

Objective: In this study, we aimed to (1) capitalize on narrative persuasion to design a web-based multisession intervention for reducing condomless anal intercourse among men who have sex with men in Hong Kong (the HeHe Talks Project) by following a systematic development process; and (2) describe the main components of the narrative intervention that potentially determine its persuasiveness.

Methods: Persuasive themes and subtopics related to reducing condomless anal intercourse were initially proposed based on epidemiological evidence. The biographic narrative interview method was used to elicit firsthand experiential stories from a maximum variation sample of local men who have sex with men with diverse backgrounds and experiences related to HIV prevention; different types of role models were established accordingly. Framework analysis was used to aggregate the original quotations from narrators into collective narratives under 6 intervention themes. A dedicated website was finally developed for intervention delivery.

Results: A series of video-based intervention messages in biographic narrative format (firsthand experiential stories shared by men who have sex with men) combined with topic-equivalent argumentative messages were produced and programmed into 6 intervention sessions. The 6-week intervention program can be automatically delivered and monitored online.

Conclusions: We systematically created a web-based HIV prevention intervention derived from peer-generated stories. Strategies used to enhance the efficacy of the narrative intervention have been discussed within basic communication components. This paper describes the methods and experiences of the rigorous development of a narrative communication intervention for HIV prevention, which enables replication of the intervention in the future.

(J Med Internet Res 2021;23(9):e22312) doi:10.2196/22312

KEYWORDS

narrative persuasion; firsthand experiential stories; online intervention; HIV prevention; sexual behavior; men who have sex with men

Introduction

Men Who Have Sex With Men as an Important Target for HIV Prevention

Although a steadily decreasing trend has been observed for the global HIV burden, the epidemic has continued to expand among men who have sex with men [1]. For example, in Hong Kong, HIV prevalence has increased to 6.54% among men who have sex with men in recent years, while the rate has remained lower than 0.1% in the general population [2]. Despite biomedical advances in antiretroviral therapy and growing access to care and treatment, there is a paradoxical rise in the HIV incidence rate among men who have sex with men that was found to be significantly associated with an increase in condomless anal intercourse [3,4]. The disproportionate risk is also fueled by a relatively low level of HIV testing [5,6]. More efforts are necessary to develop effective behavior change strategies for the promotion of HIV-preventive behaviors among this vulnerable population.

Gaps in Communicating Persuasive Messages on HIV Prevention

Although health communication is widely acknowledged as an important tool for HIV prevention [7], existing persuasive communications has failed to significantly stimulate behavior change [8]. Previous interventions drew heavily on classic health behavior change theories to improve cognitive determinants (eg, risk perceptions) [7]; however, such interventions are based on the assumption that message recipients are so-called *rational individuals*, and the influence of recipients' preexisting state in accepting and executing desirable behaviors are omitted [9,10]. Counterattitudinal messages, namely persuasive topics and arguments that differ from recipients' own values or beliefs [11], are largely used for health behavior promotion. For instance, interventions for HIV prevention are commonly developed to target high-risk groups who engage in unprotected sex or have a low intention to use condoms consistently. Recipients are prone to argue against advocated positions to protect their previously held values or beliefs [11,12], adhering to or even engaging in more of the discouraged behavior [13]. There has long been a need to overcome such resistance to persuasive messages about HIV prevention [14].

Theoretical Mechanisms Underlying Narrative Persuasion

Over the past decade, narrative persuasion has been recognized as a promising avenue for health communication [12]. A narrative refers to "a representation of connected events and characters that has an identifiable structure [15]." The traditional approach—argumentation—otherwise features expository and didactic communication, with a logical presentation of overt arguments and factual information in the form of reasons and evidence to put forth a claim [12,15,16]. The two communication styles are fundamentally different in terms of information processing. Instead of receiving carefully elaborated explicit persuasive arguments in argumentative communication [17], narrative processing typically involves engagement with narrative elements and identification with characters (ie, vicarious experience of characters' cognitive and emotional responses) as casually and chronologically related events unfold in a coherent manner [16,18,19].

Narrative persuasion is grounded in both the Extended Elaboration Likelihood Model [18] and the Transportation-Imagery Model [20]; narrative persuasion has the advantage (over argumentation) of inhibiting negative responses to counterattitudinal messages because fewer targets are available for counterarguments, given its structure, with implicit persuasive content embedded in narrative elements; (2) there is less processing capacity and motivation to critically scrutinize core arguments that are presented in narrative format; and (3) identification with narrative characters facilitates psychological distancing from the recipients' initial story-inconsistent positions [16,18,19,21].

Remarkably, the incorporation of argumentative support further bolsters narrative persuasion [22,23], facilitating cognitive elaboration, rehearsal, and repetition, which leads to an enduring persuasive impact [18,24,25]. Prefacing argumentative messages by narrative messages could potentially predispose recipients to positively process and comprehend the overtly persuasive arguments [26-28]. Recipients could first shift from existing positions temporarily through engagement with the narrative messages, and fully assimilate the advocated positions into their own value structures through positive elaboration of the subsequent argumentation messages [18].

Harnessing Narrative Persuasion for Behavioral Interventions

Existing empirical evidence supports the efficacy of narrative interventions in promoting health behaviors relative to argumentation or null controls [29,30]. Relatively fewer studies have examined the effect of a combination of narrative and argumentative messages about cancer screening [31-33], diabetes self-care [34], and vaccination [35]; most were found to be more effective in changing the target behavior and its theoretical determinants than a single message format.

Several studies [36-41] have applied narrative communication to HIV prevention, but there are some critical research gaps. First, most narratives took the form of fictional stories (eg, entertainment-education soap opera) that were scripted by researchers; its efficacy in improving HIV-preventive behaviors is controversial [36,37]. However, other narrative types, including experiential stories, have received little exploration. Second, some interventions involved multiple behavior change strategies (eg, condom distribution), and thus, made it difficult to isolate the specific effect of the narrative component [38]. Third, the majority of previous narrative interventions addressed heterosexual risk for HIV [39,40]; we suppose that efficacy of narrative interventions is higher among homosexual people because HIV prevention is more personally relevant to this population group, and personal relevance of a health topic can enhance narrative efficacy [41]. Finally, no studies have been conducted among Asian populations. More research on narrative persuasion in different contexts is warranted because shared stories inherently embody rich cultural information that shapes behaviors and responses to persuasive communication [9,42].

Utilizing Web-Based HIV Prevention Interventions for Men Who Have Sex With Men

Web-based technology has been recognized as a powerful tool to efficiently reach target populations and deliver behavioral interventions to enhance responses to the ongoing HIV epidemic among men who have sex with men [6]. A recent meta-analysis [43] of 44 studies across 12 countries and regions found that eHealth interventions effectively reduced condomless anal intercourse, reduced multiple sex partnership, and increased the uptake of HIV testing among men who have sex with men. Moreover, a considerable proportion of men who have sex with men actively use the internet for socializing and sex-seeking, which has been found to increase likelihood of condomless anal intercourse, and consequently, risk of HIV infection [44-46]. In Hong Kong, approximately 70% of men who have sex with men engage in sex-seeking online [47]. Hence, web-based interventions can seize the opportunity to promote HIV prevention among men who have sex with men at a crucial point when they tend to engage in risky sexual practices. In addition, Chinese men who have sex with men use the internet as the primary source of information about sexual health, HIV, and other sexually transmitted infections (STIs), and accept the use of eHealth in supporting their sexual health care [48].

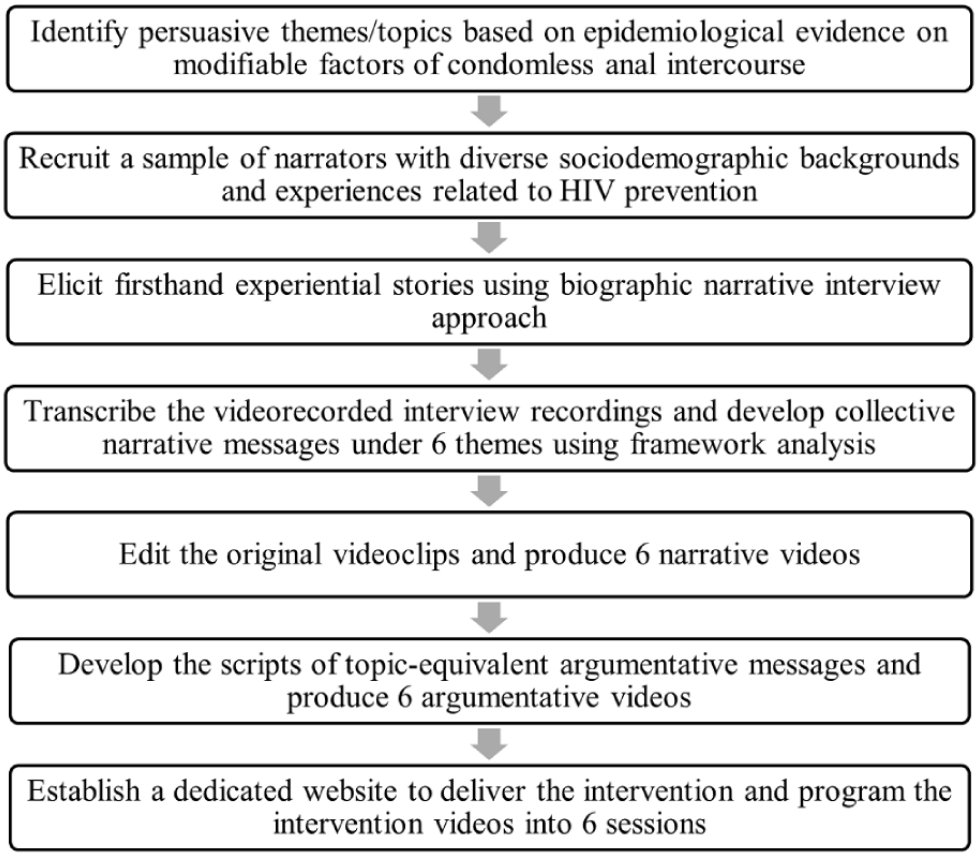
In this study, we developed a web-based narrative intervention for HIV prevention among men who have sex with men in Hong Kong (the HeHe Talks Project) to deliver a series of video-based persuasive messages to reduce condomless anal intercourse. Biographic narratives (firsthand experiential stories shared by local men who have sex with men peers) were combined with topic-equivalent argumentative messages.

Methods

Ethics and Procedure Overview

Informed consent was obtained from all participants involved in this project. Ethical approval was obtained from Joint Chinese University of Hong Kong–New Territories East Cluster Clinical Research Ethics Committee (reference number 2014.274-T). A roadmap of the development procedure is presented in Figure 1.

Figure 1. Procedure for developing the narrative intervention program (HeHe Talks).



Identifying Persuasive Themes and Subtopics

We conducted a review of modifiable and significant predictors of HIV/STI infection and condomless anal intercourse among local men who have sex with men in publications and government documents to inform topic selection. Factors were categorized into cognitive (eg, risk perceptions), interpersonal (eg, sex partnership), contextual (eg, substance use), and sexual practice–specific (eg, group sex). In addition, it has been suggested that HIV testing behavior should be addressed simultaneously, given its significant relationship with condomless anal intercourse among Chinese men who have sex with men [49,50]. Six session themes were initially proposed: knowledge about HIV/STI, condom use and partnership, condom use and risky contexts, condom use and substance use, efficacy of condom use, and HIV testing (Multimedia Appendix 1).

Qualitative evidence was obtained by conducting a focus group discussion with 5 Hong Kong men who have sex with men (age: range 28-38 years). The focus group was led by a public health researcher and a health psychologist and was used to assess knowledge, views, and information needs regarding HIV prevention and condomless anal intercourse. Most interviewees had university-level educations and above (4/5) and had full-time jobs (4/5); 2 interviewees reported consistent condom use for anal sex, and 1 had engaged in substance use–facilitated anal sex in the previous 6 months.

This focus group informed refinement of the intervention topics. For example, the participants emphasized the phenomenon of "complacency about HIV" among local men who have sex with men (ie, that strong belief about antiretroviral therapy efficacy contributed to low-risk perceptions and high motivation to engage in condomless anal intercourse). Proposed intervention topics were sent to several HIV specialists; we further revised the topics based on their feedback. For example, they suggested focusing on drug use other than alcohol consumption to address reducing substance use–facilitated condomless anal intercourse; accordingly, the subtopic *impact of alcohol use on condomless anal intercourse* was removed.

Recruiting Narrators and Operationalizing the Role Models

The Sabido entertainment-education strategy [25] was applied to establish 3 types of role models: (1) positive characters, who engage in consistent condom use while having anal intercourse and undergo regular HIV testing, and consequently, maintain sexual health; (2) negative characters, who engage in condomless anal intercourse and do not regularly undergo HIV testing, and are negatively affected (eg, infected with HIV or STI); and (3) transitional characters, who once engaged in risky sexual behaviors and suffered negative consequences and then adopted more positive attitudes and behaviors. Role modeling has been adopted in previous entertainment-education programs for HIV prevention [37,39]. Maximum variation sampling was used to recruit 2 to 3 HIV-positive and 6 to 7 HIV-negative narrators with different experiences and backgrounds [51].

Criteria for all narrators were (1) men who have sex with men; (2) age ≥18 years; (3) Hong Kong residents; (4) able to speak Cantonese fluently; and (5) of sufficient mental and physical capability to take part in a 2-hour interview. Criteria for HIV-positive narrators were (1) confirmed or high probability

of having acquired HIV by sexual transmission; (2) encountered negative experiences due to living with HIV; (3) once or currently undergoing antiretroviral treatment; and (4) having engaged in any condomless anal intercourse with male sex partners prior to the diagnosis. Criteria for HIV-negative narrators were (1) having undergone at least 1 HIV test in the previous year, with the most recent test having yielded a negative result; (2) consistently used condoms with male sex partners (in every sexual encounter in the previous year); and (3) having positive perceptions about condom use for HIV prevention. Prospective narrators were nominated and approached by the research team and a collaborating nongovernmental organization (AIDS Concern). A survey was administered to establish their eligibility and to create a profile for each individual, with background information (eg, age and sexual partnership) and specific experiences related to HIV prevention (eg, contracting HIV/STI infections, engaging in condomless anal intercourse).

Developing Persuasive Messages in Narrative and Argumentation Formats

Biographic-Narrative Interpretive Method [52,53] was employed to elicit narratives for the development of health promotion messages. A semistructured interview guide was drafted around the proposed topics and tailored to each narrator's profile. Two interviewers (MX and NK) and a professional camera operator were involved in the individual interviews with each of the selected narrators. One of the interviewers (NK) was a senior practitioner from the collaborating nongovernmental organization who had rich experience in providing community support services to local men who have sex with men; he took a leading role in the narrative interviews and preinterview negotiations, particularly with HIV-positive narrators, to ensure that the questions were delivered in a comfortable and respectful manner.

Interviewers first briefed the narrator on interview procedures and established rapport. Individual interviews started with a single question that was thematically or temporally focused, to elicit narratives of personal experiences, such as life after being diagnosed as HIV positive or an occasion of safe sex practices at a gay venue. Narrators were free to tell their story in the manner of their choosing (to provoke the narration within their own system of relevancy), and interviewers merely provided nondirectional facilitative support. Keyword notes were also taken to document the topics of interest arising in the original narrative, which helped inform follow-up questions for additional narrative to enrich the overarching story. After narrative-seeking questions, interviewers asked additional questions about topics that had not been raised or for purposes other than narrative extraction (eg, to make a persuasive appeal to the audience).

Videorecordings of the interviews were transcribed. Framework analysis was conducted to develop themed narrative messages using NVivo (version 11, QSR International) in 5 steps [54]: (1) familiarizing: carefully reading the transcripts and observational notes to gain a comprehensive understanding of the data; (2) framing: devising and refining a coding framework drawn on both predefined and emergent themes and subtopics; (3) indexing: systemically sorting original narrative quotations within the thematic framework; (4) charting: grouping indexed narratives from different narrators for each coded subtopic; (5) mapping: aggregating subtopics and corresponding quotations in a coherent manner under each of the 6 session themes to generate collective narrative messages. For each subtopic, argumentative messages were drafted based on the narrative content, by presenting didactic statements, factual claims, logical reasons, and statistical evidence.

Producing Intervention Videos and a Web-Based Delivery Platform

The videorecordings of the original narratives were segmented and re-organized to create 6 videos for the themes derived from the framework analysis. A background profile was presented for each narrator at first appearance in the videos; persuasive appeals from 1 or 2 narrators were used as epilogues. Videos were sent to the narrators to get their approval. Three male speakers (a nongovernmental organization staff and 2 research assistants) were invited to deliver scripted, videorecorded argumentative messages for each theme. Supplementary material, such as openings, subtitles, images, and background music, were added, and the finalized videos were programmed into 6 intervention sessions. Each included 2 videos with narrative (15-20 minutes) and argumentative (8-10 minutes) messages and a postvideo quiz set to enhance participants' dedication to intervention content. The narrative videos were longer than the topic-equivalent argumentative videos due to the inherent natures of storytelling (ie, portraying a series of events) and argumentation (ie, listing facts and evidence). A dedicated website was developed to deliver intervention sessions automatically, and a content management system was established for real-time monitoring of engagement in the program.

Results

Narrator Characteristics

A total of 36 men who have sex with men were screened, and 9 men were selected as narrators (Table 1). Two men, who had been diagnosed as HIV-positive 3 years prior, served as negative role-model narrators. The other men had recent HIV-negative test results (within 1 year): 3 men had never engaged in condomless anal intercourse with a male sex partner in their lifetime, and thus, served as positive role-model narrators; 4 men had once engaged in condomless anal intercourse but had consistently used condoms in the past year, and thus, served as transitional role-model narrators.

Narrators had varying sexual practices (Table 2): 3 had regular partners, and 5 had casual partners; 4 had engaged in both insertive (taking a top role) and receptive (taking a bottom role) anal intercourse; 6 had engaged in *casual fun*—anal intercourse with casual sex partners in collective sex environments (eg, gay saunas, group sex events [55]); 5 had engaged in *chem-sex*—anal intercourse while under the influence of psychotropic substances (eg, methamphetamine).

Table 1. Background characteristics of the 9 narrators.

Narrator (role model type)	Age (years)	Employment	Years as member[a]	HIV[b] status	History of other STIs[c] (type)	Type of HIV testing venues visited
A (negative)	24	Unemployed	Not disclosed	Positive	Yes (syphilis)	Community organization
B (negative)	24	Unemployed	>10	Positive	No	Public social hygiene clinic
C (transitional)	30	Full-time	>10	Negative	Yes (urethritis, pubic lice)	Public social hygiene clinic; community organization
D (positive)	38	Full-time	>10	Negative	No	Community organization; private clinic
E (transitional)	26	Full-time	>5	Negative	No	Public social hygiene clinic; community organization
F (transitional)	36	Full-time	>10	Negative	Yes (syphilis)	Community organizations
G (positive)	30	Full-time; Part-time	>10	Negative	No	Public social hygiene clinic; community organization
H (transitional)	32	Full-time	>10	Negative	No	Community organization
I (positive)	22	Student	>5	Negative	No	Community organization

[a]Period between self-identification as homosexual or bisexual and involvement in the local community of men who have sex with men.

[b]HIV: human immunodeficiency virus.

[c]STIs: sexually transmitted infections.

Table 2. Characteristics of the 9 narrators related to sexual practices (ie, anal intercourse with male sex partners prior to diagnosis or in the past year).

Narrator (role model type)	Sexual partnership	Sexual role	Consistent condom use[a]	Casual fun[b] (venue)	Chem-sex[c] (drug type[d])
A (negative)	Not recorded	Not recorded	No	No	No
B (negative)	Not recorded	Not recorded	No	Yes (gay sauna; private party)	Yes (meth, g water, foxy, rush poppers)
C (transitional)	Regular; casual	Top; bottom	Yes (recent)	Yes (gay sauna; private party)	Yes (rush poppers)
D (positive)	Regular	Top	Yes (lifetime)	No	No
E (transitional)	Casual	Top; bottom	Yes (recent)	Yes (gay sauna)	Yes (meth, g water, foxy)
F (transitional)	Casual	Bottom	Yes (recent)	Yes (gay sauna)	Yes (meth, foxy, ecstasy)
G (positive)	Regular	Top	Yes (lifetime)	No	Yes (rush poppers)
H (transitional)	Regular; casual	Top; bottom	Yes (recent)	Yes (gay sauna)	No
I (positive)	Casual	Top; bottom	Yes (lifetime)	Yes (gay sauna)	No

[a]The response *recent* referred to having once engaged in any condomless anal intercourse in lifetime but consistently using condoms within the past year; the response *lifetime* referred to having never engaged in condomless anal intercourse in lifetime.

[b]This refers to engaging in anal intercourse with casual sex partners at any local public venues.

[c]This refers to taking illicit drugs during or prior to anal intercourse.

[d]Meth, g water, foxy (ie, foxy methoxy), rush poppers, and ecstasy as slang for methamphetamine; gamma hydroxybutyrate/gamma butyrolactone; 5-methoxy-N,N-diisopropyltryptamine; inhalant alkyl nitrites; and 3,4-methylenedioxy-methamphetamine, respectively, were reported.

Topics and Contents of Persuasive Messages

There are 6 intervention sessions in total (Table 3); all sessions start with persuasive messages in the narrative format. The first session (A New Era of HIV Prevention) introduces the outcomes of HIV infection in the era of potent antiretroviral therapy. The HIV-positive peers narrate their experiences of receiving a diagnosis and living with HIV, with a focus on the severity of infections, negative influences on physical health and personal life, and co-infections with other STIs. The second session (Partnership and Protected Sex) includes experiences of condom use with different types of sexual partners, including actively negotiating condom use and making sexual decisions, resisting sexual pressure from partners, and balancing emotional relationships and safe sex practices. The third session (Casual Fun and Protected Sex) presents both obstacles to condom use experienced in high-risk contexts (eg, dark rooms in gay saunas and group sex parties) and practical skills for coping with unintentional condomless anal intercourse. The fourth session (Chem-Sex) focuses on potential exposure to psychotropic substances at local gay venues, negative influences of substance use on sexual practices and physical health, and problems related

to drug addiction. The fifth session (Efficacy of Condom Use) focuses on consistent and correct condom use for anal intercourse as well as its efficacy in preventing STIs. The last session (HIV Testing) includes experiences of undergoing HIV testing and counseling at different service sites and emphasizes the importance of regular testing while adhering to safe sex practices. Each narrative video ends with narrators appealing to the listeners. For example, the first session includes a positive message from a HIV-positive narrator; he talks about gradually accepting and recognizing his infection status, adhering to treatment, and having hope for a new future.

Table 3. An overview of topics in the narrative messages for each intervention session.

Themes and subtopics of narrative messages	Message sources (narrator and role model type)
Session 1: A New Era of HIV Prevention	
1. HIV diagnosis	A (negative) and B (negative)
2. Life impact of living with HIV	A (negative) and B (negative)
3. Problems encountered during HIV treatment	A (negative) and B (negative)
4. Co-infection with other sexually transmitted diseases	A (negative)
Session 2: Partnership and Protected Sex	
1. Active communication about condom use and assessment of partner(s)' HIV risk	D (positive), C (transitional), and E (transitional)
2. Assertive responses to unwanted condomless sex	A (negative), D (positive), G (positive), C (transitional), and E (transitional)
3. Barriers to condom use when in an emotional relationship	D (positive), G (positive), C (transitional), and E (transitional)
Session 3: Casual Fun and Protected Sex	
1. Risk of condomless sex with casual partners and norm about condom use at gay saunas	B (negative), D (positive), I (positive), C (transitional), E (transitional), and F (transitional)
2. Unintentional condomless sex encountered during casual fun	B (negative), D, I (positive), C (transitional), E (transitional), and F (transitional)
3. Contextual risk of group sex parties	B (negative), D (positive), C (transitional), and H (transitional)
Session 4: Chem-Sex	
1. Impact of drug abuse on sexual practices	B (negative), E (transitional), F (transitional), and H (transitional)
2. Impact of drug abuse on other health outcomes	B (negative), D (positive), C (transitional), E (transitional), and F (transitional)
3. Exposure to drugs when seeking sex	B (negative), D (positive), G (positive), I (positive), C (transitional), E (transitional), and F (transitional)
4. Drug addiction	B (negative), E (transitional), and F (transitional)
Session 5: Efficacy of Condom Use	
1. Protective efficacy of condom use	D (positive), G (positive), I (positive), and E (transitional)
2. Correct condom use	D (positive), I (positive), C (transitional), and E (transitional)
3. Consistent condom use	D (positive), G (positive), C (transitional), F (transitional), and G (transitional)
Session 6: HIV Testing	
1. Local HIV testing services	D (positive), G (positive), C (transitional), E (transitional), and H (transitional)
2. Regular HIV testing	A (negative), G (positive), I (positive), C (transitional), and H (transitional)

The topic-equivalent argumentative messages are delivered after narratives. For example, for the subtopic *Consistent condom use* (in Session 5), the narrative presents experiences of condomless sex and STI infections from an HIV-positive narrator, and the argumentative message is

> *If the sexual partner is HIV-positive, one could get infected through only a "single" episode of unprotected sexual encounter, ...research studies demonstrated that there was no significant difference in HIV risk between men who have sex with men who "inconsistently use condoms" and those who "have never ever used condoms."*

After videos, a quiz question is asked. For example, "According to the video messages, what type of immune cell was primarily attacked by HIV?" One correct response options ("CD4 cell") and 2 false response options ("CD3 cell" and "CD8 cell") are provided. Feedback and reinforcement are not given after the quiz, although users can return to the video during and after the quiz.

Web-Based Intervention Program

The program website can deliver the intervention automatically in a standardized manner: the 6 intervention sessions can be released sequentially on a weekly basis, and each session is accessible upon completion of the preceding session. Users can be assigned a unique password-protected account to access intervention content within a 6-week period, and activities are automatically recorded (eg, the dates and time of completing an intervention video). The program can also send out reminders when a new session is released, or users fail to engage in any activity for 2 to 3 weeks.

Discussion

General

We developed a theory-based and evidence-based narrative communication intervention for HIV prevention among men who have sex with men. The communication strategies used by this intervention program to enhance the potential narrative efficacy can be systematically discussed within the 4 basic components of health communication: message, source, recipient, and channel [12].

Message

Narratives took the form of firsthand experiential stories, in contrast to commonly used fictional stories [12]. To maintain the authenticity of personal experience narratives, which is the key to exploiting its advantages (ie, being perceived as credible, evoking high engagement and inhibiting counterarguing) over those of professionally generated stories in persuasion [15,56], we used well-developed methods for biographic narrative interviews to obtain stories [52]. The narrators were always given the space and time to narrate without any interruption; while the interviewers performed active listening by intermittently giving prompts, with nonverbal posture or empathetic and unobtrusive mirroring of the emotions expressed, throughout the storytelling. In addition, the intervention messages were produced by directly extracting biographic accounts from the videorecorded interview. Narratives conveyed in first-person are more effective than third-person narratives for STI prevention [35], because an individual is more likely to mentally embody characters' perspectives (ie, identification with characters) [57]. These strategies also strengthened the incorporation of cultural codes including visual presentations, language, and idioms that could heighten narrative engagement [58]. Furthermore, this narrative intervention featured subsequent provision of argumentative support. The argumentative messages were developed to match the themes and subtopics of the narrative messages, to reinforce the predisposition created by narrative persuasion but to avoid provoking counterargument [18].

Source

Collective narratives were extracted from a maximum variation sample of local men who have sex with men peers in this study to improve the comprehensiveness of persuasive content [15] and appeal to different audience segments [25]. The diversity of narrators allowed 3 types of characters to be established. The transitional role model was utilized the most in order to reduce selective avoidance of counterattitudinal messages about HIV prevention [19] and increase perceived homophily and realism of the characters, and thus, facilitate narrative processing [18], and enhance self-efficacy in changing risky sexual behaviors [59]. Positive and negative role models were also used to reinforce the consistency of condom use. Moreover, biographic narratives were a representation of the narrators' views of their own life-course, beliefs, interpretative schemata, and principles of judgement [52], which provided authentic and comprehensive portrayals of the characters, and in particular, avoided fostering negatively biased images of HIV-positive peers. For instance, the narrators with a history of HIV/STI shared their experiences of infection and risky sexual encounters as well as how they strived to adhere to treatments and improve their HIV-preventive behaviors.

Recipients

The targeted recipients are Chinese men who have sex with men, a highly marginalized community where members are strongly connected to each other [60]; social norms and community involvement have been found to be positively associated with their HIV-preventive behaviors [61,62]. To produce culturally authentic messages and engage the recipients [9], members of the local men who have sex with men community were actively involved throughout the intervention development, including determining persuasive topics, screening eligible narrators, assisting in narrative interviews, and generating and delivering intervention messages. Such participatory strategies can be applicable to other similar contexts where the community tend to share motives and values regarding health behavior.

Channel

The intervention comprises web-based videos. Video-based narratives have demonstrated strengths in building affection, facilitating identification with characters, and message elaboration, thereby promoting behavior changes [29,63,64]. Narrators' ability to organize and articulate stories in videorecorded interviews were carefully evaluated during the recruitment. The web-based delivery platform allows navigation of the content at the user's own pace without the presence of researchers; users are less likely to mask emotional responses and experience greater immersion when they view narratives alone [56].

Conclusions

In response to the call for innovative health communication approaches for HIV prevention, this paper describes the scientific rationale and rigorous procedure of applying narrative persuasion to promote HIV-preventive behaviors among men who have sex with men. This web-based narrative communication intervention can be easily replicated in other contexts and can be incorporated into comprehensive HIV prevention services.

Acknowledgments

This study was funded by the University Grants Committee of Hong Kong 2015-2016 General Research Fund (reference 14612615). We would like to express our heartfelt gratitude to all the narrators and participants in this project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
A summary of factors of HIV/STI infection and related behaviors among men who have sex with men in Hong Kong.
[DOCX File , 34 KB - jmir_v23i9e22312_app1.docx]

References

1. Beyrer C, Baral SD, van Griensven F, Goodreau SM, Chariyalertsak S, Wirtz AL, et al. Global epidemiology of HIV infection in men who have sex with men. Lancet 2012 Jul 28;380(9839):367-377 [FREE Full text] [doi: 10.1016/S0140-6736(12)60821-6] [Medline: 22819660]
2. HIV surveillance report 2017 update. The Virtual AIDS Office of Hong Kong. Hong Kong: Hong Kong Department of Health; 2018. URL: https://www.aids.gov.hk/english/surveillance/sur_report/hiv17.pdf [accessed 2021-05-31]
3. Phillips AN, Cambiano V, Nakagawa F, Brown AE, Lampe F, Rodger A, et al. Increased HIV incidence in men who have sex with men despite high levels of ART-induced viral suppression: analysis of an extensively documented epidemic. PLoS ONE 2013 Feb 15;8(2):e55312. [doi: 10.1371/journal.pone.0055312]
4. Recommended HIV/AIDS strategies for Hong Kong (2017-2021). Hong Kong Advisory Council on AIDS. URL: http://www.aca.gov.hk/english/strategies/pdf/strategies17-21.pdf [accessed 2021-05-31]
5. Campbell CK, Lippman SA, Moss N, Lightfoot M. Strategies to increase HIV testing among MSM: a synthesis of the literature. AIDS Behav 2018 Aug 17;22(8):2387-2412. [doi: 10.1007/s10461-018-2083-8] [Medline: 29550941]
6. Beyrer C, Baral SD, Collins C, Richardson ET, Sullivan PS, Sanchez J, et al. The global response to HIV in men who have sex with men. Lancet 2016 Jul 09;388(10040):198-206. [doi: 10.1016/S0140-6736(16)30781-4] [Medline: 27411880]
7. Storey D, Seifert-Ahanda K, Andaluz A, Tsoi B, Matsuki J, Cutler B. What is health communication and how does it affect the HIV/AIDS continuum of care? a brief primer and case study from New York City. J Acquir Immune Defic Syndr 2014 Aug 15;66 Suppl 3:S241-S249. [doi: 10.1097/QAI.0000000000000243] [Medline: 25007193]
8. Albarracín D, McNatt PS, Klein CTF, Ho RM, Mitchell AL, Kumkale GT. Persuasive communications to change actions: an analysis of behavioral and cognitive impact in HIV prevention. Health Psychol 2003 Mar;22(2):166-177 [FREE Full text] [doi: 10.1037//0278-6133.22.2.166] [Medline: 12683737]
9. Airhihenbuwa CO, Obregon R. A critical assessment of theories/models used in health communication for HIV/AIDS. J Health Commun 2000 Feb 02;5 Suppl(sup1):5-15. [doi: 10.1080/10810730050019528] [Medline: 11010357]
10. Norman P, Conner M. The theory of planned behaviour and binge drinking: assessing the moderating role of past behaviour within the theory of planned behaviour. Br J Health Psychol 2006 Feb;11(Pt 1):55-70. [doi: 10.1348/135910705X43741] [Medline: 16480555]
11. Slater M. Involvement as goal-directed strategic processing. In: The Persuasion Handbook: Developments in Theory Practice. Thousand Oaks: Sage Publications; 2002:175-194.
12. Hinyard LJ, Kreuter MW. Using narrative communication as a tool for health behavior change: a conceptual, theoretical, and empirical overview. Health Educ Behav 2007 Oct 30;34(5):777-792. [doi: 10.1177/1090198106291963] [Medline: 17200094]
13. Brehm S, Brehm J. Psychological Reactance: A Theory of Freedom and Control. New York: Academic Press; 2013.
14. Campbell RG, Babrow AS. The role of empathy in responses to persuasive risk communication: overcoming resistance to HIV prevention messages. Health Commun 2004;16(2):159-182. [doi: 10.1207/S15327027HC1602_2] [Medline: 15090283]
15. Kreuter MW, Green MC, Cappella JN, Slater MD, Wise ME, Storey D, et al. Narrative communication in cancer prevention and control: a framework to guide research and application. Ann Behav Med 2007 Jun;33(3):221-235. [doi: 10.1007/BF02879904] [Medline: 17600449]
16. Bilandzic H, Busselle R. Narrative persuasion. In: Dillard JP, Shen L, editors. The Persuasion Handbook: Developments in Theory Practice second ed. Thousand Oaks: Sage Publications; 2013:200-219.
17. O'Keefe D. The elaboration likelihood model. In: Dillard JP, Shen L, editors. The Persuasion Handbook: Developments in Theory Practice second ed. Thousand Oaks: Sage Publications; 2013:137-149.
18. Slater MD, Rouner D. Entertainment-education and elaboration likelihood: understanding the processing of narrative persuasion. Commun Theory 2002 May;12(2):173-191. [doi: 10.1111/j.1468-2885.2002.tb00265.x]
19. Moyer-Guse E. Toward a theory of entertainment persuasion: Explaining the persuasive effects of entertainment-education messages. Communication theory 2008;18(3):407-425. [doi: 10.1111/j.1468-2885.2008.00328.x]

20. Green MC, Brock TC. The role of transportation in the persuasiveness of public narratives. J Pers Soc Psychol 2000 Nov;79(5):701-721. [doi: 10.1037//0022-3514.79.5.701] [Medline: 11079236]
21. Cohen J. Defining identification: a theoretical look at the identification of audiences with media characters. Mass Communication and Society 2001 Aug;4(3):245-264. [doi: 10.1207/S15327825MCS0403_01]
22. O'Keefe DJ. Justification explicitness and persuasive effect: a meta-analytic review of the effects of varying support articulation in persuasive messages. Argumentation and Advocacy 2017 Dec 18;35(2):61-75. [doi: 10.1080/00028533.1998.11951621]
23. de Graaf A, Sanders J, Hoeken H. Characteristics of narrative interventions and health effects: a review of the content, form, and context of narratives in health-related narrative persuasion research. Rev Commun Res 2016;4:88-131. [doi: 10.12840/issn.2255-4165.2016.04.01.011]
24. Igartua J, Vega Casanova J. Identification with characters, elaboration, and counterarguing in entertainment-education interventions through audiovisual fiction. J Health Commun 2016 Jan 05;21(3):293-300. [doi: 10.1080/10810730.2015.1064494] [Medline: 26166213]
25. Singhal A, Rogers E. Entertainment-Education: A Communication Strategy for Social Change. Mahwah, New Jersey: Routledge; 1999.
26. Moyer-Gusé E, Jain P, Chung AH. Reinforcement or reactance? examining the effect of an explicit persuasive appeal following an entertainment-education narrative. J Commun 2013 Mar 20;62(6):1010-1027. [doi: 10.1111/j.1460-2466.2012.01680.x]
27. Lane R, Miller AN, Brown C, Vilar N. An examination of the narrative persuasion with epilogue through the lens of the elaboration likelihood model. Commun Q 2013 Sep;61(4):431-445. [doi: 10.1080/01463373.2013.799510]
28. Oh HJ, Larose R. Tell me a story about healthy snacking and I will follow: comparing the effectiveness of self-generated versus message-aided implementation intentions on promoting healthy snacking habits among college students. Health Commun 2015 Sep 25;30(10):962-974. [doi: 10.1080/10410236.2014.910289] [Medline: 25256795]
29. Shen F, Sheer VC, Li R. Impact of narratives on persuasion in health communication: a meta-analysis. J Advert 2015 May 05;44(2):105-113. [doi: 10.1080/00913367.2015.1018467]
30. Perrier M, Martin Ginis KA. Changing health-promoting behaviours through narrative interventions: a systematic review. J Health Psychol 2018 Dec;23(11):1499-1517. [doi: 10.1177/1359105316656243] [Medline: 27387514]
31. Dillard AJ, Fagerlin A, Dal Cin S, Zikmund-Fisher BJ, Ubel PA. Narratives that address affective forecasting errors reduce perceived barriers to colorectal cancer screening. Soc Sci Med 2010 Jul;71(1):45-52 [FREE Full text] [doi: 10.1016/j.socscimed.2010.02.038] [Medline: 20417005]
32. McGregor LM, von Wagner C, Vart G, Yuen WC, Raine R, Wardle J, et al. The impact of supplementary narrative-based information on colorectal cancer screening beliefs and intention. BMC Cancer 2015 Mar 21;15(1):162 [FREE Full text] [doi: 10.1186/s12885-015-1167-3] [Medline: 25884168]
33. McQueen A, Caburnay C, Kreuter M, Sefko J. Improving adherence to colorectal cancer screening: a randomized intervention to compare screener vs. survivor narratives. J Health Commun 2019 Mar 29;24(2):141-155 [FREE Full text] [doi: 10.1080/10810730.2019.1587109] [Medline: 30924402]
34. Campbell T, Dunt D, Fitzgerald JL, Gordon I. The impact of patient narratives on self-efficacy and self-care in Australians with type 2 diabetes: stage 1 results of a randomized trial. Health Promot Int 2015 Sep 28;30(3):438-448. [doi: 10.1093/heapro/dat058] [Medline: 23985247]
35. Nan X, Dahlstrom MF, Richards A, Rangarajan S. Influence of evidence type and narrative type on HPV risk perception and intention to obtain the HPV vaccine. Health Commun 2015 Jul 25;30(3):301-308. [doi: 10.1080/10410236.2014.888629] [Medline: 25061715]
36. Downs JS, Murray PJ, Bruine de Bruin W, Penrose J, Palmgren C, Fischhoff B. Interactive video behavioral intervention to reduce adolescent females' STD risk: a randomized controlled trial. Soc Sci Med 2004 Oct;59(8):1561-1572. [doi: 10.1016/j.socscimed.2004.01.032] [Medline: 15279915]
37. Jones R, Hoover DR, Lacroix LJ. A randomized controlled trial of soap opera videos streamed to smartphones to reduce risk of sexually transmitted human immunodeficiency virus (HIV) in young urban African American women. Nurs Outlook 2013;61(4):205-215.e3 [FREE Full text] [doi: 10.1016/j.outlook.2013.03.006] [Medline: 23743482]
38. The CDC AIDS Community Demonstration Projects Research Group. Community-level HIV intervention in 5 cities: final outcome data from the CDC AIDS Community Demonstration Projects. Am J Public Health 1999 Mar;89(3):336-345. [doi: 10.2105/ajph.89.3.336] [Medline: 10076482]
39. Vaughan PW, Rogers EM, Singhal A, Swalehe RM. Entertainment-education and HIV/AIDS prevention: a field experiment in Tanzania. J Health Commun 2000 Feb 02;5 Suppl(sup1):81-100. [doi: 10.1080/10810730050019573] [Medline: 11010359]
40. Zhang J, Chen GM, Chock TM, Wang Y, Ni L, Schweisberger V. A psychophysiological study of processing HIV/AIDS public service announcements: the effects of novelty appeals, sexual appeals, narrative versus statistical evidence, and viewer's sex. Health Commun 2016 Jul 07;31(7):853-862. [doi: 10.1080/10410236.2015.1012629] [Medline: 26642917]
41. Zhang J, Chock TM, Chen GM, Schweisberger V, Yi W. Persuasiveness of HIV/AIDS public service announcements as a function of argument quality, personal relevance, and evidence form. Soc Behav Pers 2014 Nov 18;42(10):1603-1612. [doi: 10.2224/sbp.2014.42.10.1603]

42. Green M, Brock T. The narrative integration of personal and collective identity in social movements. In: Jacobs RN, editor. Narrative Impact: Social and Cognitive Foundations. Mahwah, NJ: Lawrence Erlbaum Associates; 2002:205-228.
43. Xin M, Viswanath K, Li AY, Cao W, Hu Y, Lau JT, et al. The effectiveness of electronic health interventions for promoting HIV-preventive behaviors among men who have sex with men: meta-analysis based on an integrative framework of design and implementation features. J Med Internet Res 2020 May 25;22(5):e15977 [FREE Full text] [doi: 10.2196/15977] [Medline: 32449685]
44. Zou H, Fan S. Characteristics of men who have sex with men who use smartphone geosocial networking applications and implications for HIV interventions: a systematic review and meta-analysis. Arch Sex Behav 2017 May;46(4):885-894. [doi: 10.1007/s10508-016-0709-3] [Medline: 27043837]
45. Pan S, Xu J, Han X, Zhang J, Hu QH, Chu ZX, et al. Internet-based sex-seeking behavior promotes HIV infection risk: a 6-year serial cross-sectional survey to MSM in Shenyang, China. Biomed Res Int 2016;2016:2860346-2860346 [FREE Full text] [doi: 10.1155/2016/2860346] [Medline: 28105415]
46. Lewnard JA, Berrang-Ford L. Internet-based partner selection and risk for unprotected anal intercourse in sexual encounters among men who have sex with men: a meta-analysis of observational studies. Sex Transm Infect 2014 Jun 11;90(4):290-296. [doi: 10.1136/sextrans-2013-051332] [Medline: 24518249]
47. Kwan TH, Lee SS. Bridging awareness and acceptance of pre-exposure prophylaxis among men who have sex with men and the need for targeting chemsex and HIV testing: cross-sectional survey. JMIR Public Health Surveill 2019 Jul 03;5(3):e13083 [FREE Full text] [doi: 10.2196/13083] [Medline: 31271148]
48. Muessig KE, Bien CH, Wei C, Lo EJ, Yang M, Tucker JD, et al. A mixed-methods study on the acceptability of using eHealth for HIV prevention and sexual health care among men who have sex with men in China. J Med Internet Res 2015 Apr 21;17(4):e100 [FREE Full text] [doi: 10.2196/jmir.3370] [Medline: 25900881]
49. Lau JT, Gu J, Tsui HY, Wang Z. Prevalence and associated factors of intention to participate in HIV voluntary counseling and testing for the first time among men who have sex with men in Hong Kong, China. Prev Med 2013 Dec;57(6):813-818. [doi: 10.1016/j.ypmed.2013.09.005] [Medline: 24045009]
50. Lau JTF, Li D, Wang Z, Lai CHY. Repeated HIV voluntary counseling and testing increased risk behaviors among men who have sex with men in China: a prospective cohort study. AIDS Behav 2015 Nov;19(11):1966-1977. [doi: 10.1007/s10461-015-1034-x] [Medline: 25863465]
51. Ziebland S, Herxheimer A. How patients' experiences contribute to decision making: illustrations from DIPEx (personal experiences of health and illness). J Nurs Manag 2008 May;16(4):433-439. [doi: 10.1111/j.1365-2834.2008.00863.x] [Medline: 18405260]
52. Wengraf T. Preparing lightly-structured depth interviews: a design for a BNIM-type biographic-narrative interview. In: Wengraf T, editor. Qualitative Research Interviewing: Biographic Narrative and Semi-structured Methods. London: Sage Publications; 2001:111-152.
53. Kreuter MW, Holmes K, Alcaraz K, Kalesan B, Rath S, Richert M, et al. Comparing narrative and informational videos to increase mammography in low-income African American women. Patient Educ Couns 2010 Dec;81 Suppl:S6-14 [FREE Full text] [doi: 10.1016/j.pec.2010.09.008] [Medline: 21071167]
54. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Analyzing Qualitative Data. London: Routledge; 2002:187-208.
55. Frank K. Rethinking risk, culture, and intervention in collective sex environments. Arch Sex Behav 2019 Jan 10;48(1):3-30. [doi: 10.1007/s10508-018-1153-3] [Medline: 29748787]
56. van Laer T, Feiereisen S, Visconti LM. Storytelling in the digital era: a meta-analysis of relevant moderators of the narrative transportation effect. J Bus Res 2019 Mar;96:135-146. [doi: 10.1016/j.jbusres.2018.10.053]
57. Brunyé TT, Ditman T, Mahoney CR, Augustyn JS, Taylor HA. When you and I share perspectives: pronouns modulate perspective taking during narrative comprehension. Psychol Sci 2009 Jan;20(1):27-32. [doi: 10.1111/j.1467-9280.2008.02249.x] [Medline: 19076318]
58. Miller-Day M, Hecht ML. Narrative means to preventative ends: a narrative engagement framework for designing prevention interventions. Health Commun 2013 Oct;28(7):657-670 [FREE Full text] [doi: 10.1080/10410236.2012.762861] [Medline: 23980613]
59. de Graaf A. The effectiveness of adaptation of the protagonist in narrative impact: similarity influences health beliefs through self-referencing. Hum Commun Res 2013 Oct 03;40(1):73-90. [doi: 10.1111/hcre.12015]
60. Gao MY, Wang S. Participatory communication and HIV/AIDS prevention in a Chinese marginalized (MSM) population. AIDS Care 2007 Jul 14;19(6):799-810. [doi: 10.1080/09540120601114832] [Medline: 17573601]
61. Wong C, Tang CS. Sexual practices and psychosocial correlates of current condom use among Chinese gay men in Hong Kong. Arch Sex Behav 2004 Apr;33(2):159-167. [doi: 10.1023/b:aseb.0000014330.67201.1b]
62. Gu J, Lau J, Tsui H. Psychological factors in association with uptake of voluntary counselling and testing for HIV among men who have sex with men in Hong Kong. Public Health 2011 May;125(5):275-282. [doi: 10.1016/j.puhe.2011.01.010] [Medline: 21419466]

63. Occa A, Suggs LS. Communicating breast cancer screening with young women: an experimental test of didactic and narrative messages using video and infographics. J Health Commun 2016 Jul 06;21(1):1-11. [doi: 10.1080/10810730.2015.1018611] [Medline: 26147625]
64. Shaffer VA, Owens J, Zikmund-Fisher BJ. The effect of patient narratives on information search in a web-based breast cancer decision aid: an eye-tracking study. J Med Internet Res 2013;15(12):e273 [FREE Full text] [doi: 10.2196/jmir.2784] [Medline: 24345424]

Abbreviations

HIV: human immunodeficiency virus
STI: sexually transmitted infections

Edited by R Kukafka; submitted 16.10.20; peer-reviewed by A Queiroz, B Serrano Zárate; comments to author 04.02.21; revised version received 01.06.21; accepted 05.07.21; published 16.09.21.

Please cite as:
Xin M, Coulson NS, Jiang CL, Sillence E, Chidgey A, Kwan NNM, Mak WWS, Goggins W, Lau JTF, Mo PKH
Web-Based Behavioral Intervention Utilizing Narrative Persuasion for HIV Prevention Among Chinese Men Who Have Sex With Men (HeHe Talks Project): Intervention Development
J Med Internet Res 2021;23(9):e22312
URL: https://www.jmir.org/2021/9/e22312
doi:10.2196/22312
PMID:34528889

Original Paper

A Web-Based Integrated Management Program for Improving Medication Adherence and Quality of Life, and Reducing Readmission in Patients With Atrial Fibrillation: Randomized Controlled Trial

Hui-Ling Hsieh[1], MSN; Chi-Wen Kao[2], PhD; Shu-Meng Cheng[3], MD, PhD; Yue-Cune Chang[4], PhD

[1]Graduate Institute of Medical Sciences, National Defense Medical Center, Taipei, Taiwan

[2]School of Nursing, National Defense Medical Center, Taipei, Taiwan

[3]Division of Cardiology, Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan

[4]Department of Mathematics, Tamkang University, Taipei, Taiwan

Corresponding Author:
Chi-Wen Kao, PhD
School of Nursing
National Defense Medical Center
No 161, Sec 6, Minchuan E Rd
Neihu District
Taipei, 11490
Taiwan
Phone: 886 2 8792 3100
Email: chiwenkao@ndmctsgh.edu.tw

Abstract

Background: Atrial fibrillation (AF) is related to a variety of chronic diseases and life-threatening complications. It is estimated that by 2050, there will be 72 million patients with AF in Asia, of which 2.9 million will have AF-associated stroke. AF has become a major issue for health care systems.

Objective: We aimed to evaluate the effects of a web-based integrated management program on improving coping strategies, medication adherence, and health-related quality of life (HRQoL) in patients with AF, and to detect the effect on decreasing readmission events.

Methods: The parallel-group, single-blind, prospective randomized controlled trial recruited patients with AF from a medical center in northern Taiwan and divided them randomly into intervention and control groups. Patients in the intervention group received the web-based integrated management program, whereas those in the control group received usual care. The measurement tools included the Brief Coping Orientation to Problems Experienced (COPE) scale, Medication Adherence Rating Scale (MARS), the three-level version of the EuroQoL five-dimension self-report questionnaire (EQ-5D-3L), and readmission events 2 years after initiating the intervention. Data were collected at 4 instances (baseline, 1 month, 3 months, and 6 months after initiating the intervention), and analyzed with generalized estimating equations (GEEs).

Results: A total of 231 patients were recruited and allocated into an intervention (n=115) or control (n=116) group. The mean age of participants was 73.08 (SD 11.71) years. Most participants were diagnosed with paroxysmal AF (171/231, 74%), and the most frequent comorbidity was hypertension (162/231, 70.1%). Compared with the control group, the intervention group showed significantly greater improvement in approach coping strategies, medication adherence, and HRQoL at 1, 3, and 6 months (all $P<.05$). In addition, the intervention group showed significantly fewer readmission events within 2 years (OR 0.406, $P=.03$), compared with the control group.

Conclusions: The web-based integrated management program can significantly improve patients' coping strategy and medication adherence. Therefore, it can empower patients to maintain disease stability, which is a major factor in improving their HRQoL and reducing readmission events within 2 years.

Trial Registration: ClinicalTrials.gov NCT04813094; https://clinicaltrials.gov/ct2/show/NCT04813094.

(J Med Internet Res 2021;23(9):e30107) doi:10.2196/30107

KEYWORDS

web-based program; atrial fibrillation; coping strategy; medication adherence; readmission; health-related quality of life

Introduction

Atrial fibrillation (AF) is related to a variety of chronic diseases and life-threatening complications [1]. Owing to progressive aging of the population and the force of unhealthy lifestyles, the global incidence of AF is expected to achieve a twofold increase in the next decade [2]. According to the 2010 Global Burden of Disease Study report, approximately 33 million people worldwide suffer from AF [3]. It is estimated that by 2050, there will be 72 million AF patients in Asia, of which 2.9 million will have AF-associated stroke [4]. Compared with Europe and the United States, the population of AF has increased more rapidly in Asia [5-7], suggesting that AF has become a major issue for Asian health care systems.

The asymptomatic nature of AF often makes patients unaware of the subsequent stroke [2]. Nearly 40% of AF patients suffer from cerebrovascular accidents for this reason [1,8,9]. AF is the most common cause of cardioembolic stroke, which is particularly severe in terms of high mortality and serious disability [2]. The complications of AF lead to frequent visits to the emergency department, hospitalization and readmission, and impaired quality of life [10-12].

Anticoagulation is the preferred medical intervention to prevent stroke in patients with AF [7]. Nonetheless, a European clinical trial showed that nearly 40% of patients with AF withdrew anticoagulants within 1 year following the initiation of the treatment [13]. In fact, compliance to anticoagulants decreases with time in patients with AF [14], which is a major factor contributing to stroke in this population [15,16]. As such, anticoagulant incompliance worsens the prognosis of AF in terms of increased readmission and mortality rates [17-19]. Ensuring anticoagulant adherence in patients with AF is an important issue for stroke prevention and disease recovery.

AF frequently occurs in patients with multiple concurrent chronic diseases including hypertension, ischemic heart disease, and heart failure, which increase the number of medications used. Patients taking several medications have many adverse effects and tend to be noncompliant to treatments [20]. Therefore, advanced and integrated disease management programs are essential for AF patients to achieve successful disease control [21]. Prior studies support the use of eHealth to assist patients in conducting advanced disease management. A few randomized controlled trials had investigated the impact of eHealth on medication adherence in patients with AF for 3 months [22,23]. Although the use of eHealth management significantly improved medication adherence at 3 months after intervention, the long-term effects on medication adherence remain to be investigated [24]. Again, integrated management aims to improve patients' disease knowledge and empower patients to manage their disease [9,24,25]. Integrated management needs the involvement of patients and multidisciplinary AF teams. Multidisciplinary professionals using technological support such as eHeath provide disease information and self-care skills to AF patients [9]. As a result, advanced and integrated disease management programs are expected to improve medication adherence and disease control in patients with AF.

In this study, we aimed to evaluate the effects of a web-based integrated management program on improving coping strategies, medication adherence, and health-related quality of life (HRQoL) in patients with AF, and to detect its effect on decreasing readmission events. Therefore, we have two hypotheses. The first hypothesis is that the AF patients receiving the web-based integrated management program in the intervention group would use more approach coping strategies and improve their medication adherence and HRQoL compared with the patients in the control group. The second hypothesis is that the patients in the intervention group would have fewer 2-year readmission events than the patients in the control group.

Methods

Design

This study was a prospective, single-blind, randomized controlled trial with repeated measurements to determine the effects of a 6-month web-based integrated management program on improving the coping strategy, medication adherence, and HRQoL, and decreasing 2-year readmission events in patients with AF. Data were collected from October 2018 to January 2021. We collected the data at 4 time points: baseline (prior to randomization, T0), 1 month after beginning the intervention (T1), after completing 3 months of the intervention (T2), and after completing 6 months of the intervention (T3).

Participants

Participants were recruited from the cardiovascular outpatient department at a medical center in northern Taiwan through convenience sampling. Patients were included if they met the following criteria: (1) diagnosed with AF by cardiologists, (2) receiving anticoagulant treatment, (3) aged above 20 years, (4) able to speak and read Taiwanese or Mandarin to understand and follow instructions, and (5) able to use a mobile phone or computer correctly. The exclusion criteria included the following: (1) diagnosed with mental disorders or (2) involved in other clinical trials.

In total, 324 patients were screened in the cardiovascular outpatient department, and 258 patients were eligible to be included. Among the eligible patients, 26 patients refused to participate. The remaining 232 participants were randomly assigned into 2 groups by blocked randomization using a web-based system [26]. Finally, 116 participants were assigned into the intervention group and 116 participants were assigned into the control group. The patients in the intervention group received the web-based integrated management program, whereas those in the control group received consultations and were coached thrice by a research nurse over telephone.

Ethical Considerations

This study was approved by the Institutional Review Board of the participating hospital (IRB: 2-107-05-081; ClinicalTrials.gov NCT04813094) and conducted in accordance with the principles of the Declaration of Helsinki revised in 2000. A research nurse explained the research process to each participant and obtained informed consent from them. We ensured that the participants were kept anonymous. In addition, participants were informed that they had the right to withdraw from the study at any time for any reason.

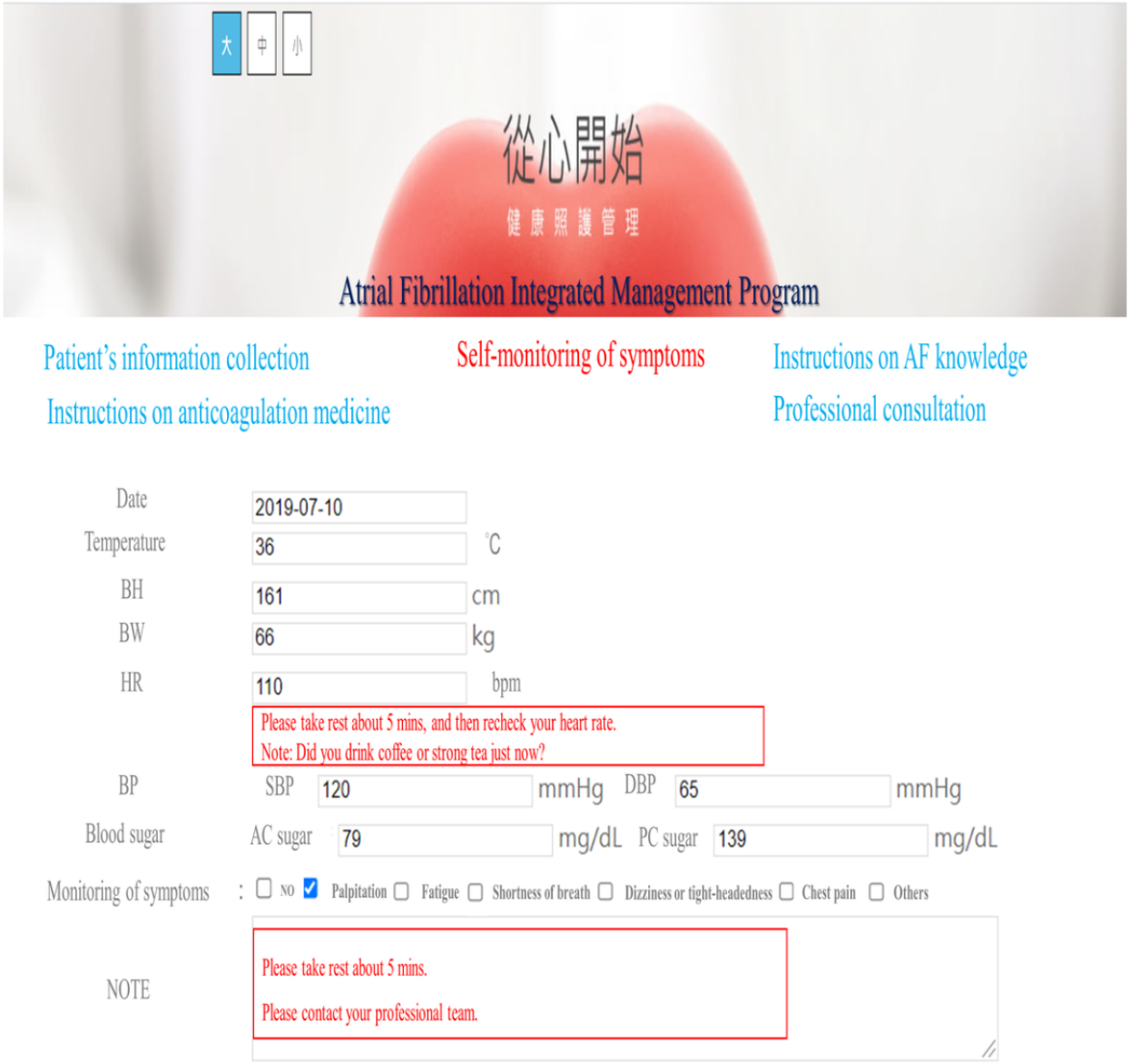
Study Interventions

The web-based integrated management program was conducted in the cardiovascular outpatient department. Participants had their own account and password to log in to the web-based program via mobile phones or computers. The program included five domains: patient information collection, instructions on AF knowledge, instructions on anticoagulation medicine, self-monitoring of symptoms, and professional consultation. In the patient information collection domain, the participants were able to provide and read their information, including age, gender, years of education, type of AF, modified European Heart Rhythm Association (mEHRA) classification, comorbidities, and use of anticoagulation medication. In the instructions on AF knowledge domain, participants were able to receive information about AF through texts and videos, including an introduction to AF, the risk factors and symptoms of AF, how to prevent stroke in patients with AF, and how to manage a healthy lifestyle. The videos were also available in Taiwanese and Mandarin. In the instructions on anticoagulation medicine domain, the participants were able to obtain descriptions about a variety of anticoagulant medicines. There was textual information about anticoagulant medicines, such as the effectiveness and adverse effects of these medicines, and precautions to be taken, as well as pictures of the corresponding medicines to serve as references for the participants. In the self-monitoring of symptoms domain, participants could provide record their symptoms every day, which could assist them in monitoring their disease progress and provide cardiologists with references to regulate their treatment. In the professional consultation domain, participants could receive consultations from multidisciplinary professionals on any issues related to AF at any time. All the participants had their own private consultations with clinical professionals. The research nurse also sent messages every day to monitor the participants' condition through the messaging function of this domain. When the participants had an emergency event, they could receive help to manage this situation through textual information or telephonic coaching (Figures 1 and 2).

After face-to-face presentations, we ensured that the participants fully understood the instructions for using this web-based program by the return demonstration method. We reviewed their understanding of the program at the next outpatient visit. In addition, we instructed participants on managing AF and anticoagulant treatment through the web-based program and provided the AF management manual to them.

Figure 1. Screenshot of the Atrial Fibrillation Integrated Management Program. BH: body height; BW: body weight; po: per os (Latin term meaning orally) QD: qaque die (Latin term meaning one a day).

Figure 2. Screenshot showing self-monitoring of symptoms. AF: atrial fibrillation; BH: body height; BW: body weight; HR: heart rate; BP: blood pressure; DBP: diastolic blood pressure; SBP: systolic blood pressure.



Control Group

Patients in the control group received standard instructions. The AF management manual was provided to them with explanations. In addition, we provided telephonic coaching thrice, which taught participants how to manage their disease at 1 month, 3 months, and 6 months after random assignment.

Measurements

The Clinical and Demographic Information Form

Demographic and clinical characteristics, including age, gender, education, marital status, type of AF, mEHRA classification, comorbidities, and use of anticoagulation medicine, were recorded before random assignment. The mEHRA classification was used to categorize AF patients according to their symptom severity. The mEHRA classification has five levels including class I, none; class IIa, mild; class IIb, moderate; class III, severe; and class IV, disabling [9].

Coping Strategies

The participants' coping strategies were measured using the Brief Coping Orientation to Problems Experienced (COPE) scale [27]. This 28-item instrument contains 2 subscales: approach coping (8 items) and avoidance coping (6 items). The approach coping strategies show how individuals actively seek resources to deal with their health problems (active coping, planning, positive reframing, acceptance, humor, religion, and using instrumental and emotional support). The avoidance coping strategies show how individuals attempt to divert attention away from events (self-distraction, denial, venting, substance use, behavioral disengagement, and self-blame) [28]. A higher score indicated that the patient often uses this coping strategy. In this study, the Brief COPE was translated into Chinese. The content validity for the Chinese version of the Brief COPE was examined by five cardiology experts. The

content validity index was 0.83. A previous study reported that the Cronbach α of each item ranged between .50 and .90 [27].

Medication Adherence

The participants' medication adherence was measured by the Medication Adherence Rating Scale (MARS) [29]. This instrument includes three domains: medication adherence behavior, attitude toward taking medication, and attitude toward the adverse effects of medication. Each item has two options. Participants choose an answer based on their prescription. The total possible score ranges from 0 to 10. A higher score indicates better medication adherence. In this study, the MARS was translated into its Chinese version. The content validity of the Chinese version was examined by five cardiology experts. The content validity index was 0.92. The Cronbach α of the MARS was .75 [29].

HRQoL Measurement

The participants' HRQoL was measured by the three-level version of the EuroQol five-dimension self-report questionnaire (EQ-5D-3L). The instrument includes 2 subscales: the EQ-5D descriptive system (5 items) and EuroQol visual analog scale (EQ-VAS). The EQ-5D descriptive system had five domains: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each item was rated at 3 levels: no problems at level 1 and extreme problems at level 3. The scores were converted into a single summary score. The EQ-VAS was a 20 cm visual analog scale to let participants self-assess their health status. The range of the EQ-VAS was from 0 to 100. A high score indicated the best state of health. We used the Chinese version of the EQ-5D-3L, which has adequate validity and reliability [30].

Readmission Events Within 2 Years

After recruiting the patients in the study, we followed their readmission events for 2 years. We collected these data through chart reviews of each participant.

Data Analysis

Statistical analyses were conducted using SPSS (version 23.0; IBM Corp). Categorical variables were reported as frequencies and percentages, and continuous variables with normal distributions were reported as means with SDs. The initial differences between groups for demographic characteristics and the baseline scores of each scale were examined with independent *t* and chi-square tests. The effects of the web-based integrated management program on the coping strategy, medication adherence, HRQoL, and readmission events were examined using a generalized estimating equation (GEE) [31]. The significance level was defined by a two-tailed *t* test with $P<.05$.

We used univariate logistic regression to assess the association between patient characteristics, including group assignment, and readmission events within 2 years and identify the predictor variables. The readmission event within 2 years was coded as the dependent variable with 1 denoting a readmission event within 2 years after intervention and 0 denoting no readmission event within 2 years after intervention. The predictor variables with $P<.20$ were eligible for inclusion in the multivariable logistic regression model for measuring the outcome. In the multivariable logistic regression model, predictor variables with two-tailed *t* tests and $P<.05$ were considered significant [32]. We used the GPower (Version 3.1, Heinrich-Heine-Universität [33]) procedure with an α value of .05, a medium effect size of 0.25, and sample size of 231 to conduct a post hoc statistical power analysis [33]. Finally, the statistical power of this study was greater than 0.8.

Results

Participant Characteristics

A total of 258 patients were recruited from the outpatient department of the study hospital according to the inclusion criteria. Of these patients, 26 refused to participate. At the end of the study, the complete data of 115 participants in the intervention group and 116 participants in the control group were included for the statistical analysis. One participant in the intervention group was lost to follow-up owing to emigration (Figure 3). The mean age of the participants was 73.08 (SD 11.71) years. Half of the participants were male (116/231, 50.2%), and over half of the participants were married (155/231, 67.1%). Most participants were diagnosed with paroxysmal AF (171/231, 74%), and the most frequent comorbidity was hypertension (162/231, 70.1%). Assessing the classification of AF symptoms through mEHRA indicated that 44.6% (103/231) of the participants were class I, 22.9% (53/231) were class IIa, 14.7% (34/231) were class IIb, and 17.7% (41/231) were class III. Approximately 36.4% (84/231) of the participants received rivaroxaban, and 27.3% (63/231) received apixaban. There were no significant differences in the demographic and clinical characteristics, and the baseline scores of each scale between the two groups (Table 1).

Figure 3. Flow diagram showing the inclusion of patients with atrial fibrillation in the two groups for the randomized control trial.

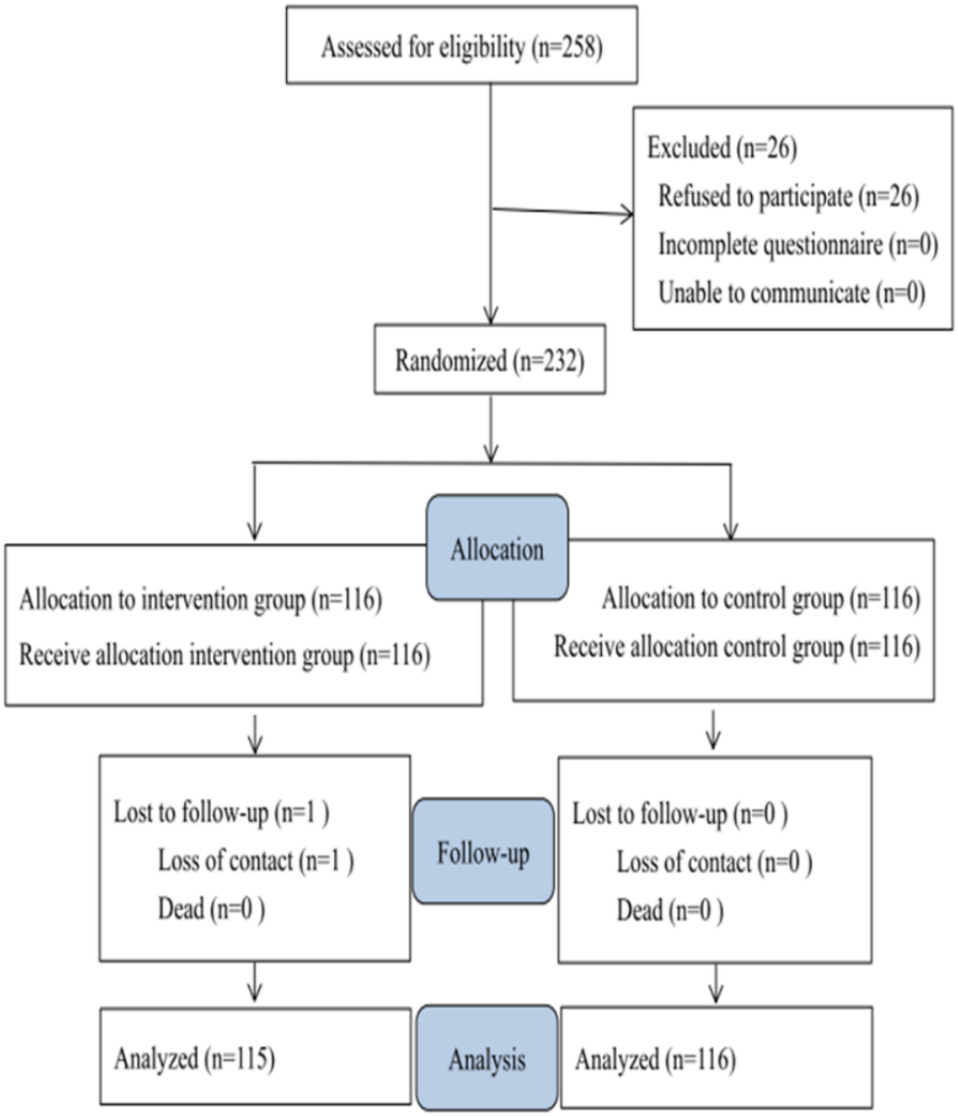


Table 1. Baseline characteristics of the study participants (N=231) and differences between groups.

Characteristic	Total (N=231)	Intervention group (n=115)	Control group (n=116)	*P* value
Age (years), mean (SD)	73.08 (11.71)	71.57 (13.16)	74.57 (9.9)	.052
Gender, n (%)				.19
Male	116 (50.2)	63 (54.8)	53 (45.7)	
Female	115 (49.8)	52 (45.2)	63 (54.3)	
Education (years), n (%)				.19
<6 years	98 (42.4)	42 (36.5)	56 (48.3)	
6-12 years	84 (36.4)	46 (40)	38 (32.8)	
>12 years	49 (21.2)	27 (23.5)	22 (19)	
Marital status, n (%)				.75
Single	6 (2.6)	4 (3.5)	2 (1.7)	
Married	155 (67.1)	78 (67.8)	77 (66.4)	
Divorced	7 (3)	4 (3.5)	3 (2.6)	
Widowed	63(27.3)	29 (25.2)	34 (29.3)	
Type of AF,[a] n (%)				.17
Paroxysmal AF	171 (74)	85 (73.9)	86 (74.1)	
Persistent AF	21 (9.1)	14 (12.2)	7 (6)	
Permanent AF	39 (16.9)	16 (13.9)	23 (19.8)	
mEHRA,[b] n (%)				.88
mEHRA I	103 (44.6)	52 (45.2)	51 (44)	
mEHRA IIa	53 (22.9)	28 (23.4)	25 (21.6)	
mEHRA IIb	34 (14.7)	15 (13)	19 (16.4)	
mEHRA III	41 (17.7)	20 (17.4)	21 (18.1)	
mEHRA IV	0 (0)	0 (0)	0 (0)	
Comorbidities, n (%)				
Hypertension	162 (70.1)	77 (67)	85 (73.3)	.32
Diabetes mellitus	61 (26.4)	26 (22.6)	35 (30.2)	.23
MI[c]	14 (6.1)	7 (6.1)	7 (6)	.99
Valve heart disease	77 (33.3)	36 (31.3)	41 (35.3)	.58
Valve heart disease	77 (33.3)	36 (31.3)	41 (35.3)	.58
Heart failure	78 (33.8)	39 (33.9)	39 (33.6)	.99
CAD[d]	85 (36.8)	41 (35.7)	44 (37.9)	.79
Hyperthyroid	14 (6.1)	5 (4.3)	9 (7.8)	.41
Cancer	10 (4.3)	5 (4.3)	5 (4.3)	.99
Medication, n (%)				
Ticlopidine hydrochloride	1 (0.4)	1 (0.9)	0 (0)	.50
Dipyridamole	6 (2.6)	4 (3.5)	2 (1.7)	.45
Aspirin	17 (7.4)	10 (8.7)	7 (6)	.46
Apixaban	63 (27.3)	32 (27.8)	31 (26.7)	.88
Edoxaban	47 (20.3)	28 (24.3)	19 (16.4)	.14
Clopidogrel	10 (4.3)	3 (2.6)	7 (6)	.33
Dabigatran	7 (3)	3 (2.6)	4 (3.4)	.99

Characteristic	Total (N=231)	Intervention group (n=115)	Control group (n=116)	*P* value
Rivaroxaban	84 (36.4)	38 (33)	46 (39.7)	.34
Questionnaires, mean (SD)				
Brief COPE[e]				
Approach coping	26.1 (7.68)	25.6 (7.58)	26.6 (7.77)	.33
Avoidance coping	14.32 (6.04)	15.02 (6.6)	13.63 (5.36)	.08
MARS[f]	7.07 (1.80)	7.17 (1.79)	6.97 (1.80)	.40
HRQoL[g]				
EQ-5D[h]	0.62 (0.12)	0.61 (0.12)	0.63 (0.12)	.16
EQ-VAS[i]	81.22 (6.48)	80.57 (7.079)	81.85 (5.79)	.13

[a]AF: atrial fibrillation.

[b]mEHRA: modified European Heart Rhythm Association.

[c]MI: myocardial infarction.

[d]CAD: coronary arterial disease.

[e]Brief COPE: Brief Coping Orientation to Problems Experienced.

[f]MARS: Medication Adherence Rating Scale.

[g]HRQoL: health-related quality of life.

[h]EQ-5D: EuroQol five-dimension self-report questionnaire.

[i]EQ-VAS: EuroQol visual analog scale.

Effects on Coping Strategies

The GEE model analysis showed a significant difference in the scores of the approach coping subscale between the intervention and control groups at 1 month (β=.792; 95% CI 0.143-1.172; *P*<.001), 3-months (β=1.297; 95% CI 0.441-2.153; *P*=.003), and 6 months (β=1.902; 95% CI 0.882-2.922; *P*<.001) (Table 2 and Figure 4).

In addition, the intervention group showed a significantly higher decrease in the scores of the avoidance coping subscale at 1 month (β=–2.284; 95% CI –2.885 to –1.683; *P*<.001), 3 months (β=–2.602; 95% CI –3.511 to –1.694; *P*<.001), and 6 months (β=–2.982; 95% CI –4.096 to –1.869; *P*<.001) compared with the control group (Table 2 and Figure 4). These findings indicated that the web-based integrated management program helped participants to use more approach coping strategies and less avoidance coping strategies in managing their AF-related issues.

Table 2. Generalized estimating equation analysis of the intervention effect on the Brief Coping Orientation to Problems Experienced scale (N=231).

Variable	β	95% CI	SE[a]	χ^2	*P* value
Approach coping					
Group (intervention)[b]	–.995	–2.966 to 0.976	1.006	0.979	.32
Time effects					
Time (1-month, T1)[c]	.129	–0.070 to –0.329	0.102	1.618	.20
Time (3-months, T2)	.155	–0.479 to 0.789	0.323	0.230	.63
Time (6-months, T3)[c]	–.241	–0.989 to 0.506	0.381	0.401	.53
Interaction effects					
Intervention × T1[d]	.792	0.143 to 1.172	0.194	16.778	<.001
Intervention × T2[d]	1.297	0.441 to 2.153	0.437	8.816	.003
Intervention × T3[d]	1.902	0.882 to 2.922	0.520	13.362	<.001
Avoidance coping					
Group (intervention)[b]	1.388	–0.157 to 2.933	0.788	3.102	.078
Time effects					
Time (1 month, T1)[c]	2.388	1.856 to 2.920	0.272	77.383	<.001
Time (3 months, T2)	2.724	1.966 to 3.483	0.387	49.552	<.001
Time (6 months, T3)[c]	3.043	2.240 to 3.846	0.410	55.215	<.001
Interaction effects					
Intervention × T1[d]	–2.284	–2.885 to –1.683	0.307	55.471	<.001
Intervention × T2[d]	–2.602	–3.511 to –1.694	0.464	31.505	<.001
Intervention × T3[d]	–2.982	–4.096 to –1.869	0.568	27.554	<.001

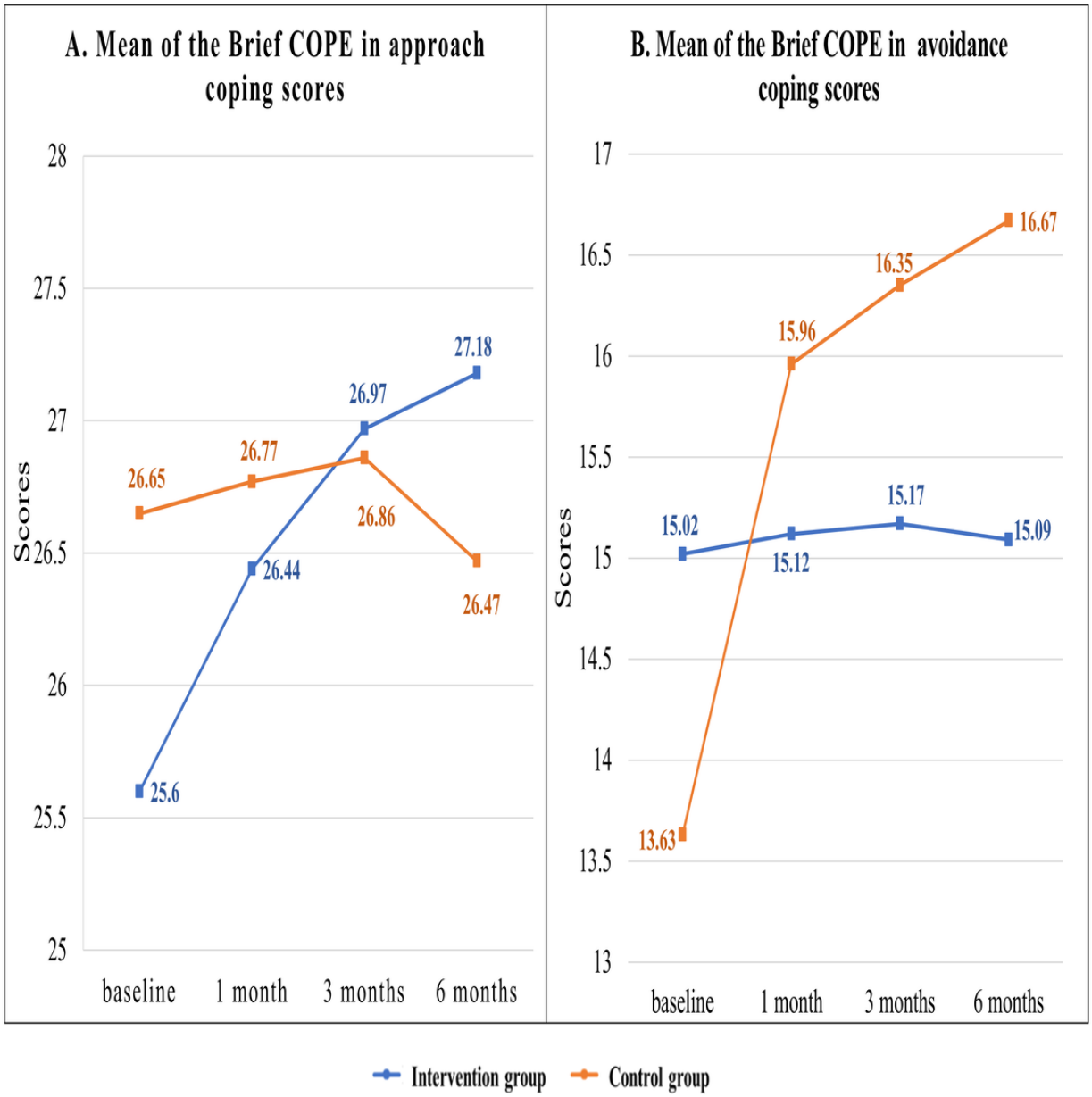
[a]SE: standard error.

[b]Reference group: control group.

[c]Reference group: time (baseline).

[d]Reference group: group (control) × time (baseline).

Figure 4. Mean scores of the approach and avoidance coping subscales for participants in the intervention and control groups. Brief COPE: Brief Coping Orientation to Problems Experienced.



Effects on Medication Adherence

The GEE model analysis showed significant differences in the scores of the MARS between the intervention and control groups at 1 month (β=.326; 95% CI 0.034-0.618; *P*=.03), 3 months (β=.595; 95% CI 0.24-0.95; *P*=.001), and 6 months (β=.606; 95% CI 0.253-0.96; *P*=.001) (Table 3 and Figure 5). These results indicated that after receiving the web-based integrated management program, the AF patients showed significantly better medication adherence than that shown by the patients in the control group.

Table 3. Generalized estimation equation analysis of the intervention effect on Medication Adherence Rating Scale (N=231).

Variable	β	95% CI	SE[a]	χ^2	*P* value
Group (intervention)[b]	.2	–0.262 to 0.661	0.2354	0.720	.40
Time effects					
Time (1-month, T1)[c]	.5	0.337 to 0.663	0.831	36.172	<.001
Time (3-months, T2)	.422	0.181 to 0.664	0.1232	11.752	.001
Time (6-months, T3)[c]	.716	0.499 to 0.932	0.1103	42.106	<.001
Interaction effects					
Intervention × T1[d]	.326	0.034 to 0.618	0.149	4.784	.03
Intervention × T2[d]	.595	0.240 to 0.950	0.181	10.818	.001
Intervention × T3[d]	.606	0.253 to 0.960	0.180	11.306	.001

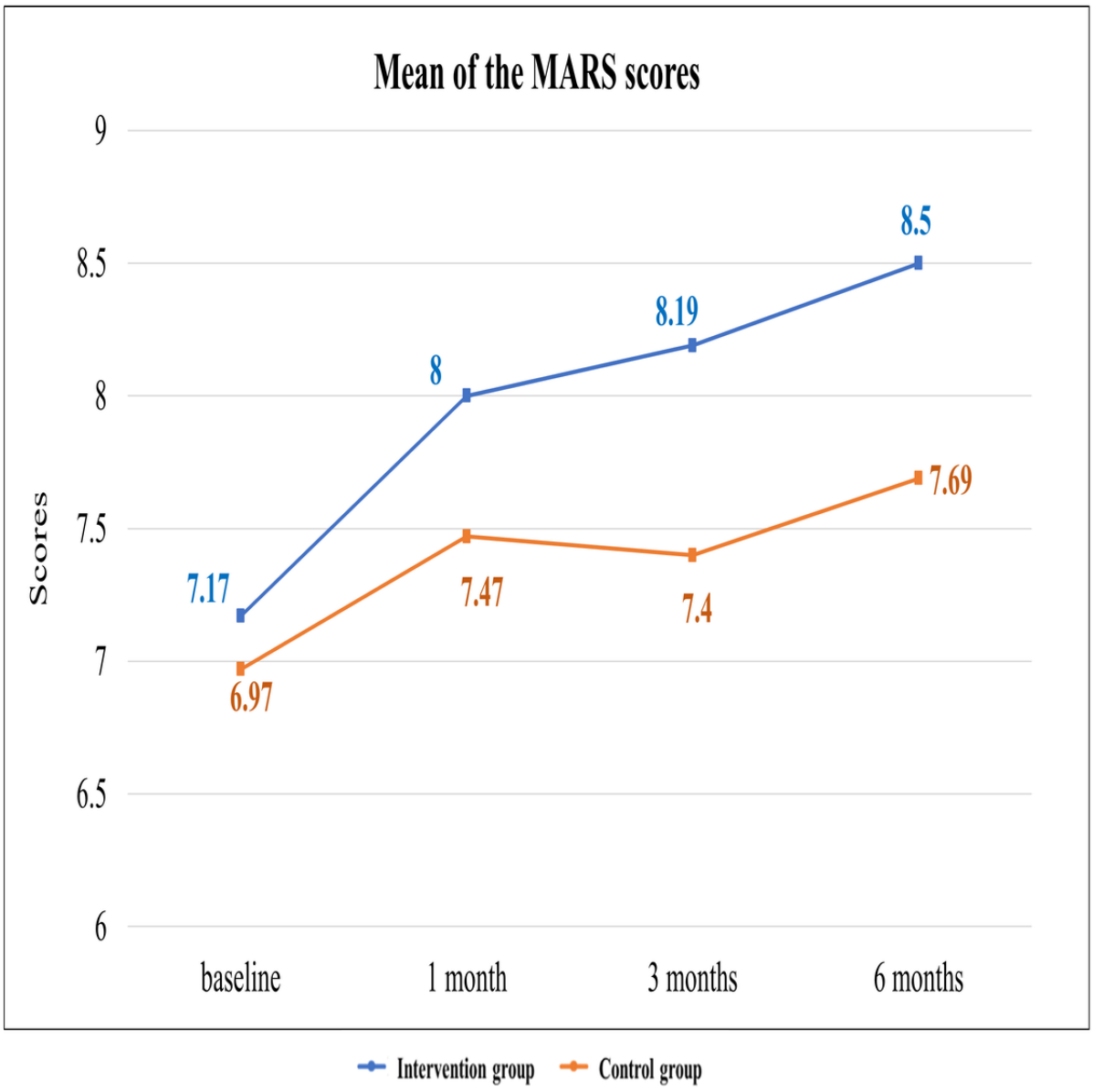
[a]SE: standard error.

[b]Reference group: control group.

[c]Reference group: time (baseline).

[d]Reference group: group (control) × time (baseline).

Figure 5. Mean scores of the Medication Adherence Rating Scale for participants in the intervention and control groups. MARS: Medication Adherence Rating Scale.



Effects on HRQoL

The GEE model analysis showed no significant difference in the EQ-5D scores between the 2 groups at 1 month (β=.042; 95% CI –0.014 to 0.980; *P*=.14). However, the patients in the intervention group showed significantly higher EQ-5D scores at 3 months (β=.082; 95% CI 0.028-0.137; *P*=.003) and 6 months (β=.19; 95% CI 0.13-0.25; *P*<.001) than the patients in the control group (Table 4 and Figure 6).

In addition, the GEE model analysis revealed significant differences between the 2 groups in the EQ-VAS scores at 1 month (β=2.071; 95% CI 0.551-3.592; *P*=.008), 3 months (β=3.838; 95% CI 2.169-5.506; *P*<.001), and 6 months (β=5.782; 95% CI 3.837-7.727; *P*<.001) (Table 4 and Figure 6). These results indicated that the AF patients receiving the web-based integrated management program showed a significantly higher HRQoL improvement than that shown by the patients in the control group.

Table 4. Generalized estimating equation analysis of the intervention effect on health-related quality of life (N=231).

Variable	β	95% CI	SE[a]	χ^2	*P* value
EQ-5D score[b]					
Group (intervention)[c]	–.022	–0.053 to 0.010	0.016	1.717	.16
Time effects					
Time (1-months, T1)[d]	–.020	–0.055 to 0.015	0.018	1.255	.26
Time (3-months, T2)[d]	–.012	–0.041 to 0.017	0.015	0.642	.42
Time (6-months, T3)[d]	–.010	–0.059 to 0.040	0.025	0.153	.70
Interaction effects					
Intervention × T1[e]	.042	–0.014 to 0.098	0.285	2.168	.14
Intervention × T2[e]	.082	0.028 to 0.137	0.028	8.820	.003
Intervention × T3[e]	.190	0.130 to 0.250	0.031	38.222	<.001
EQ-VAS[f] score					
Group (intervention)[c]	–1.280	–2.941 to 0.382	0.848	2.278	.13
Time effects					
Time (1-months, T1)[d]	–1.802	–2.918 to –0.685	0.570	10.008	.002
Time (3-months, T2)[d]	–1.672	–2.800 to –0.545	0.575	8.447	.004
Time (6-months, T3)[d]	–1.043	–2.353 to 0.267	0.668	2.436	.12
Interaction effects					
Intervention × T1[e]	2.071	0.551 to 3.592	0.776	7.128	.008
Intervention × T2[e]	3.838	2.169 to 5.506	0.851	20.316	<.001
Intervention × T3[e]	5.782	3.837 to 7.727	0.992	33.946	<.001

[a]SE: standard error.

[b]EQ-5D: EuroQol five-dimension self-report questionnaire.

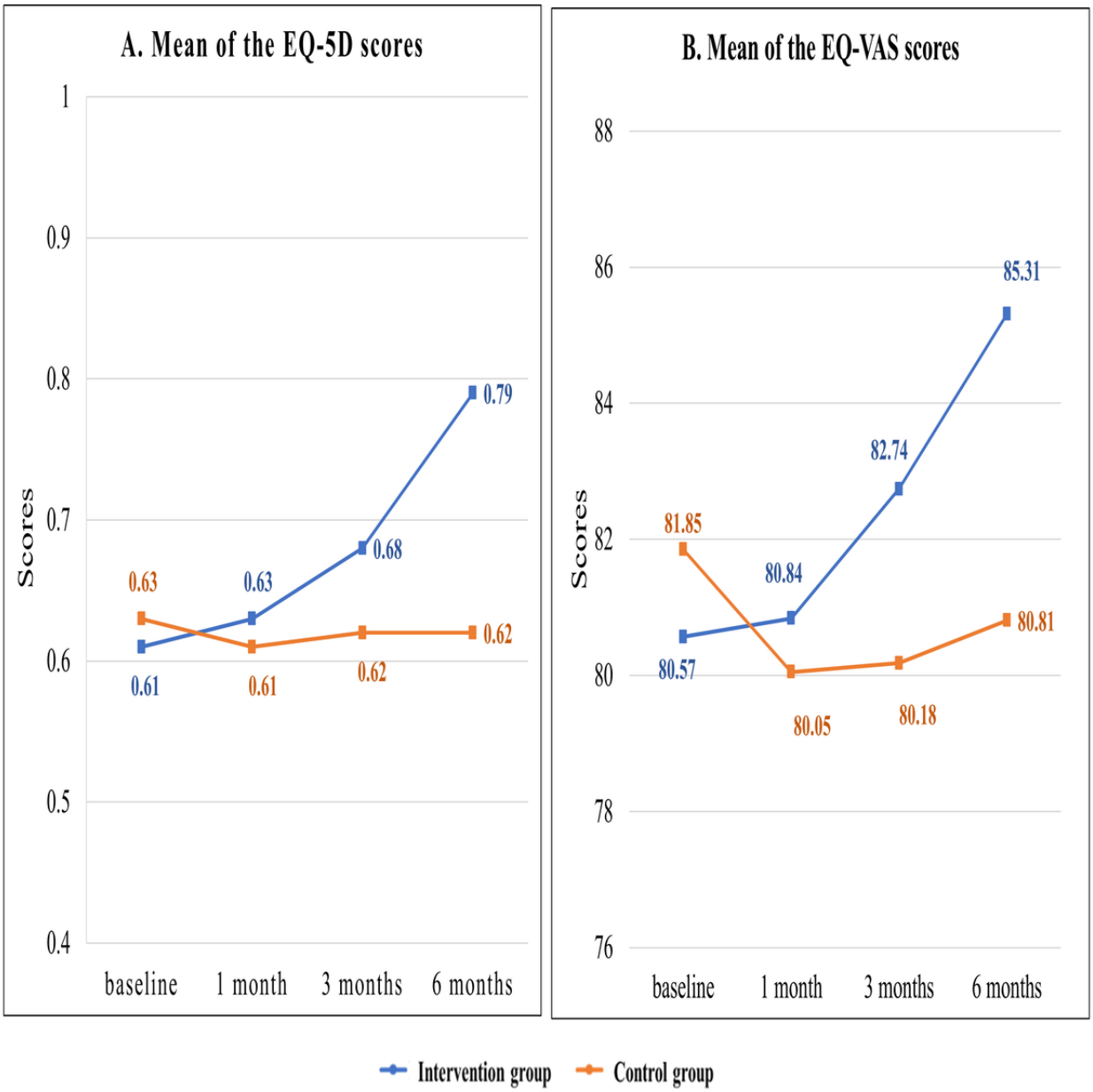
[c]Reference group: control group.

[d]Reference group: time (baseline).

[e]Reference group: group (control) × time (baseline).

[f]EQ-VAS: EuroQol visual analog scale.

Figure 6. Mean scores of the EuroQoL five-dimension self-report questionnaire and EuroQoL visual analog scale for participants in the intervention and control groups. EQ-5D: EuroQoL five-dimension self-report questionnaire; EQ-VAS: EuroQoL visual analog scale.



Effects on Readmission Events Within 2 Years

During the study, we followed the readmission events within 2 years after intervention. A total of 34 readmission events occurred. There were 11 readmission events in the intervention group and 23 in the control group. The predictor variables with $P<.20$ were based on univariate logistic regression. The predictor variables meeting the criteria were patients aged over 65 years, the mEHRA IIa class, comorbidities such as valvular heart disease and heart failure, receiving clopidogrel, and group assignment (Table 5). According to the multivariate logistic regression model, the intervention group showed significantly fewer readmission events within 2 years (OR 0.406; 95% CI 0.178-0.926; $P=.03$) compared with the control group (Table 5). These results indicated that the patients with AF receiving the web-based integrated management program had a lower probability of readmission within 2 years, compared with the control group.

Table 5. Association between the web-based integrated management program and readmission events within 2 years (N=231).

Characteristic	Univariate			Multivariate		
	OR[a] (%)	95% CI	*P* value	OR (%)	95% CI	*P* value
Gender: male	1.136	0.548-2.356	.73	—[b]	—	—
Aged over 65 years	0.574	0.253-1.300	.18	0.608	0.240 to –1.542	.30
Type of AF[c]						
Persistent AF	0.929	0.255-3.382	.91	—	—	—
Permanent AF	0.820	0.294-2.291	.71	—	—	—
mEHRA[d]						
mEHRA IIa	0.444	0.141-1.402	.17	0.378	0.112-1.269	.16
mEHRA IIb	0.725	0.225-2.340	.59	—	—	—
mEHRA III	1.754	0.720-4.272	.22	—	—	—
Comorbidities						
Hypertension	1.026	0.462-2.280	.95	—	—	—
Diabetes mellitus	1.004	0.440-2.291	.99	—	—	—
MI[e]	0.964	0.206-4.509	.96	—	—	—
Valvular heart disease	1.982	0.947-4.416	.07	2.200	0.994-4.870	.05
Heart failure	2.230	1.067-4.659	.03	1.907	0.851-4.272	.12
CAD[f]	1.075	0.508-2.275	.85	—	—	—
Hyperthyroidism	1.038	0.222-4.857	.96	—	—	—
Cancer	1.477	0.300-7.271	.63	—	—	—
Dyslipidemia	1.123	0.530-2.379	.76	—	—	—
Medication						
Ticlopidine hydrochloride	0	—	.99	—	—	—
Dipyridamole	1.164	0.132-10.279	.89	—	—	—
Aspirin	0.343	0.044-2.674	.31	—	—	—
Apixaban	1.132	0.508-2.525	.76	—	—	—
Edoxaban	1.247	0.524-2.965	.62	—	—	—
Clopidogrel	4.244	1.131-15.927	.03	4.240	0.843-21.324	.08
Dabigatran	0	—	.99	—	—	—
Rivaroxaban	0.693	0.314-1.529	.36	—	—	—
Intervention group[g]	0.428	0.198-0.925	.03	0.406	0.178-0.926	.03

[a]OR: odds ratio.

[b]Not applicable.

[c]AF: atrial fibrillation. Reference group: Paroxysmal AF.

[d]mEHRA: modified European Heart Rhythm Association. Reference group: mEHRA I.

[e]MI: myocardial infarction.

[f]CAD: coronary arterial disease.

[g]Reference group: control group.

Discussion

Principal Findings

The main findings of this study were that the web-based integrated management program significantly improved coping strategies, medication adherence, and HRQoL in patients with AF after 1, 3, and 6 months of intervention. In addition, the AF patients receiving the web-based integrated management program had a lower probability of readmission within 2 years after intervention compared with the patients not receiving this program.

The findings of this study identified that the AF patients receiving this program increased the usage of approach coping strategies and decreased the usage of avoidance coping strategies, as opposed to the patients in the control group. Individuals who used approach coping strategies were able to actively solve problems and conduct positive reappraisals [34]. As the patients were able to receive information immediately and appropriately through this web-based program, they became more confident in solving their problems. On the other hand, the patients in the control group received the manual for AF management to address their problems. They always complained that they did not have sufficient disease information and immediate discussions with clinical professionals when they faced problems. They resorted to using more avoidance coping strategies. Although avoidance coping strategies may be useful in temporarily reducing stress, they increase suffering and mortality, and decrease medication adherence and quality of life [35-37].

Furthermore, in our study, patients receiving the web-based program showed significant improvements in medication adherence and HRQoL, which resulted from the improved disease knowledge and empowerment to solve their problems. Wang et al pointed out three reasons why knowledge may improve medication adherence in patients with AF; first, they become more knowledgeable about the importance of anticoagulant treatment; second, they develop the ability to detect the adverse effects of anticoagulant medicines; third, they received more self-care information about managing the symptoms of AF and adverse effects of the medicines [38]. In addition, Ammenwerth et al suggested that integrated management with eHealth schemes can increase motivated lifestyles and adherence to medical recommendations, thereby improving the health status and HRQoL [39]. Our findings were consistent with those of previous studies [40,41]. All the findings proved that patients were able to manage their treatment and lifestyle by themselves through a web-based program providing positive coping strategies and sufficient disease knowledge and could therefore improve their medication adherence and HRQoL.

To test the second hypothesis that the patients in the intervention group would experience fewer 2-year readmission events than those in the control group, we followed the readmission status of AF patients for 2 years. Our findings showed significantly fewer readmission events within 2 years in patients who received the web-based integrated management program compared to those in the control group. Marcolino et al proposed that implementing web-based intervention programs can improve communication between patients and medical providers and provide assistance in disease management to promote health recovery of patients [42]. In our study, through the web-based integrated management program, direct communication between the patients/their families and health professionals was possible when the patients went back home. A randomized control trial involving patients with heart failure showed that using a remote medication monitoring system can greatly improve patients' medication adherence and reduce all-cause readmission events. It reduced the risk of readmission by 80% [43]. Our findings were consistent with these previous ones [44,45]. Because of the web-based integrated management program, we empowered patients to change their coping strategies and increase their compliance to medical treatment. Ultimately, it improved the HRQoL and long-term beneficial clinical effects in patients with AF in terms of reducing 2-year readmission events.

Conclusions

The study provides useful information for health care professionals regarding web-based integrated management programs for optimizing AF management. The most important findings of the study are summarized as follows: The web-based integrated management program can significantly improve patients' coping strategies and medication adherence, and help patients maintain disease stability, which has a major influence on improving the HRQoL and reducing adverse clinical events and readmission episodes. Overall, the program may improve the quality of patient care while reducing medical costs.

Limitations

Some limitations of this study should be considered. First, the sample may not reflect the entire AF population owing to the disease severity of most patients being mild or moderate. If AF patients had severe symptoms, they were hospitalized to receive medical interventions. Therefore, most of the AF patients in the community only had mild or moderate symptoms. However, they still had very limited knowledge on self-managing their symptoms. After receiving this program, they were able to take care of themselves, leading to decreased readmission events. Second, the single-center design limited the generalization of the findings. Third, factors that affect medication adherence and readmission, such as social support and economic status, were not included in the present study. As there were no differences in the educational levels and marital statuses of the patients in the intervention and control groups, we assumed that both groups of patients had similar social support levels and economic statuses. Nevertheless, future studies should examine these two factors and control their influence.

Acknowledgments

This study was supported by a grant from the Ministry of Science and Technology, Taiwan (MOST grant 107-2314-B016-013-MY3). This funding agency had no influence on the study design, data collection or analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 98 KB - jmir_v23i9e30107_app1.pdf]

References

1. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, ESC Scientific Document Group. 2020 ESC guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association of Cardio-Thoracic Surgery (EACTS). Eur Heart J 2020 Aug;42(5):373-498. [doi: 10.1093/eurheartj/ehaa612] [Medline: 32860505]
2. Pistoia F, Sacco S, Tiseo C, Degan D, Ornello R, Carolei A. The epidemiology of atrial fibrillation and stroke. Cardiol Clin 2016 May;34(2):255-268. [doi: 10.1016/j.ccl.2015.12.002] [Medline: 27150174]
3. Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ, et al. Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 study. Circulation 2014 Feb;129(8):837-847 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.113.005119] [Medline: 24345399]
4. Chiang CE, Wang KL, Lip GYH. Stroke prevention in atrial fibrillation: an Asian perspective. Thromb Haemost 2014 May;111(5):789-797. [doi: 10.1160/TH13-11-0948] [Medline: 24500243]
5. Miyasaka Y, Barnes ME, Gersh BJ, Cha SS, Bailey KR, Abhayaratna WP, et al. Secular trends in incidence of atrial fibrillation in Olmsted County, Minnesota, 1980 to 2000, and implications on the projections for future prevalence. Circulation 2006 Jul;114(2):119-125. [doi: 10.1161/CIRCULATIONAHA.105.595140] [Medline: 16818816]
6. Krijthe BP, Kunst A, Benjamin EJ, Lip GYH, Franco OH, Hofman A, et al. Projections on the number of individuals with atrial fibrillation in the European Union, from 2000 to 2060. Eur Heart J 2013 Sep;34(35):2746-2751 [FREE Full text] [doi: 10.1093/eurheartj/eht280] [Medline: 23900699]
7. Chiang CE, Wang KL, Lin SJ. Asian strategy for stroke prevention in atrial fibrillation. Europace 2015 Oct;17 Suppl 2:ii31-ii39. [doi: 10.1093/europace/euv231] [Medline: 26842113]
8. Chiang CE, Wu TJ, Ueng KC, Chao TF, Chang KC, Wang CC, et al. 2016 guidelines of the Taiwan Heart Rhythm Society and the Taiwan Society of Cardiology for the management of atrial fibrillation. J Formos Med Assoc 2016 Nov;115(11):893-952 [FREE Full text] [doi: 10.1016/j.jfma.2016.10.005] [Medline: 27890386]
9. Kirchhof P, Benussi S, Kotecha D, Ahlsson A, Atar D, Casadei B, et al. 2016 ESC guidelines for the management of atrial fibrillation developed in collaboration with EACTS. Eur Heart J 2016 Oct;37(38):2893-2962. [doi: 10.1093/eurheartj/ehw210] [Medline: 27567408]
10. Rienstra M, Lubitz SA, Mahida S, Magnani JW, Fontes JD, Sinner MF, et al. Symptoms and functional status of patients with atrial fibrillation: state of the art and future research opportunities. Circulation 2012 Jun;125(23):2933-2943 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.111.069450] [Medline: 22689930]
11. Lippi G, Sanchis-Gomar F, Cervellin G. Global epidemiology of atrial fibrillation: an increasing epidemic and public health challenge. Int J Stroke 2021 Feb;16(2):217-221. [doi: 10.1177/1747493019897870] [Medline: 31955707]
12. Son YJ, Baek KH, Lee SJ, Seo EJ. Health-related quality of life and associated factors in patients with atrial fibrillation: an integrative literature review. Int J Environ Res Public Health 2019 Aug;16(17):3042 [FREE Full text] [doi: 10.3390/ijerph16173042] [Medline: 31443363]
13. Skeppholm M, Friberg L. Adherence to warfarin treatment among patients with atrial fibrillation. Clin Res Cardiol 2014 Dec;103(12):998-1005. [doi: 10.1007/s00392-014-0742-y] [Medline: 25080281]
14. Banerjee A, Benedetto V, Gichuru P, Burnell J, Antoniou S, Schilling RJ, et al. Adherence and persistence to direct oral anticoagulants in atrial fibrillation: a population-based study. Heart 2020 Jan;106(2):119-126 [FREE Full text] [doi: 10.1136/heartjnl-2019-315307] [Medline: 31601729]
15. Ozaki AF, Choi AS, Le QT, Ko DT, Han JK, Park SS, et al. Real-world adherence and persistence to direct oral anticoagulants in patients with atrial fibrillation: a systematic review and meta-analysis. Circ Cardiovasc Qual Outcomes 2020 Mar;13(3):e005969. [doi: 10.1161/CIRCOUTCOMES.119.005969] [Medline: 32148102]
16. Yamashiro K, Kurita N, Tanaka R, Ueno Y, Miyamoto N, Hira K, et al. Adequate adherence to direct oral anticoagulant is associated with reduced ischemic stroke severity in patients with atrial fibrillation. J Stroke Cerebrovasc Dis 2019 Jun;28(6):1773-1780. [doi: 10.1016/j.jstrokecerebrovasdis.2018.09.019] [Medline: 30318259]
17. Komen JJ, Heerdink ER, Klungel OH, Mantel-Teeuwisse AK, Forslund T, Wettermark B, et al. Long-term persistence and adherence with non-vitamin K oral anticoagulants in patients with atrial fibrillation and their associations with stroke risk. Eur Heart J Cardiovasc Pharmacother 2021 Apr;7(FI1):f72-f80. [doi: 10.1093/ehjcvp/pvaa017] [Medline: 32324233]
18. Rivera-Caravaca JM, Esteve-Pastor MA, Roldán V, Marín F, Lip GYH. Non-vitamin K antagonist oral anticoagulants: impact of non-adherence and discontinuation. Expert Opin Drug Saf 2017 Sep;16(9):1051-1062. [doi: 10.1080/14740338.2017.1351542] [Medline: 28678553]
19. Caldeira D, Nunes-Ferreira A, Rodrigues R, Vicente E, Pinto FJ, Ferreira JJ. Non-vitamin K antagonist oral anticoagulants in elderly patients with atrial fibrillation: a systematic review with meta-analysis and trial sequential analysis. Arch Gerontol Geriatr 2019;81:209-214. [doi: 10.1016/j.archger.2018.12.013] [Medline: 30612074]

20. Shaikh F, Pasch LB, Newton PJ, Bajorek BV, Ferguson C. Addressing multimorbidity and polypharmacy in individuals with atrial fibrillation. Curr Cardiol Rep 2018 Mar;20(5):32. [doi: 10.1007/s11886-018-0975-x] [Medline: 29574524]
21. Berti D, Hendriks JML, Brandes A, Deaton C, Crijns HJGM, Camm AJ, et al. A proposal for interdisciplinary, nurse-coordinated atrial fibrillation expert programmes as a way to structure daily practice. Eur Heart J 2013 Sep;34(35):2725-2730. [doi: 10.1093/eurheartj/eht096] [Medline: 23520187]
22. Desteghe L, Vijgen J, Koopman P, Dilling-Boer D, Schurmans J, Dendale P, et al. Telemonitoring-based feedback improves adherence to non-vitamin K antagonist oral anticoagulants intake in patients with atrial fibrillation. Eur Heart J 2018 Apr;39(16):1394-1403. [doi: 10.1093/eurheartj/ehx762] [Medline: 29300888]
23. Guo Y, Chen Y, Lane DA, Liu L, Wang Y, Lip GYH. Mobile health technology for atrial fibrillation management integrating decision support, education, and patient involvement: mAF app trial. Am J Med 2017 Dec;130(12):1388-1396 [FREE Full text] [doi: 10.1016/j.amjmed.2017.07.003] [Medline: 28847546]
24. Guo Y, Guo J, Shi X, Yao Y, Sun Y, Xia Y, mAF-App II Trial investigators. Mobile health technology-supported atrial fibrillation screening and integrated care: a report from the mAFA-II trial Long-term extension cohort. Eur J Intern Med 2020 Dec;82:105-111 [FREE Full text] [doi: 10.1016/j.ejim.2020.09.024] [Medline: 33067121]
25. Brørs G, Pettersen TR, Hansen TB, Fridlund B, Hølvold LB, Lund H, et al. Modes of e-Health delivery in secondary prevention programmes for patients with coronary artery disease: a systematic review. BMC Health Serv Res 2019 Jun;19(1):364 [FREE Full text] [doi: 10.1186/s12913-019-4106-1] [Medline: 31182100]
26. Create a blocked randomisation list. Sealed Envelope. URL: https://www.sealedenvelope.com/simple-randomiser/v1/lists [accessed 2021-02-27]
27. Carver CS. You want to measure coping but your protocol's too long: consider the brief COPE. Int J Behav Med 1997;4(1):92-100. [doi: 10.1207/s15327558ijbm0401_6] [Medline: 16250744]
28. Franks HM, Roesch SC. Appraisals and coping in people living with cancer: a meta-analysis. Psychooncology 2006 Dec;15(12):1027-1037. [doi: 10.1002/pon.1043] [Medline: 16602072]
29. Thompson K, Kulkarni J, Sergejew AA. Reliability and validity of a new Medication Adherence Rating Scale (MARS) for the psychoses. Schizophr Res 2000 May;42(3):241-247. [doi: 10.1016/s0920-9964(99)00130-9] [Medline: 10785582]
30. Wu C, Gong Y, Wu J, Zhang S, Yin X, Dong X, et al. Chinese version of the EQ-5D preference weights: applicability in a Chinese general population. PLoS One 2016;11(10):e0164334 [FREE Full text] [doi: 10.1371/journal.pone.0164334] [Medline: 27711169]
31. Hardin J. Generalized estimating equations (GEE). In: Encyclopedia of Statistics in Behavioral Science. Chichester, UK: John Wiley & Sons, Ltd; 2005:721-729.
32. Nick TG, Campbell KM. Logistic regression. Methods Mol Biol 2007;404:273-301. [doi: 10.1007/978-1-59745-530-5_14] [Medline: 18450055]
33. Erdfelder E, Faul F, Buchner A. GPOWER: A general power analysis program. Behavior Research Methods, Instruments, & Computers 1996 Mar;28(1):1-11. [doi: 10.3758/bf03203630]
34. Cotton SM, McCann TV, Gleeson JF, Crisp K, Murphy BP, Lubman DI. Coping strategies in carers of young people with a first episode of psychosis. Schizophr Res 2013 May;146(1-3):118-124. [doi: 10.1016/j.schres.2013.02.008] [Medline: 23490761]
35. Levasseur M, Couture M. Coping strategies associated with participation and quality of life in older adults. Can J Occup Ther 2015 Feb;82(1):44-53. [doi: 10.1177/0008417414552188] [Medline: 25803947]
36. Smith MM, Saklofske DH, Keefer KV, Tremblay PF. Coping strategies and psychological outcomes: the moderating effects of personal resiliency. J Psychol 2016;150(3):318-332. [doi: 10.1080/00223980.2015.1036828] [Medline: 25951375]
37. Taylor SE, Stanton AL. Coping resources, coping processes, and mental health. Annu Rev Clin Psychol 2007;3:377-401. [doi: 10.1146/annurev.clinpsy.3.022806.091520] [Medline: 17716061]
38. Wang Y, Kong MC, Lee LH, Ng HJ, Ko Y. Knowledge, satisfaction, and concerns regarding warfarin therapy and their association with warfarin adherence and anticoagulation control. Thromb Res 2014 Apr;133(4):550-554. [doi: 10.1016/j.thromres.2014.01.002] [Medline: 24448058]
39. Ammenwerth E, Woess S, Baumgartner C, Fetz B, van der Heidt A, Kastner P, et al. Evaluation of an integrated telemonitoring surveillance system in patients with coronary heart disease. Methods Inf Med 2015;54(5):388-397 [FREE Full text] [doi: 10.3414/ME15-02-0002] [Medline: 26395147]
40. Sherwood A, Blumenthal JA, Koch GG, Hoffman BM, Watkins LL, Smith PJ, et al. Effects of coping skills training on quality of life, disease biomarkers, and clinical outcomes in patients with heart failure: a randomized clinical trial. Circ Heart Fail 2017 Jan;10(1):e003410 [FREE Full text] [doi: 10.1161/CIRCHEARTFAILURE.116.003410] [Medline: 28062537]
41. Krack G, Holle R, Kirchberger I, Kuch B, Amann U, Seidl H. Determinants of adherence and effects on health-related quality of life after myocardial infarction: a prospective cohort study. BMC Geriatr 2018 Jun;18(1):136 [FREE Full text] [doi: 10.1186/s12877-018-0827-y] [Medline: 29898677]
42. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]

43. Hale TM, Jethwani K, Kandola MS, Saldana F, Kvedar JC. A remote medication monitoring system for chronic heart failure patients to reduce readmissions: a two-arm randomized pilot study. J Med Internet Res 2016 Apr;18(5):e91 [FREE Full text] [doi: 10.2196/jmir.5256] [Medline: 27154462]
44. Guo Y, Lane DA, Wang L, Zhang H, Wang H, Zhang W, mAF-App II Trial Investigators. Mobile health technology to improve care for patients with atrial fibrillation. J Am Coll Cardiol 2020 Apr;75(13):1523-1534. [doi: 10.1016/j.jacc.2020.01.052] [Medline: 32241367]
45. Lin MH, Yuan WJ, Huang TC, Zhang HF, Mai JT, Wang JF. Clinical effectiveness of telemedicine for chronic heart failure: a systematic review and meta-analysis. J Investig Med 2017 Jun;65(5):899-911. [doi: 10.1136/jim-2016-000199] [Medline: 28330835]

Abbreviations

AF: atrial fibrillation
EQ-5D-3L: three-level version of EuroQol five-dimension self-report questionnaire
EQ-VAS: EuroQol visual analog scale
GEE: generalized estimating equation
HRQoL: health-related quality of life
MARS: Medication Adherence Rating Scale
mEHRA: modified European Heart Rhythm Association

Edited by G Eysenbach; submitted 01.05.21; peer-reviewed by J Kvedar; comments to author 01.06.21; revised version received 17.06.21; accepted 26.07.21; published 22.09.21.

Please cite as:
Hsieh HL, Kao CW, Cheng SM, Chang YC
A Web-Based Integrated Management Program for Improving Medication Adherence and Quality of Life, and Reducing Readmission in Patients With Atrial Fibrillation: Randomized Controlled Trial
J Med Internet Res 2021;23(9):e30107
URL: https://www.jmir.org/2021/9/e30107
doi:10.2196/30107
PMID:34550084

Original Paper

Facilitator Contact, Discussion Boards, and Virtual Badges as Adherence Enhancements to a Web-Based, Self-guided, Positive Psychological Intervention for Depression: Randomized Controlled Trial

Judith Tedlie Moskowitz[1], MPH, PhD; Elizabeth L Addington[1], PhD; Eva Shiu[2], MPH; Sarah M Bassett[1], PhD; Stephanie Schuette[3], MA; Ian Kwok[4], MA; Melanie E Freedman[4], BA; Yan Leykin[5], PhD; Laura R Saslow[6], PhD; Michael A Cohn[7], PhD; Elaine O Cheung[1], PhD

[1]Department of Medical Social Sciences, Osher Center for Integrative Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[2]Department of Obstetrics and Gynecology, Biological Sciences Division, University of Chicago, Chicago, IL, United States

[3]Department of Psychology and Neuroscience, Duke University, Durham, NC, United States

[4]Department of Psychiatry and Behavioral Sciences, Osher Center for Integrative Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[5]Department of Psychology, Palo Alto University, Palo Alto, CA, United States

[6]Department of Health Behavior and Biological Sciences, University of Michigan School of Nursing, Ann Arbor, MI, United States

[7]Osher Center for Integrative Medicine, University of California San Francisco, San Francisco, CA, United States

Corresponding Author:
Judith Tedlie Moskowitz, MPH, PhD
Department of Medical Social Sciences
Osher Center for Integrative Medicine
Northwestern University Feinberg School of Medicine
625 N Michigan Ave
Suite 2700
Chicago, IL, 60611
United States
Phone: 1 3125037712
Email: judith.moskowitz@northwestern.edu

Abstract

Background: Adherence to self-guided interventions tends to be very low, especially in people with depression. Prior studies have demonstrated that enhancements may increase adherence, but little is known about the efficacy of various enhancements in comparison to, or in combination with, one another.

Objective: The aim of our study is to test whether 3 enhancements—facilitator contact (FC), an online discussion board, and virtual badges (VB)—alone, or in combination, improve adherence to a self-guided, web-based intervention for depression. We also examined whether age, gender, race, ethnicity, comfort with technology, or baseline depression predicted adherence or moderated the effects that each enhancement had on adherence.

Methods: Participants were recruited through web-based sources and, after completing at least 4 out of 7 daily emotion reports, were sequentially assigned to 1 of 9 conditions—the intervention alone; the intervention plus 1, 2, or all 3 enhancements; or an emotion reporting control condition. The intervention was a positive psychological program consisting of 8 skills that specifically targeted positive emotions, and it was delivered over 5 weeks in a self-guided, web-based format. We operationalized adherence as the number of skills accessed.

Results: A total of 602 participants were enrolled in this study. Participants accessed, on average, 5.61 (SD 2.76) of 8 skills. The total number of enhancements participants received (0-3) did not predict the number of skills accessed. Participants who were assigned to the VB+FC condition accessed significantly more skills than those in the intervention only conditions. Furthermore, participants in arms that received the combination of both the VB and FC enhancements (VB+FC and VB+FC+online discussion board) accessed a greater number of skills relative to the number of skills accessed by participants who received either VB or FC

without the other. Moderation analyses revealed that the receipt of VB (vs no VB) predicted higher adherence among participants with moderately severe depression at baseline.

Conclusions: The results suggested that the VB+FC combination significantly increased the number of skills accessed in a self-guided, web-based intervention for elevated depression. We have provided suggestions for refinements to these enhancements, which may further improve adherence.

Trial Registration: ClinicalTrials.gov NCT02861755; http://clinicaltrials.gov/ct2/show/NCT02861755

(J Med Internet Res 2021;23(9):e25922) doi:10.2196/25922

KEYWORDS

mHealth; adherence; depression; discussion board; gamification; positive psychological intervention; mobile phone

Introduction

Despite the fact that up to 20% of people in the United States will experience depression over the course of their lives [1], more than half will remain undiagnosed and untreated [2] for a variety of reasons including treatment cost and geographic inaccessibility [3,4]. Web-based self-help or self-guided (eHealth) programs can overcome potential barriers associated with conventional modes of treatment and may be particularly useful for individuals who might not otherwise seek out conventional face-to-face therapy [5]. However, eHealth interventions for depression also come with significant challenges. Retention in these web-based, self-guided interventions is low compared with other types of trials [6]. For example, previous studies reported retention rates as low as 43% posttreatment, with even lower rates with longer-term follow-up [7]. A recent review of retention in studies of smartphone-delivered interventions for mental health concluded that researchers can expect attrition of up to one-third of participants who enroll in the study [8]. In addition, adherence to intervention content—operationalized as the number of log-ins, duration of web exposure, or number of modules or exercises completed—is as low as 50% for web-based programs [7], and complete adherence to all components of smartphone-delivered interventions for depression was achieved by an average of only 34% of participants [8]. Predictors of poor adherence in web-based depression trials include higher baseline levels of depression [7], male gender [9,10], and lower levels of education [9]. Older age is sometimes found to be associated with better adherence [9,10] and sometimes poorer adherence [7]. Those with less frequent internet use at baseline (weekly to monthly compared with daily) may be more adherent to self-guided programs [10].

Social enhancements such as discussion boards that encourage users to generate and share content with each other can improve engagement [11-14]. Opportunities to earn points and badges can also increase motivation by providing virtual rewards to participants in the form of positive feedback [15-17]. More conventional enhancements, such as reminders via postcards or brief phone calls, can also increase participation [18,19], although other studies have found that similar enhancements such as personalization, interactivity, reminders, and text messages were not associated with better intervention adherence [20]. The discrepancy in these results suggests that specific automated enhancements, or combinations of automated and human-supported enhancements, may be necessary to maximize adherence to self-guided programs.

In this study, we tested the following three enhancements to a self-guided web-based intervention for people experiencing depressive symptoms: brief facilitator contact (FC), an online discussion board (ODB), and virtual badges (VB). These features were developed in previous phases of the program [21,22] and were selected based on research suggesting that personal contact, interactive components, and incentivizing participation increased engagement in web-based programs [7,23-29].

One enhancement was brief facilitator phone contact. Contact from study staff, a mental health professional, or another type of coach can provide a sense of obligation or accountability on the part of the participant and increase adherence to intervention content [29]. In eHealth interventions, this social presence, or sense that there is another human behind the intervention who is aware of the participant's engagement and may provide reminders or prompts, significantly improves adherence [30,31]. FC may increase user engagement by building rapport and connection to the intervention, providing feedback, or supporting meaningful use by tailoring the content or application of the material to participants' needs [32].

The second enhancement we tested was an ODB. Asynchronous discussion boards where participants can reflect on their experiences of learning the skills presented in the intervention and interact with other participants may provide a sense of peer support that improves the efficacy of the intervention. Although web-based support groups alone have not been particularly effective [33], from a collaborative learning perspective, the addition of a forum that allows participants to interact asynchronously with other participants and intervention content may help facilitate understanding and retention of the content and increase engagement with the targeted behaviors [34].

VB are a common form of gamification that increases engagement with an intervention [35] through the inclusion of playful, enjoyable, or competitive elements that boost adherence to a designated activity [36]. Although VB have been criticized for their emphasis on extrinsic motivation at the expense of intrinsic [35], there is evidence that gamified interventions are more engaging than those that do not contain such elements [37].

We hypothesize that the enhancements would increase adherence, operationalized as the number of sessions accessed,

and aim to explore whether the combination of enhancements had a significantly greater impact than any one alone. We also examine whether age, gender, race or ethnicity, comfort with technology, or baseline depression predict adherence or moderate the effects of each enhancement on adherence. Finally, we explore enjoyment ratings as predictors of use or engagement with each enhancement.

Methods

Participants

Detailed methods are described in the study by Cheung et al [22]. The study was approved by the institutional review board and preregistered through ClinicalTrials.gov (trial number: NCT02861755). All participants were recruited on the web through platforms such as ResearchMatch, Craigslist, and Reddit. Eligibility was determined using a web-based screener. To be eligible for the study, participants had to (1) have at least mild levels of depression, as indicated by a Patient Health Questionnaire-8 (PHQ-8) depression score of ≥5 [38]; (2) be aged ≥18 years; (3) have daily access to the internet; (4) own a mobile phone; (5) live in the United States; and (6) be able to read and write in English.

Procedures

Baseline and Randomization

After obtaining web-based consent, participants completed a baseline questionnaire and a 7-day run-in period in which they were asked to complete brief daily assessments of emotion. If participants completed the baseline assessment and at least 4 of 7 days of the run-in period, they were allocated to one of the nine study arms. We stratified participant allocation to a condition based on gender and the level of depressive symptom severity (PHQ-8 score: 5-9 [mild]; 10-14 [moderate]; 15-19 [moderately severe]; >20 [severe]) to ensure sufficient numbers of each group within each condition, and participants were sequentially assigned to the study arm after stratification.

The nine study arms varied according to the type and number of adherence enhancements. The nine conditions are as follows: (1) intervention alone, (2) intervention+FC, (3) intervention+ODB, (4) intervention+VB, (5) intervention+FC+ODB, (6) intervention+FC+VB, (7) intervention+ODB+VB, (8) intervention+FC+ODB+VB, and (9) the emotion reporting control condition.

Enhancements

In the FC conditions, the study staff called participants once per week to check their progress through the course. Specifically, following a prescribed set of questions, the staff member asked about the participant's experience learning that week's skills, using the home practice, and completing the daily emotion surveys; they probed for questions about the skills, accomplishments, barriers, and any difficulties with the technology. If the participant could not be reached by phone, the staff member sent an email with the same content in place of the call and encouraged the participant to reply if they had questions. Staff members were explicitly trained not to provide supportive counseling or skill training, but instead focused on discussing progress, challenges, and technology issues.

Participants who were assigned to the ODB had access to a virtual discussion board where they could participate anonymously using pseudonyms. The study staff provided prompts to seed the discussion board and were notified when a participant posted something. Posts were reviewed for inappropriate content and signs of extreme distress or suicidality.

In the VB condition, participants could earn flower badges that they could arrange in a garden plot. They received badges for completing activities such as creating a profile, reading a lesson, or logging in to the website for 7 consecutive days.

MARIGOLD Intervention

Conditions 1-8 received the MARIGOLD program, a positive psychological intervention that consisted of eight skills that specifically targeted positive emotions: noticing positive events, capitalizing, gratitude, behavioral activation, mindfulness, positive reappraisal, personal strengths, and acts of kindness. Additional details of the positive emotion skills intervention can be found elsewhere [21,22,39-43]. The skills were delivered over the course of 5 weeks in a self-guided, web-based format, and each week had associated home practice activities. Participants did not have access to all the skills at once; instead, 1 to 3 skills were released in each of the 5 weeks. Participants in conditions 1 to 8 also completed the same daily emotion reporting as the control condition (described in the following section).

Emotion Reporting Control

Participants in arm 9 were asked to log in to the website once per day for 49 days (the length of the other eight conditions) to report their emotions. Participants in the emotion reporting control arm completed approximately 20 of the 49 days of daily emotion reporting during the study period (mean 19.90, SD 18.01; range 0-50). They did not receive any of the intervention content and thus did not have adherence data as we have defined it; participants in the control condition will therefore not be included in the present analyses (we have included greater detail on the rates of retention for participants in the emotion reporting control arm and the comparison of daily emotion reporting in the emotion reporting control arm versus the other eight conditions in the supplemental analyses; Multimedia Appendix 1).

Measures

Adherence and Enhancement Ratings

The primary adherence outcome was operationalized as the number of skills (out of eight) accessed (see supplemental analyses in Multimedia Appendix 1 for analyses with other operationalizations of adherence [number of pages viewed and home practice completed] and study retention. Predictors of alternative operationalizations of adherence and of retention were essentially the same). In addition, we explored the evaluations of the enhancements, measured both quantitatively and qualitatively. Upon completion of the 5-week MARIGOLD study, participants completed postcourse surveys in which they provided feedback about the intervention content and

enhancements. Participants were asked to rate whether they enjoyed each enhancement they were assigned to receive (ie, FC, ODB, and VB) on a slider scale from 0 (*definitely not*) to 100 (*definitely yes*). Participants also asked open-ended questions regarding the benefits, drawbacks, and impact of each of the enhancements they received.

Predictors of Adherence and Enhancement Ratings

We also examined demographics, baseline depression, and comfort with technology as predictors of adherence and enjoyment of enhancements. Gender was assessed as *male*, *female*, *or other*, or *prefer not to answer*. Participants were asked whether they considered themselves Hispanic or Latino or Latina and how they self-identified their race (*Black*, *White*, *Asian*, *Native American*, *Pacific Islander*, and *mixed or other*). Age was calculated based on the date of birth. Depression was assessed using the PHQ-8 [38]. Finally, comfort with technology was assessed with six items that tapped a participant's overall frequency of and confidence with technology use (eg, "I am confident I can navigate websites") with response options ranging from 0 (*not at all confident*) to 10 (*totally confident*).

Analyses

Baseline Predictors of Adherence and Engagement With Enhancements

We first conducted an overdispersed (or quasi-) Poisson regression [44,45] to examine whether the number of enhancements received predicted the number of skills accessed. We then conducted the analyses predicting the number of skills accessed from the intervention arm, with the intervention only condition as the reference category.

We then conducted overdispersed Poisson regressions predicting the number of skills accessed from age, gender (1=female; male=0), race or ethnicity (with indicator variables to represent each race or ethnicity category), comfort with technology, and baseline depressive symptom severity (categorized as mild, moderate, moderately severe, and severe).

To further examine whether adherence to the study differed as a function of the combination of enhancements received, we conducted an additional overdispersed Poisson regression predicting each adherence outcome as a function of whether the participant received each enhancement (FC, ODB, and VB), the FCODB, FCVB, and ODBVB two-way interactions and the FCODBVB three-way interaction. The two-way and three-way interactions allowed us to explore whether receiving a specific combination of enhancements had a significantly greater impact than receiving any one enhancement alone.

Moderators of Adherence

We examined whether the effect of each enhancement on adherence was moderated by the following characteristics: age, gender (female vs male), ethnicity (Hispanic vs non-Hispanic), race (Asian, Black or African American, White, other race [Pacific Islander, Native American, and mixed or other]), comfort with technology, and baseline depressive symptom severity (mild, moderate, moderately severe, and severe). To examine this, overdispersed Poisson regressions predict adherence by whether the participants received each enhancement, with dummy variables representing each enhancement, the moderator of interest, and the two-way interactions between the moderator and each enhancement type (FC×moderator, ODB×moderator, and VB×moderator). Owing to power limitations, we did not test all 4 moderators in the same model or explored interactions among the moderators and combinations of enhancements.

Power Analysis

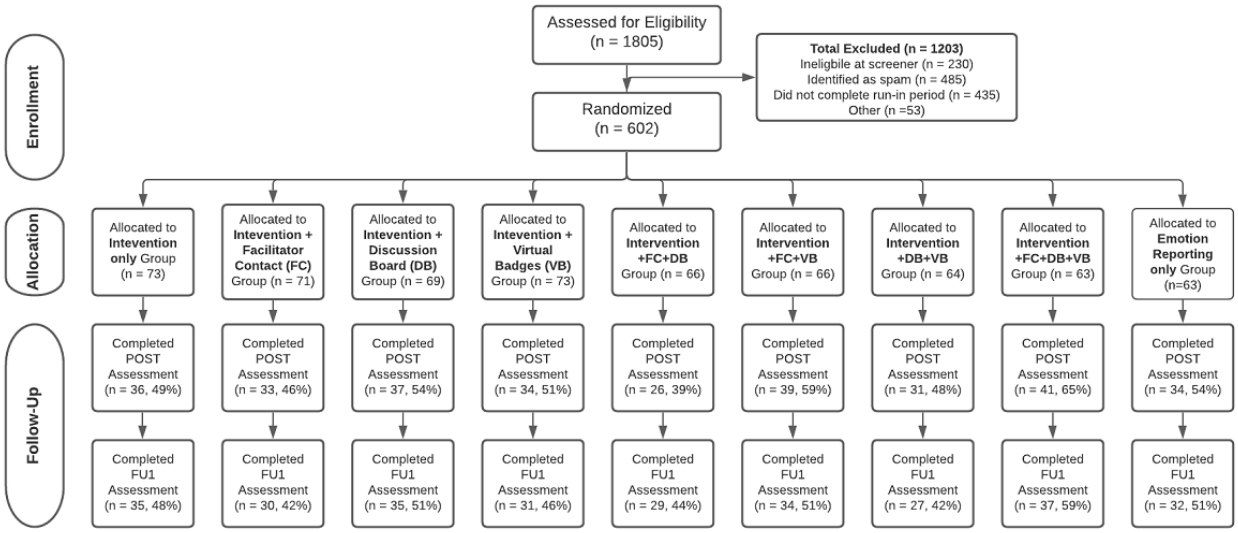
A sensitivity power analysis using the G*Power 3.1.9.6 software [46] revealed that the sample size for this study (N=539 participants across the eight intervention arms) was sufficient to detect a small effect size with adequate power. The minimum effect size necessary for this study was w=0.12, assuming an α significance criterion of P=.05 and power=0.80.

Results

Screening, Enrollment, and Randomization

Of the 1805 respondents who completed the screener, 1037 (57.45%) were deemed eligible and entered the run-in period. Of these, 58.05% (602/1037) completed at least 4 of 7 days of the emotion reports for the run-in and were sequentially assigned to one of the nine study conditions. Figure 1 shows the CONSORT (Consolidated Standards of Reporting Trials) diagram. There were 485 eligibility forms that were deemed spam. Most of these cases were identified because they enrolled in quick succession on the same day and had suspicious repeating patterns of first or last names. We sent additional screener questions to these cases and none completed them and therefore did not proceed to enrollment.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram. FU: follow-up.



Baseline Characteristics

Across the nine conditions, there were no differences in baseline demographics, depression, or comfort with technology (Table 1). The average age of the participants was 38 years (range 18-80 years). Three-fourths of the sample were female (422/573, 73.6%), three-fourths were White (421/573, 73.4%), and 43% (244/568) were college graduates. Compared with those who did not complete at least 4 of the 7 days of the run-in period, participants who completed the run-in were older (mean 37.95, SD 13.68 vs mean 36.03, SD 12.77; *P*=.04), more likely to be female (467/633, 73.7% vs 210/323, 65%; *P*=.008), and had lower baseline depressive symptom severity (PHQ-8: mean 13.85, SD 5.18 vs mean 15.34, SD 5.11); *P*<.001; Table S1 in Multimedia Appendix 1).

Table 1. Sample demographics and baseline variables by condition.

Baseline variables	Total sample (N=602)	Condition									*P* value[a]
		Intervention alone (n=73)	FC[b] (n=71)	ODB[c] (n=69)	VB[d] (n=67)	FC+ODB (n=66)	FC+VB (n=66)	ODB+VB (n=64)	FC+ODB +VB (n=63)	Control (n=63)	
Age (years), mean (SD)	37.8 (13.5)	39.1 (13.4)	37.8 (14.5)	38.9 (14.1)	38.4 (15.5)	38.0 (12.6)	38.4 (14.2)	36.1 (12.1)	38.6 (12.8)	34.7 (12.1)	.67
Gender, n (%)											.85
Male	146 (26)	18 (25)	17 (25)	18 (26)	16 (25)	16 (25)	17 (27)	15 (25)	13 (25)	16 (25)	
Female	422 (74)	51 (72)	48 (72)	49 (72)	49 (75)	47 (75)	46 (73)	46 (75)	39 (75)	47 (75)	
Baseline PHQ-8[e], n (%)											.99
Mild	139 (23)	17 (23)	16 (23)	16 (23)	16 (24)	16 (24)	16 (24)	14 (22)	14 (22)	14 (22)	
Moderate	193 (32)	23 (32)	23 (32)	21 (30)	21 (31)	21 (32)	21 (32)	21 (33)	21 (33)	21 (33)	
Moderately severe	175 (29)	21 (29)	21 (30)	21 (30)	19 (28)	19 (29)	19 (29)	19 (30)	18 (29)	18 (29)	
Severe	95 (16)	12 (16)	11 (16)	11 (16)	11 (16)	10 (15)	10 (15)	10 (16)	10 (16)	10 (16)	
Ethnicity, n (%)											.89
Hispanic	100 (17)	13 (18)	8 (12)	11 (16)	10 (15)	13 (21)	11 (18)	12 (20)	12 (23)	10 (16)	
Non-Hispanic	473 (83)	58 (82)	59 (88)	57 (84)	55 (85)	50 (79)	52 (83)	49 (80)	40 (77)	53 (84)	
Race, n (%)											
Black	93 (16)	8 (11)	7 (10)	15 (22)	9 (14)	13 (21)	12 (19)	10 (16)	9 (17)	10 (16)	.62
White	421 (74)	(73)	(73)	(74)	(79)	(68)	(73)	(67)	(79)	(75)	.86
Asian	49 (9)	9 (13)	11 (16)	4 (6)	4 (6)	4 (6)	6 (10)	5 (8)	3 (6)	3 (5)	.26
Native American	27 (5)	2 (3)	5 (8)	4 (6)	4 (6)	3 (5)	4 (6)	2 (3)	0 (0)	3 (5)	.71
Pacific Islander	2 (1)	0 (0)	0 (0)	1 (2)	0 (0)	0 (0)	1 (2)	0 (0)	0 (0)	0 (0)	.56
Mixed or other	43 (8)	4 (6)	5 (8)	5 (7)	7 (11)	5 (8)	2 (3)	8 (13)	1 (2)	6 (10)	.38
Education, n (%)											.33
<High school	4 (1)	0 (0)	0 (0)	1 (2)	0 (0)	2 (3)	0 (0)	1 (2)	0 (0)	0 (0)	
High school	40 (7)	4 (6)	2 (3)	10 (15)	6 (9)	5 (8)	7 (11)	2 (3)	2 (4)	2 (3)	
Some college	170 (30)	23 (32)	21 (32)	15 (22)	23 (35)	16 (26)	18 (29)	23 (38)	16 (31)	15 (25)	
College graduate	244 (43)	23 (32)	34 (52)	32 (47)	26 (40)	27 (44)	28 (44)	21 (34)	21 (40)	32 (53)	
Professional degree	98 (17)	18 (25)	8 (12)	9 (13)	9 (14)	12 (19)	8 (13)	12 (20)	12 (23)	10 (17)	
Other	12 (2)	3 (4)	1 (2)	1 (2)	1 (2)	0 (0)	2 (3)	2 (3)	1 (2)	1 (2)	
Income (US $), n (%)											.24
<10,000	48 (8)	10 (14)	4 (6)	8 (12)	3 (5)	8 (13)	6 (10)	5 (8)	2 (4)	2 (3)	
10,000 to 19,999	65 (11.5)	11 (16)	6 (9)	9 (13)	11 (17)	10 (16)	4 (7)	4 (7)	4 (8)	6 (10)	
20,000 to 29,999	92 (16)	11 (15)	10 (15)	13 (19)	12 (19)	11 (18)	8 (13)	7 (12)	10 (19)	10 (15)	
30,000 to 49,999	102 (18)	10 (14)	15 (23)	16 (24)	13 (20)	9 (15)	10 (16)	10 (16)	10 (19)	9 (15)	
50,000 to 74,999	105 (19)	10 (14)	13 (20)	11 (16)	10(15)	7 (11)	13 (21)	9 (15)	13 (25)	19 (31)	
75,000 to 99,999	72 (13)	10 (14)	6 (9)	5 (8)	7 (11)	7 (11)	10 (16)	17 (28)	3 (6)	7 (12)	
100,000 to 199,999	68 (12)	7 (10)	10 (15)	4 (6)	9 (14)	7 (11)	8 (13)	9 (15)	7 (14)	7 (12)	
>200,000	14 (2)	1 (1)	2 (3)	1 (2)	0 (0)	3 (5)	3 (5)	0 (0)	3 (6)	1 (2)	
Comfort with technology, mean (SD)	9.2 (1.3)	9.1 (1.6)	9.3 (1.2)	9.2 (1.2)	9.1 (1.5)	9.2 (1.2)	9.0 (1.3)	9.2 (1.3)	9.1 (1.7)	9.5 (0.7)	.57

[a]*P* values were based on analysis of variance (continuous) and chi-square tests (categorical outcomes).

[b]FC: facilitator contact.

[c]ODB: online discussion board.

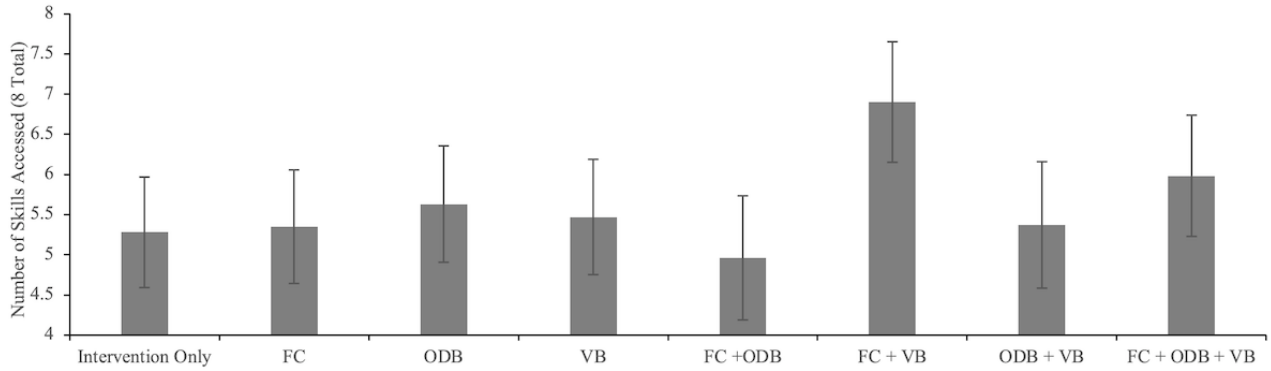
[d]VB: virtual badges.

[e]PHQ-8: Patient Health Questionnaire-8.

Adherence

Participants accessed, on average, 5.61 (SD 2.76) of 8 skills. The total number of enhancements that participants received (0-3) did not predict the number of skills accessed (1 enhancement vs 0: *P*=.63; 2 enhancements vs 0: *P*=.25; 3 enhancements vs 0: *P*=.19). We conducted an analysis predicting the number of skills accessed as a function of whether the participant received each enhancement and their two- and three-way interactions. There were no significant main effects for the FC (*P*=.24) or ODB (*P*=.36) enhancements. There was a significant main effect of receiving the VB enhancement, such that participants in the conditions that received VB (VB, VB+FC, VB+ODB, and VB+FC+ODB) had a greater number of skills (mean 5.90, 95% CI 5.52-6.31) relative to participants in groups that did not receive the VB enhancement (intervention only, FC, ODB, and FC+ODB; mean 5.30; range 4.96-5.67; χ^2_1=5.0; *P*=.03). This main effect, however, appeared to be driven by a significant VB by FC two-way interaction predicting the number of skills accessed (χ^2_1=5.5; *P*=.02) such that participants in arms that received the *combination* of both the VB and FC enhancements together (VB+FC and VB+FC+ODB) accessed a greater number of skills (mean 6.42; range 5.87-7.03) than participants who received either enhancement without the other (VB without FC: mean 5.42; range 4.92-5.98; *P*=.01; FC without VB: mean 5.15; range 4.67-5.68; *P*=.001). Furthermore, participants who received either VB or FC enhancement without the other did not differ in the number of skills accessed compared with participants who received neither enhancement (neither VB or FC: mean 5.45; range 4.98-5.98; *P*=.40 to *P*=.93). No other two-way or three-way interactions emerged with statistical significance (FC by ODB interaction: *P*=.17; ODB by VB interaction: *P*=.44; and FC by ODB by VB interaction: *P*=.94). Taken together, these findings suggest that receiving the VB and FC enhancements *in combination* increased the number of skills accessed, but receiving either the VB or FC enhancement without the other did not (Figure 2).

Figure 2. Number of skills accessed (n=8) as a function of condition. Error bars denote 95% CIs. FC: facilitator contact; ODB: online discussion board; VB: virtual badges.



The patterns of results were the same for other indicators of adherence (eg, pages viewed or home practice completed) and retention in the study. The details of these analyses are provided in Multimedia Appendix 1.

Moderators of Adherence

We examined age, gender, race, ethnicity, comfort with technology, and baseline depression as predictors of adherence and as moderators of the effects of each individual enhancement on adherence. As seen in Table 2, none of these baseline characteristics significantly predicted participants' adherence to the intervention content (*P*=.19 to *P*=.87).

Table 2. Overdispersed Poisson regression predicting adherence from baseline predictors.

Predictor	Number of skills accessed (n=8)	
	Odds ratio (95% CI)	*P* value
Age (years)	1.00 (0.99-1.00)	.79
Gender		
Female	1.08 (0.96-1.21)	.20
Male (reference)	—[a]	—
Ethnicity		
Hispanic	1.04 (0.91-1.19)	.60
Non-Hispanic (reference)	—	—
Race		
Asian	0.94 (0.78-1.13)	.50
Black or African American	1.09 (0.96-1.23)	.19
Other race[b]	0.97 (0.82-1.15)	.70
White (reference)	—	—
Comfort with technology	1.01 (0.97-1.04)	.78
Baseline PHQ-8[c]		
Moderate	1.05 (0.92-1.19)	.47
Moderately severe	0.99 (0.86-1.13)	.87
Severe	1.04 (0.90-1.22)	.58
Mild (reference)	—	—

[a]Not available (indicates reference value).

[b]Other race: the other race category constitutes those who identified as Pacific Islander, Native American, mixed, or other.

[c]PHQ-8: Patient Health Questionnaire-8.

Next, we explored whether the enhancements had a greater effect on adherence depending on demographics, comfort with technology, or baseline depression. There was a significant baseline depressive symptom severity by VB two-way interaction in predicting the number of skills accessed, χ^2_3=8.6 (*P*=.04). Specifically, among participants with moderately severe depression (scores between 15 and 19 on the PHQ-8), receiving VB enhancement was associated with better adherence compared with those who did not receiving VB enhancement (B=1.09; *P*=.01). Two trends in the data that did not reach conventional levels of statistical significance are worth noting: for those in the mild (5-9) and moderate (10-14) range in the PHQ, there was also a positive association between the VB group and adherence (mild: B=1.00; *P*=.06; moderate: B=0.80; *P*=.10). At the highest level of depressive symptoms (scores >20); however, the association was in the opposite direction, with those who received the VB enhancement showing *lower* adherence compared with those who did not receive the VB enhancement (B=−1.21; *P*=.08). Given that the effects among participants in 3 of the 4 depression groups (mild, moderate, and severe but not moderately severe) did not reach statistical significance, caution is warranted in drawing conclusions based on these data.

Exploratory Analyses on Engagement With Enhancements

To better understand the responses to the three enhancements, we further explored the level of engagement and enjoyment ratings for each one.

Analysis of FC

For FC (n=266), study staff were able to contact 45.5% (121/266) of the participants in the FC enhancement conditions by phone at least once, and the average number of weeks that staff were able to reach participants by phone was 1 (SD 1.42; range 0-6). Facilitators were able to reach participants who were older, female, and more comfortable with technology and those who had baseline depression scores in the moderate and moderately severe range more frequently than those who were younger, male, less comfortable with technology, and had baseline depression scores in the mild range (*P*=.001 to *P*=.04; Table 3). The number of weeks that participants could be reached by phone was positively correlated with the number of skills accessed (Spearman r_{206}=0.32; *P*<.001).

Table 3. Overdispersed Poisson regressions predicting engagement with the virtual badges, facilitator contact, and ODB[a] enhancements from baseline predictors.

Predictor	Number of virtual badges		Facilitator contact phone contact		ODB posts	
	OR[b] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Age (years)	1.01 (1.00-1.02)	.05	1.06 (1.03-1.08)	<.001	1.01 (0.99-1.02)	.29
Gender						
Female	1.23 (0.88-1.71)	.24	4.08 (1.44-11.58)	.008	0.61 (0.40-0.92)	.02
Male (reference)	—[c]	—	—	—	—	—
Ethnicity						
Hispanic	1.60 (1.10-2.32)	.01	1.52 (0.55-4.22)	.42	1.19 (0.67-2.13)	.55
Non-Hispanic	—	—	—	—	—	—
Race						
Asian	0.84 (0.48-1.46)	.53	0.86 (0.16-4.68)	.86	1.21 (0.63-2.34)	.56
Black or African American	1.01 (0.70-1.47)	.96	1.56 (0.66-3.70)	.31	1.05 (0.62-1.77)	.86
Other race[d]	0.56 (0.33-0.95)	.06	0.44 (0.10-2.03)	.29	0.49 (0.21-1.14)	.10
White (reference)	—	—	—	—	—	—
Comfort with technology	1.05 (0.95-1.16)	.36	1.44 (1.02-2.06)	.04	1.00 (0.87-1.15)	.99
Baseline PHQ-8[e]						
Moderate	1.25 (0.84-1.85)	.27	4.88 (1.37-17.34)	.01	0.77 (0.46-1.29)	.32
Moderately severe	1.46 (0.99-2.17)	.06	5.34 (1.47-19.38)	.01	0.85 (0.51-1.43)	.54
Severe	1.06 (0.65-1.72)	.81	3.10 (0.67-14.33)	.15	0.59 (0.29-1.19)	.14
Mild (reference)	—	—	—	—	—	—

[a]ODB: online discussion board.

[b]OR: odds ratio.

[c]Not available (indicates reference value).

[d]Other race: the other race category constitutes those identified as Pacific Islander, Native American, mixed, or other.

[e]PHQ-8: Patient Health Questionnaire-8.

Participants rated the facilitator phone calls as moderately enjoyable (mean 47.20, SD 32.64; range 0-100), and the qualitative feedback supported a wide range of responses such as “I liked hearing from a real person, because then I knew someone was paying attention to what I was doing in the course” versus “It got irritating to answer the same questions and it was frustrating and embarrassing because I hadn't done the skills.” Enjoyment ratings of FC were significantly correlated with staff contact by phone (Spearman r_{107}=0.27; P=.006) and the FC enjoyment ratings were positively associated with the number of skills accessed (Spearman r_{98}=0.25; P=.01).

Analysis of the ODB

Among participants assigned to receive the ODB enhancement (n=262), 25.2% (66/262) posted on the discussion board at least once, and slightly less than half 48.5% (127/262) liked or commented on a post at least once. Male participants posted a greater number of times on the ODB than female participants (P=.02). The number of times that participants posted on the ODB was positively correlated with the number of skills accessed (Spearman r_{198}=0.49; P<.001).

Enjoyment ratings for the ODB were also in the midrange (mean 45.53, SD 30.69; range 0-100), although the ratings varied greatly. Some participants appreciated connecting with other participants through the discussion board: “It was motivational to see others progress and be able to communicate with others.” However, for others, the ODB fell short of expectations: “There doesn’t seem to be much activity there.” Still others were in the middle ground and neither liked or disliked it: “I was able to connect with other people in the program but it wasn’t very personal, in the sense that I would just comment a few blurbs here and there but didn’t feel like I was doing much.” Those who rated the ODB as more enjoyable had more posts on the board (Spearman r_{101}=0.29; P=.003). However, participants’ ODB enjoyment ratings were not associated with the number of skills accessed in the program (Spearman r_{91}=0.10; P=.34).

Analysis of VB

Among participants who were assigned to receive a VB enhancement (n=260), the number of badges earned ranged from 9 to 192, with a mean of 17.29 (SD 16.48). All participants assigned to receive a VB enhancement earned at least one badge. As seen in Table 3, older participants and Hispanic participants

earned a greater number of VB than younger and non-Hispanic participants (P=.01 to P=.05). None of the other baseline characteristics significantly predicted the total number of VB (all P=.06 to P=.96).

For enjoyment ratings, participants rated the VB as moderately enjoyable (mean 50.90, SD 33.17; range 0-100). However, there was wide variability in enjoyment ratings, and the qualitative feedback for the VB reflected this range: "It is a visual point system that is cute and fun. It seems silly but I felt like I accomplished something when I got each flower." Others did not appreciate the whimsical garden plot calling it *childish* and *a waste of time*. Although the number of badges earned and enjoyment ratings of the VB enhancement were significantly correlated (Spearman r_{115}=.30; P=.001), participants' VB enjoyment ratings did not predict the number of skills accessed (Spearman r_{109}=0.14; P=.16).

Discussion

Principal Findings

Web-based, self-guided interventions hold significant promise for people with elevated depressive symptoms. Attaining acceptable levels of adherence to these programs is critical to effectiveness, yet this has proven to be a challenge. We developed and tested 3 enhancements that we hypothesized would, either alone or in combination, improve adherence to a self-guided positive psychological intervention for people with elevated depression. The enhancements were an ODB, VB, and FC. Participants in the eight conditions that received the intervention content also received 1, 2, or all 3 or no enhancements. Our results suggest that the combination of VB and FC is especially impactful and points to areas for future focus to improve adherence across the three enhancements.

For the ODB enhancement, participants had access to an online forum where they could anonymously post questions, share their experiences, and offer encouragement to other participants. Previous findings regarding the efficacy of discussion boards for improving adherence to web-based programs have been mixed. For example, one study found that an internet discussion board or support group resulted in *lower* adherence to the self-guided program compared with the self-guided program alone [47], although participants who were assigned to receive the discussion board had a greater reduction in depression at 6 months.

The engagement on the discussion board was quite low. Only a quarter of the participants posted on the board at least once, and qualitative feedback indicated that this low level of activity further discouraged participants from making use of the ODB. Therefore, it is perhaps not surprising that there was no indication that the ODB increased adherence to the program, either alone or in combination with other enhancements. Simple availability without engagement is not sufficient for an ODB to increase adherence to the intervention content. There are a number of ways that we could have increased engagement in the ODB, such as requiring participants to post as part of the home practice assignments or making it easier to navigate to the discussion board to lower the barriers to engagement.

VB is a form of gamification and is intended to increase adherence to the intervention content by providing rewards for completing activities on the platform, for example, reaching a goal for the number of log-ins or consecutive homework completions. Participants in the arm that received both VB and FC completed significantly more sessions than did those in the no-enhancement arm (Figure 2). Furthermore, participants in arms that received VB and FC together (VB+FC; VB+FC+ODB) accessed, on average, one more skill (mean 6.42) compared with participants who received either VB or FC without the other (mean 5.42; FC alone, VB alone, ODB+VB, and ODB+FC). Participants were of a mixed opinion on whether the VB were enjoyable, with 33.9% (39/115) giving the VB low enjoyment ratings (below 31 on a 0-100 scale) and 33% (38/115) providing high enjoyment ratings (between 70 and 100). Previous studies document similar challenges to the successful use of gamification elements, such as a lack of clarity on which behaviors are being rewarded or confusion over how progress was being illustrated [48-50]. Future work to make the badges more interpretable, enjoyable, and clearly linked to target behaviors will help to increase the impact of the VB on adherence to the intervention.

Our FC enhancement consisted of staff contacting participants by phone once per week for a brief check-in regarding any challenges they were experiencing while completing that week's skills and answering questions or concerns they had. We included the FC component based on studies showing that contact with a person improves adherence to self-guided, web-based interventions [18,23,27,51,52]. Although receipt of the FC alone did not significantly improve adherence to the program, when combined with VB, there was an improvement such that those participants who received *both* FC and VB completed more of the intervention sessions compared with those who did not. The average number of times a facilitator was able to reach a participant was one, and fewer than half of the participants were ever reached by a facilitator by phone. The data indicate that the more participants enjoyed the FC, the more skills they completed, and as with VB, there was a wide range of enjoyment ratings for FC, leaving a lot of room for improvement. One easy change that may increase participants' enjoyment of the FC condition, and as a result improve the impact of FC on adherence, is to vary the questions that are asked each week and word them so they do not come across as judgmental. Qualitative feedback indicated that participants found these questions repetitive and, especially if they had not completed the home practice that week, somewhat shaming. Furthermore, some participants may prefer email over phone contact, and providing an email contact option could increase the impact of the enhancement. Other ways to improve the impact of the FC condition could include improving the *supportive accountability* [29], such as setting clear expectations and goals for engagement with the platform, and tailored performance monitoring that is supportive rather than shaming. In addition, clarifying the expertise of the facilitators and increasing trust could enhance the impact of FC [29].

Aside from the content of the enhancement, a simple count of the number of enhancements the participant was assigned to (up to 3) was not predictive of adherence. On the basis of our

data, it is not necessarily better. However, the combination of VB and FC increased adherence compared with either enhancement alone. It may be that the FC helps encourage people to engage with the platform, with the phone call providing a reminder and the facilitator providing someone to be accountable. Once participants are on the platform, the VB reinforce that engagement and provide some extrinsic motivation to remain engaged [29].

We explored potential moderators of the effect of each enhancement and found that depression influenced whether VB enhancement was associated with adherence. Specifically, among those with moderately severe baseline depression (PHQ-8 scores between 15 and 19), those who received the VB enhancement demonstrated better adherence than those who were not assigned to a VB condition. For those in the mild and moderate range of the PHQ-8 (scores between 5 and 14), the direction of the effect was the same but did not reach statistical significance. For participants with severe baseline depression (>20), the effect was in the opposite direction, such that VB were associated with (nonsignificantly) poorer adherence. Although this interaction should be interpreted with caution given the marginal significance and the very small number of participants in this group (those with the highest levels of depression who received the VB), this pattern was also apparent in our analyses with the proportion of intervention completed as an outcome (the number of pages viewed as a proportion of the total possible number of pages; Multimedia Appendix 1). It may be that for people with severe clinical depression, the badges were off-putting, and extra thought should be put into this form of gamification for people with the highest levels of depression. Of note, other studies of web-based depression interventions tend to exclude people who have scores in the severe range [10,20] so would not have picked up this potential demotivating effect of VB for people with the highest levels of depression.

Limitations

There are several limitations to consider in this study. First, the sample was recruited from all web-based sources and likely differed from clinic-based samples or those referred by medical providers. The sample was predominantly made up of non-Hispanic White individuals and females, which significantly limits the generalizability of the findings. Furthermore, participants needed to complete 4 of 7 daily emotion reports as part of a *run-in* to be assigned to a condition in which only 58.05% (602/1037) of the participants were able to accomplish. Although run-in periods are often used in clinical trials [53], by including only those who demonstrated adherence to the study protocol, we may have artificially enriched our sample with participants who were more likely to be adherent to the intervention. Comparison of those who completed the run-in to those who did not indicate that those who completed were older, more likely to be female, and less depressed than those who consented but did not complete the run-in. We also ran a large number of statistical tests and did not control for multiple comparisons. However, our analyses were hypothesis-driven and appropriate for this stage of intervention development. Finally, we tested the enhancements in the context of a positive psychological intervention, and the content may have influenced adherence, although we do not have a way to test this possibility in this study.

These findings lead to several suggestions for future self-guided interventions and potential adherence enhancements. First, the program and enhancements were feasible even for people with very high levels of depression. Unlike most previous studies of web-based programs for people with depression, we included participants with the highest levels of depression. The findings demonstrate that people with severe levels of depression can also engage in programs like this, and their inclusion in the study sample suggested important differences in terms of response to enhancements. Second, the combination of VB and FC was associated with better adherence than either enhancement alone. Researchers should go beyond single enhancements to improve adherence and consider including combinations of enhancements that may be more effective than any one enhancement alone. Furthermore, it may be helpful to offer participants a choice of which enhancements they want to access. Providing participant control over which enhancements they receive from the outset could increase feelings of engagement and investment in the program. Finally, as the technology for intervention delivery progresses, there are new possibilities for better-engaging participants. Future work may consider more immersive approaches such as virtual reality [54] and approaches such as SMS text messages [55] or chatbots [56] that foster a sense of social connection and may encourage stronger engagement with the intervention content.

Self-guided web-based interventions hold great promise to help people living with depression, and creative approaches to better engage participants will provide greater benefit to more people as the programs are more widely disseminated. Despite low levels of engagement with the FC enhancement, the results of our randomized trial suggest that the combination of VB and FC may be especially effective at improving adherence to the intervention, which in turn, may increase the impact of the intervention on well-being and bode well for future programs that incorporate enhancements such as these.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental analyses.
[DOCX File , 41 KB - jmir_v23i9e25922_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 406 KB - jmir_v23i9e25922_app2.pdf]

References

1. Kessler RC, Bromet EJ. The epidemiology of depression across cultures. Annu Rev Public Health 2013 Mar 18;34(1):119-138 [FREE Full text] [doi: 10.1146/annurev-publhealth-031912-114409] [Medline: 23514317]
2. Ghio L, Gotelli S, Marcenaro M, Amore M, Natta W. Duration of untreated illness and outcomes in unipolar depression: a systematic review and meta-analysis. J Affect Disord 2014 Jan;152-154:45-51. [doi: 10.1016/j.jad.2013.10.002] [Medline: 24183486]
3. Brenes GA, Danhauer SC, Lyles MF, Hogan PE, Miller ME. Barriers to mental health treatment in rural older adults. Am J Geriatr Psychiatry 2015 Nov;23(11):1172-1178 [FREE Full text] [doi: 10.1016/j.jagp.2015.06.002] [Medline: 26245880]
4. Hollon SD, Muñoz RF, Barlow DH, Beardslee WR, Bell CC, Bernal G, et al. Psychosocial intervention development for the prevention and treatment of depression: promoting innovation and increasing access. Biolog Psychiat 2002 Sep;52(6):610-630. [doi: 10.1016/s0006-3223(02)01384-7]
5. Wallin E, Maathz P, Parling T, Hursti T. Self-stigma and the intention to seek psychological help online compared to face-to-face. J Clin Psychol 2018 Jul 08;74(7):1207-1218. [doi: 10.1002/jclp.22583] [Medline: 29315545]
6. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
7. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
8. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: A systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
9. Karyotaki E, Kleiboer A, Smit F, Turner DT, Pastor AM, Andersson G, et al. Predictors of treatment dropout in self-guided web-based interventions for depression: an 'individual patient data' meta-analysis. Psychol. Med 2015 Apr 17;45(13):2717-2726. [doi: 10.1017/s0033291715000665]
10. Fuhr K, Schröder J, Berger T, Moritz S, Meyer B, Lutz W, et al. The association between adherence and outcome in an Internet intervention for depression. J Affect Disord 2018 Mar 15;229:443-449. [doi: 10.1016/j.jad.2017.12.028] [Medline: 29331706]
11. Horgan A, McCarthy G, Sweeney J. An evaluation of an online peer support forum for university students with depressive symptoms. Arch Psychiatr Nurs 2013 Apr;27(2):84-89. [doi: 10.1016/j.apnu.2012.12.005] [Medline: 23540518]
12. Griffiths KM, Calear AL, Banfield M. Systematic review on Internet Support Groups (ISGs) and depression (1): Do ISGs reduce depressive symptoms? J Med Internet Res 2009 Sep 30;11(3):e40 [FREE Full text] [doi: 10.2196/jmir.1270] [Medline: 19793719]
13. Ellis L, Campbell A, O'Dea B. Comparative randomized trial of an online cognitive-behavioral therapy program and an online support group for depression and anxiety. J CyberTher Rehab 2011;4(4):461-467 [FREE Full text]
14. Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014 Feb 14;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
15. Dias LP, Barbosa JL, Vianna HD. Gamification and serious games in depression care: A systematic mapping study. Telemat Informat 2018 Apr;35(1):213-224. [doi: 10.1016/j.tele.2017.11.002]
16. Deterding S, Sicart M, Nacke L, O'Hara K, Dixon D. Gamification. using game-design elements in non-gaming contexts. In: Proceddings of the Extended Abstracts on Human Factors in Computing Systems. 2011 Presented at: CHI '11: CHI Conference on Human Factors in Computing Systems; May 7 - 12, 2011; Vancouver BC Canada p. 2425-2428. [doi: 10.1145/1979742.1979575]
17. Brown M, O'Neill N, van Woerden H, Eslambolchilar P, Jones M, John A. Gamification and adherence to web-based mental health interventions: a systematic review. JMIR Ment Health 2016 Aug 24;3(3):e39 [FREE Full text] [doi: 10.2196/mental.5710] [Medline: 27558893]
18. Clarke G, Eubanks D, Reid E, Kelleher C, O'Connor E, DeBar LL, et al. Overcoming Depression on the Internet (ODIN) (2): a randomized trial of a self-help depression skills program with reminders. J Med Internet Res 2005 Jun 21;7(2):e16 [FREE Full text] [doi: 10.2196/jmir.7.2.e16] [Medline: 15998607]
19. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016 Dec 8;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
20. Kelders SM, Bohlmeijer ET, Pots WT, van Gemert-Pijnen JE. Comparing human and automated support for depression: Fractional factorial randomized controlled trial. Behav Res Ther 2015 Sep;72:72-80. [doi: 10.1016/j.brat.2015.06.014] [Medline: 26196078]

21. Addington EL, Cheung EO, Bassett SM, Kwok I, Schuette SA, Shiu E, et al. The MARIGOLD study: Feasibility and enhancement of an online intervention to improve emotion regulation in people with elevated depressive symptoms. J Affect Disord 2019 Oct 01;257:352-364 [FREE Full text] [doi: 10.1016/j.jad.2019.07.049] [Medline: 31302525]
22. Cheung E, Addington EL, Bassett SM, Schuette SA, Shiu EW, Cohn MA, et al. A self-paced, web-based, positive emotion skills intervention for reducing symptoms of depression: protocol for development and pilot testing of MARIGOLD. JMIR Res Protoc 2018 Jun 05;7(6):e10494 [FREE Full text] [doi: 10.2196/10494] [Medline: 29871853]
23. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009 Dec 15;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
24. Christensen H. Delivering interventions for depression by using the internet: randomised controlled trial. Br Med J 2004 Jan 31;328(7434):265. [doi: 10.1136/bmj.37945.566632.ee]
25. Christensen H, Griffiths KM, Korten A. Web-based cognitive behavior therapy: analysis of site usage and changes in depression and anxiety scores. J Med Internet Res 2002 Feb 15;4(1):e3 [FREE Full text] [doi: 10.2196/jmir.4.1.e3] [Medline: 11956035]
26. Clarke G, Kelleher C, Hornbrook M, Debar L, Dickerson J, Gullion C. Randomized effectiveness trial of an Internet, pure self-help, cognitive behavioral intervention for depressive symptoms in young adults. Cogn Behav Ther 2009 May 13;38(4):222-234 [FREE Full text] [doi: 10.1080/16506070802675353] [Medline: 19440896]
27. Proudfoot J, Goldberg D, Mann A, Everitt B, Marks I, Gray JA. Computerized, interactive, multimedia cognitive-behavioural program for anxiety and depression in general practice. Psychol Med 2003 Feb 14;33(2):217-227. [doi: 10.1017/s0033291702007225] [Medline: 12622301]
28. van den Berg S, Shapiro DA, Bickerstaffe D, Cavanagh K. Computerized cognitive-behaviour therapy for anxiety and depression: a practical solution to the shortage of trained therapists. J Psychiatr Ment Health Nurs 2004 Oct;11(5):508-513. [doi: 10.1111/j.1365-2850.2004.00745.x] [Medline: 15450016]
29. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
30. Tate DF, Jackvony EH, Wing RR. A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an Internet weight loss program. Arch Intern Med 2006 Aug 14;166(15):1620-1625. [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]
31. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009 May 14;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]
32. Jonathan GK, Pivaral L, Ben-Zeev D. Augmenting mHealth with human support: Notes from community care of people with serious mental illnesses. Psychiatr Rehabil J 2017 Sep;40(3):336-338 [FREE Full text] [doi: 10.1037/prj0000275] [Medline: 28891660]
33. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013 Jul;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
34. Francescato D, Mebane M, Porcelli R, Attanasio C, Pulino M. Developing professional skills and social capital through computer supported collaborative learning in university contexts. Int J Hum Comp Stud 2007 Feb;65(2):140-152. [doi: 10.1016/j.ijhcs.2006.09.002]
35. Cheng VW, Davenport T, Johnson D, Vella K, Hickie IB. Gamification in apps and technologies for improving mental health and well-being: systematic review. JMIR Ment Health 2019 Jun 26;6(6):e13717 [FREE Full text] [doi: 10.2196/13717] [Medline: 31244479]
36. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Jul;71:31-48 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]
37. Kelders SM, Sommers-Spijkerman M, Goldberg J. Investigating the direct impact of a gamified versus nongamified well-being intervention: an exploratory experiment. J Med Internet Res 2018 Jul 26;20(7):e247 [FREE Full text] [doi: 10.2196/jmir.9923] [Medline: 30049669]
38. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
39. Cohn MA, Pietrucha ME, Saslow LR, Hult JR, Moskowitz JT. An online positive affect skills intervention reduces depression in adults with type 2 diabetes. J Posit Psychol 2014 Jan 01;9(6):523-534 [FREE Full text] [doi: 10.1080/17439760.2014.920410] [Medline: 25214877]
40. Dowling GA, Merrilees J, Mastick J, Chang VY, Hubbard E, Moskowitz JT. Life enhancing activities for family caregivers of people with frontotemporal dementia. Alzheimer Dis Assoc Disord 2014;28(2):175-181. [doi: 10.1097/WAD.0b013e3182a6b905] [Medline: 24113564]
41. Moskowitz J, Carrico A, Cohn M, Duncan L, Bussolari C, Layous K, et al. Randomized controlled trial of a positive affect intervention to reduce stress in people newly diagnosed with HIV; protocol and design for the IRISS study. Open Access J Clin Trials 2014 Sep;2014(6):85-100. [doi: 10.2147/OAJCT.S64645]

42. Moskowitz JT, Carrico AW, Duncan LG, Cohn MA, Cheung EO, Batchelder A, et al. Randomized controlled trial of a positive affect intervention for people newly diagnosed with HIV. J Consult Clin Psychol 2017 May;85(5):409-423 [FREE Full text] [doi: 10.1037/ccp0000188] [Medline: 28333512]
43. Moskowitz JT, Cheung EO, Snowberg KE, Verstaen A, Merrilees J, Salsman JM, et al. Randomized controlled trial of a facilitated online positive emotion regulation intervention for dementia caregivers. Health Psychol 2019 May;38(5):391-402 [FREE Full text] [doi: 10.1037/hea0000680] [Medline: 31045422]
44. Coxe S, West SG, Aiken LS. The analysis of count data: a gentle introduction to poisson regression and its alternatives. J Pers Assess 2009 Mar 17;91(2):121-136. [doi: 10.1080/00223890802634175] [Medline: 19205933]
45. Gardner W, Mulvey EP, Shaw EC. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. Psychol Bull 1995 Nov;118(3):392-404. [doi: 10.1037/0033-2909.118.3.392] [Medline: 7501743]
46. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
47. Griffiths KM, Mackinnon AJ, Crisp DA, Christensen H, Bennett K, Farrer L. The effectiveness of an online support group for members of the community with depression: a randomised controlled trial. PLoS One 2012 Dec 28;7(12):e53244 [FREE Full text] [doi: 10.1371/journal.pone.0053244] [Medline: 23285271]
48. Zuckerman O, Gal-Oz A. Deconstructing gamification: evaluating the effectiveness of continuous measurement, virtual rewards, and social comparison for promoting physical activity. Pers Ubiquit Comput 2014 Jul 5;18(7):1705-1719. [doi: 10.1007/s00779-014-0783-2]
49. Beltran A, O'Connor T, Hughes S, Baranowski J, Nicklas TA, Thompson D, et al. Alpha test of a videogame to increase children's vegetable consumption. Games Health J 2012 Jun;1(3):219-222 [FREE Full text] [doi: 10.1089/g4h.2011.0027] [Medline: 24761315]
50. Rapp A. A qualitative investigation of gamification: Motivational factors in online gamified services and applications, in Gamification: Concepts, Methodologies, Tools, and Applications. In: Gamification: Concepts, Methodologies, Tools, and Applications. Pennsylvania, United States: IGI Global; 2015:32-48.
51. Sander L, Paganini S, Lin J, Schlicker S, Ebert DD, Buntrock C, et al. Effectiveness and cost-effectiveness of a guided internet- and mobile-based intervention for the indicated prevention of major depression in patients with chronic back pain-study protocol of the PROD-BP multicenter pragmatic RCT. BMC Psychiatry 2017 Jan 21;17(1):36 [FREE Full text] [doi: 10.1186/s12888-017-1193-6] [Medline: 28109247]
52. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
53. Laursen DR, Paludan-Müller AS, Hróbjartsson A. Randomized clinical trials with run-in periods: frequency, characteristics and reporting. Clin Epidemiol 2019 Feb;Volume 11:169-184. [doi: 10.2147/clep.s188752]
54. Maples-Keller JL, Bunnell BE, Kim S, Rothbaum BO. The Use of Virtual Reality Technology in the Treatment of Anxiety and Other Psychiatric Disorders. Harv Rev Psychiatry 2017;25(3):103-113. [doi: 10.1097/hrp.0000000000000138]
55. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
56. Greer S, Ramo D, Chang Y, Fu M, Moskowitz J, Haritatos J. Use of the chatbot "Vivibot" to deliver positive psychology skills and promote well-being among young people after cancer treatment: randomized controlled feasibility trial. JMIR Mhealth Uhealth 2019 Oct 31;7(10):e15018 [FREE Full text] [doi: 10.2196/15018] [Medline: 31674920]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
FC: facilitator contact
ODB: online discussion board
PHQ-8: Patient Health Questionnaire-8
VB: virtual badges

Edited by R Kukafka; submitted 20.11.20; peer-reviewed by A Parks, S Schueller, M Nahum, A Bedendo; comments to author 26.01.21; revised version received 16.04.21; accepted 24.05.21; published 22.09.21.

Please cite as:
Moskowitz JT, Addington EL, Shiu E, Bassett SM, Schuette S, Kwok I, Freedman ME, Leykin Y, Saslow LR, Cohn MA, Cheung EO Facilitator Contact, Discussion Boards, and Virtual Badges as Adherence Enhancements to a Web-Based, Self-guided, Positive Psychological Intervention for Depression: Randomized Controlled Trial
J Med Internet Res 2021;23(9):e25922
URL: https://www.jmir.org/2021/9/e25922
doi:10.2196/25922
PMID:34550076

Original Paper

Effect of a Brief Web-Based Educational Intervention on Willingness to Consider Human Papillomavirus Vaccination for Children in Japan: Randomized Controlled Trial

Yukio Suzuki[1,2], MD, PhD; Akiko Sukegawa[1], MD, PhD; Yutaka Ueda[3], MD, PhD; Masayuki Sekine[4], MD, PhD; Takayuki Enomoto[4], MD, PhD; Etsuko Miyagi[1], MD, PhD

[1]Department of Obstetrics and Gynecology, Graduate School of Medicine, Yokohama City University, Yokohama, Japan

[2]Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, United States

[3]Department of Obstetrics and Gynecology, Graduate School of Medicine, Osaka University, Suita, Japan

[4]Department of Obstetrics and Gynecology, Graduate School of Medical and Dental Sciences, Niigata University, Niigata, Japan

Corresponding Author:
Yukio Suzuki, MD, PhD
Department of Obstetrics and Gynecology
Graduate School of Medicine
Yokohama City University
3-9 Fukuura, Kanazawa-ku
Yokohama, 236-0004
Japan
Phone: 81 45 787 2800
Fax: 81 45 701 3536
Email: yetii@yokohama-cu.ac.jp

Abstract

Background: The human papillomavirus (HPV) vaccination rate in Japan has fallen to nearly zero since the suspension of governmental proactive recommendations in 2013, owing to the development of purported adverse events.

Objective: This study aimed to evaluate the effects of a brief web-based educational intervention using the theory of behavioral insights on the willingness of adults to consider the HPV vaccine for their daughters and sons.

Methods: We recruited 1660 participants aged 20 years or older in March 2018 via a webpage and provided them with a 10-item questionnaire related to the following aspects: awareness regarding HPV infection and vaccination, willingness for immunization, and actions for prevention. We randomly stratified participants based on sex and age with or without a brief educational intervention involving scientific information presented in an easy-to-read format.

Results: Only 484 (29.2%) of the respondents were aware of the benefits of HPV vaccination. Although only 352 (21.2%) of the respondents displayed a willingness for immunization of their daughters, there were 40 (4.8%) more respondents in the intervention group with this willingness (adjusted odds ratio [aOR] 1.32, 95% CI 1.04-1.69). In a subanalysis, the willingness toward vaccination for daughters in men was significantly higher in the intervention group (aOR 1.46, 95% CI 1.05-2.02). However, such a difference was not observed among women (aOR 1.20, 95% CI 0.83-1.73).

Conclusions: This study suggests that a brief web-based educational intervention increases the willingness of adults to consider the HPV vaccine for their children, especially among men. Thus, providing adequate information to men may be a useful strategy to improve the currently low rates of HPV vaccination.

Trial Registration: UMIN Clinical Trials Registry UMIN000049745 (UMIN-CTR); https://upload.umin.ac.jp/cgi-open-bin/ctr_e/ctr_view.cgi?recptno=R000049745

(J Med Internet Res 2021;23(9):e28355) doi:10.2196/28355

KEYWORDS

human papillomavirus; human papillomavirus vaccination; behavioral insights; behavioral change; web-based randomized controlled trial

Introduction

Background

In Japan, the human papillomavirus (HPV) nonavalent vaccine was approved by the Pharmaceuticals and Medical Devices Agency (PMDA) for girls aged 9 years or older in July 2020 [1]. The corresponding quadrivalent inoculation was ratified for boys in the same age group in December 2020 [1]. Nevertheless, the level of awareness regarding HPV, the resultant cancer (cervical cancer [CC]), and the vaccine needs to be higher in the Japanese population [2,3]. Although vaccination is the exclusive means of preventing HPV infection, the immunization rate has fallen below 1% owing to subsequent adverse events [4], which are regarded as functional disorders. These cases were reported repeatedly in Japanese media in sensational ways [2,4]. As a result of the dissemination of misinformation and the misunderstanding of the HPV vaccine, most Japanese people have distrust toward the HPV vaccine [2]. Thus, the immunization rate of the bivalent or quadrivalent HPV vaccine for the target population from the 6th grade of elementary school students to the 1st grade of high school students was 0.8% in 2018 [5]. This has persisted from 2013 for almost 8 years [4-6].

According to the strategy of the World Health Organization devised in 2019 for controlling CC [7], by 2030, 90% of girls worldwide would be vaccinated with the HPV vaccine by the age of 15 years. Malaysia, Mexico, Bhutan, Brunei, and Rwanda have achieved an immunization rate of 90% or higher in the target population [7]. Furthermore, in Australia, the success of two national programs, the National Cervical Screening Program (NCSP) and the Australian National HPV Vaccination Program (NHVP), resulted in achievement of the threshold for rare cancer in 2020 [8]. By 2028, the estimated number of cases would be less than 4 per 100,000 [9], which is the threshold for elimination [8].

In addition to HPV-related cancers, vulvar, anal, and throat malignancies may be prevented by the vaccine [9,10]. A Japanese research group provided evidence of a significant reduction in the incidence of cervical intraepithelial neoplasia grade 3 or higher in vaccinated women [11].

Specific information regarding the effects and adverse events of HPV vaccines needs to be fully disseminated to the Japanese population. Considering this, approximately 40% of the population is willing to be vaccinated [2]. Therefore, it is critical to disseminate adequate scientific knowledge regarding the beneficial effects of the vaccine so that people in the target age group actually take the vaccine.

Goal of the Study

Vaccine awareness programs are necessary, and campaigns through the media and social network services can play significant roles [2,4]. In an information-overloaded society, people frequently make decisions related to health issues based on a bunch of information [12]. Therefore, it is critical to consider the influence of behavioral insights to promote change [13]. This broadly refers to concrete approaches based on the knowledge of behavioral science and economics. The Easy, Attractive, Social, and Timely (EAST) principles are a simple way of applying behavioral insights to interventions and have been used to change human awareness and behavior [13]. This study aimed to assess the effects of these behavioral insights utilizing brief scientific information on vaccine benefits, along with statistics on CC.

Methods

Study Design and Participants

We recruited a total of 1660 participants in March 2018 via a specially designed webpage for this study. These were registered members of the research panel owned by Macromill Inc (Tokyo, Japan). The participants were 20 years old or above as on March 12-13, 2018. They were recruited until the target sample size was fulfilled. We randomly assigned each participant to respond to an identical questionnaire after (intervention group) or prior to (control group) providing behavioral insights material (BI-material) featuring brief scientific information presented in an easy-to-read format (as displayed in Figures 1 and 2).

Figure 1. Flow diagram of the randomization. BI-material: behavioral insights material featuring brief scientific information presented in an easy-to-read format.

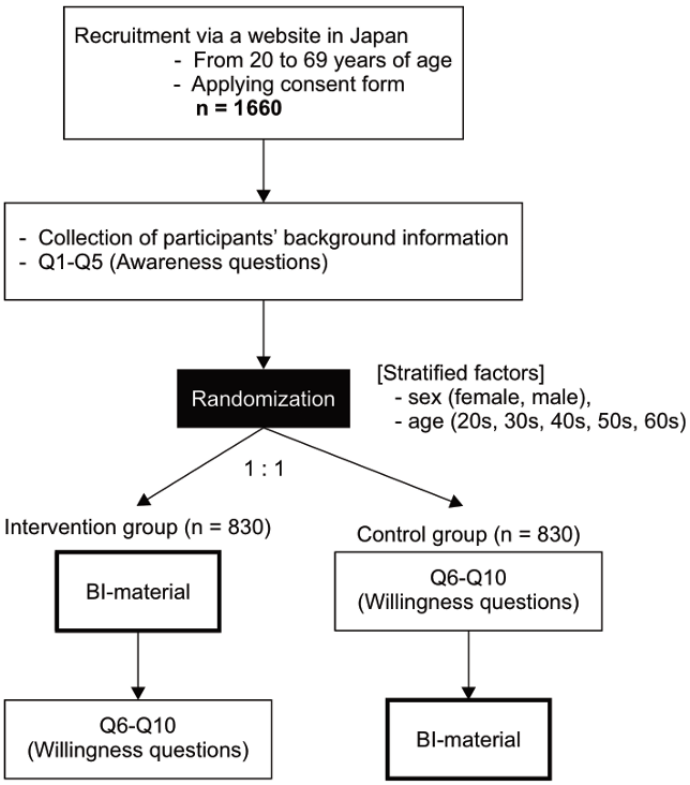
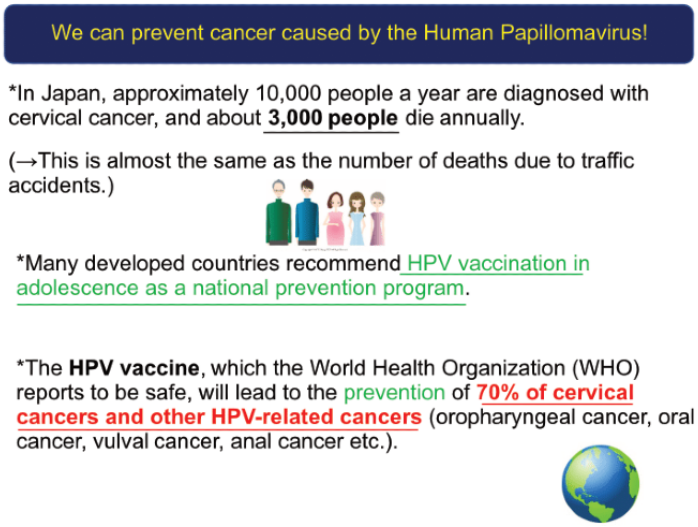


Figure 2. Behavioral insights material.



Informational Material and Behavioral Insights

Informational materials are designed primarily to increase effectiveness; they are based on a specific purpose rather than a template. In this study, first, we provided one-page BI-material containing three sentences, which were deemed to be in line with scientific evidence and objective facts associated with the theory of behavioral change and economics. This theory was suggested and represented by Richard Thaler [14]. The first sentence of our material reflected the Japanese epidemiological data on CC (Figure 2), which was intended for the framing effect [15,16]. For example, a statement like “Cervical cancer is the 11th most frequent malignancy among women in Japan” or “2.8% of women diagnosed with cervical cancer die” does not emphasize the deterioration in health. Therefore, this would not contribute to general behavior change. Additionally, we applied the prospect theory by drawing similarities between the number of deaths due to CC in Japan and fatalities due to traffic accidents. Second, to promote a change in the consciousness of participants, the presentation of social norms was included as part of the EAST framework (S-social) in the second sentence of the BI-material (Figure 2). This was employed for devising behavioral insights of greater effectiveness [17]. Third, heuristics (awareness of prejudice and intuition) [18] hypothesized a reduced interest in male participants in case the HPV immunization focused exclusively on CC. Therefore, it was conveyed that non-CC HPV malignancies are preventable

through HPV vaccination [9]. The first author (YS) is one of the specialists of the behavioral design team in Yokohama, which was established as the first nongovernmental nudge unit in Japan [19]. Thus, this BI-material was created given the behavioral insights methodology.

Randomization

Participants were randomly allocated (1:1) to each group. The intervention was performed using an automatic web-based allocation system stratified by the sex (female/women and male/men) and age (20s, 30s, 40s, 50s, and 60s) of the participants. Randomization was performed by the web-research system of Macromill Inc. Participants and investigators were blinded to the distribution (double-blinded). Once the upper limit of each stratum was reached, new participants could not be added to the web system. This ensured uniform distribution of the stratification factors. In the intervention group, we provided the BI-material prior to answering questions related to preventive awareness, following consent for the online study. The control group was provided with the same material as that provided to the intervention group after all responses were completed.

Questionnaire

The participant demographics included age, sex, marital status, number of children, sexual experience, household income, medical background, education, and tobacco use. For female participants, we also collected information regarding HPV vaccination history and previous pap screening. Medical background was defined as follows: a respondent who had a history of working as a medical professional (medical doctor, dentist, nurse, or allied health care professional) or who attended medical school.

The respondents completed a 10-item questionnaire. The first half determined HPV awareness as background information, and the second half addressed willingness to receive HPV vaccines and undergo screening tests. The respondents were instructed to answer “Yes” or “No” for each question.

The awareness questions were as follows:

Q1. It is possible to detect both cancer and precancerous lesions through CC screening.

Q2. Sexual experience is associated with HPV infection.

Q3. CC screening is necessary for women even after vaccination.

Q4. I have heard of the benefits of the HPV vaccine.

Q5. I have heard of the adverse events associated with the HPV vaccine.

The willingness questions were as follows:

Q6. If you have/had a daughter, do/would you consider getting her vaccinated against HPV?

Q7. If you have/had a son, do/would you consider getting him vaccinated against HPV?

Q8. Would you consider undergoing a pap smear? If male, will you want your family or partner to undergo a smear test?

Q9. Would you undergo the cancer screening tests recommended by the government?

Q10. Do you plan informing family members, friends, or others about cancer prevention and screening (through Facebook, LINE, Twitter, etc)?

Statistical Analysis

Tests Performed

Statistical evaluation comprised the Student *t* test, the chi-square (χ^2) test, and multiple regression analyses. These were performed using SPSS, version 27 (IBM Corp).

Power and Statistical Significance

The sample size was calculated as 80% powered to detect a 10% effect in the intervention group (increased from 40% in the control group to 50% in the intervention group) with a two-sided *P* value of .05. *P* values less than .05 were regarded as significant. The hypothetical baseline willingness rate in the control group was determined based on our previous study [2]. The sample size was calculated as 776 when the effect of the intervention estimated a 10% increase. The number of participants recruited was double of the calculated sample size because of the difficulty in estimating the baseline willingness and the intervention effect.

Odds Ratio

The adjusted odds ratios (aORs) related to the background knowledge level from the binominal logistic regression were analyzed to assess differences in both groups. In the subanalysis, questions Q2 and Q4, for which the responses showed a significant difference in the rate of awareness between the two groups (data not shown), were included as covariates in the aOR in the analysis for men. Questions Q1, Q2, and Q5, for which the responses showed a significant difference in the rate of awareness between the two groups (data not shown), were included as covariates in the aOR in the analysis for women.

Ethical Approval and Funding

The study protocol was approved by the Institutional Research Ethics Committee of Yokohama City University School of Medicine (A180200004). The trial registration number is UMIN000049745. We received research funding from the Japan Agency for Medical Research and Development (grant number 15ck0106103h0102). The survey was outsourced to Macromill Inc.

Results

Analysis of Participant Demographics

From March 12 to 13, 2018, 1660 participants were recruited. Stratifying factors, such as sex and age, were evenly allocated. The following variables displayed no significant differences between the intervention and control groups: marital status (*P*=.96), children (*P*=.84), household income (*P*=.58), sexual experience (*P*=.26), education (*P*=.44), medical background (*P*=.50), and tobacco use (*P*=.64) (Table 1).

Table 1. Characteristics and knowledge level of the participants recruited.

Characteristic	All (N=1660)	Intervention group (n=830)	Control group (n=830)	*P* value[a]
Sex, n (%)				>.99
Male	830 (50.0)	415 (50.0)	415 (50.0)	
Female	830 (50.0)	415 (50.0)	415 (50.0)	
Age (years), mean (SD)	44.7 (14.0)	44.6 (14.1)	44.8 (14.0)	.77
Age range (years)	20-69	20-69	20-69	
Age groups (years), n (%)				>.99
20-29	332 (20.0)	166 (20.0)	166 (20.0)	
30-39	332 (20.0)	166 (20.0)	166 (20.0)	
40-49	332 (20.0)	166 (20.0)	166 (20.0)	
50-59	332 (20.0)	166 (20.0)	166 (20.0)	
≥60	332 (20.0)	166 (20.0)	166 (20.0)	
Marital status, n (%)				.96
Married	1013 (61.0)	507 (61.1)	506 (61.0)	
Unmarried	647 (39.0)	323 (38.9)	324 (39.0)	
Children, n (%)				.84
Yes	904 (54.5)	454 (54.7)	450 (54.2)	
No	756 (45.5)	376 (45.3)	380 (45.8)	
Sexual experience, n (%)				.26
Experienced	1298 (78.2)	659 (79.4)	639 (77.0)	
Not experienced	145 (8.7)	64 (7.7)	81 (9.8)	
Declined to answer	217 (13.1)	107 (12.9)	110 (13.3)	
Household income (million yen/year)[b], n (%)				.58
<2	120 (7.2)	61 (7.3)	59 (7.1)	
2-4	316 (19.0)	147 (17.7)	169 (20.4)	
4-6	384 (23.1)	199 (24.0)	185 (22.3)	
6-8	208 (12.5)	105 (12.7)	103 (12.4)	
8-10	153 (9.2)	80 (9.6)	73 (8.8)	
>10	126 (7.7)	60 (7.2)	66 (8.0)	
Declined to answer	353 (21.2)	178 (21.4)	175 (21.1)	
Medical background, n (%)				.50
Medical	116 (7.0)	54 (6.5)	62 (7.5)	
Nonmedical	1544 (93.0)	776 (93.5)	769 (92.5)	
Education, n (%)				.44
Less than high school graduate	34 (2.0)	17 (2.0)	17 (2.0)	
High school graduate	467 (28.1)	237 (28.6)	230 (27.7)	
More than high school graduate	1159 (69.8)	576 (69.4)	583 (70.2)	
Tobacco use, n (%)				.64
Smoker	312 (18.8)	162 (19.5)	150 (18.1)	
Nonsmoker	896 (54.0)	439 (52.9)	457 (55.1)	
Previous smoker	452 (27.2)	229 (27.6)	223 (26.9)	
Q1. It is possible to detect both cancer and precancerous lesions through cervical cancer (CC) screening, n (%)				.11

Characteristic	All (N=1660)	Intervention group (n=830)	Control group (n=830)	*P* value[a]
Already known	670 (40.4)	319 (38.4)	351 (42.3)	
Not known	990 (59.6)	511 (61.6)	479 (57.7)	
Q2. Sexual experience is associated with HPV[c] infection, n (%)				.001
Already known	911 (54.9)	423 (51.0)	488 (58.8)	
Not known	749 (45.1)	407 (49.0)	342 (41.2)	
Q3. Cervical cancer screening is necessary for women even after vaccination, n (%)				.06
Already known	518 (31.2)	277 (33.4)	241 (29.0)	
Not known	1142 (68.8)	553 (66.6)	589 (71.0)	
Q4. Benefits of HPV vaccination, n (%)				.004
Already known	484 (29.2)	269 (32.4)	215 (25.9)	
Not known	1176 (70.8)	561 (67.6)	615 (74.1)	
Q5. Adverse events associated with HPV vaccination, n (%)				.004
Already known	543 (32.7)	299 (36.0)	244 (29.4)	
Not known	1117 (67.3)	531 (64.0)	586 (70.6)	
Last pap[d] screening[e], n (%)				.87
<2 years	341 (41.1)	178 (42.9)	163 (39.3)	
2-5 years	107 (12.9)	54 (13.0)	53 (12.8)	
>5 years	125 (15.1)	57 (13.7)	68 (16.4)	
Never	232 (28.0)	113 (27.2)	119 (28.7)	
Unknown	25 (3.0)	13 (3.1)	12 (2.9)	
HPV vaccination[e], n (%)				.70
Already vaccinated	36 (4.3)	21 (5.1)	15 (3.6)	
Not yet vaccinated	595 (71.7)	303 (73.0)	292 (70.4)	
Unknown	199 (24.0)	91 (21.9)	108 (26.0)	

[a]*P* values are estimated using the chi-square and Student *t* tests.

[b]1 USD=110 JPY.

[c]HPV: human papillomavirus.

[d]Pap: Papanicolaou test.

[e]Only female participants aged 20 years or older (n=830).

HPV Awareness Analysis

For questions Q1 to Q5 regarding HPV awareness, the recognition rate for Q2 was significantly higher (by 7.8%) in the control group than in the intervention group. For questions Q4 and Q5, the values in the intervention group were significantly higher (by 6.5% and 6.6%, respectively) than those in the control group (Table 1).

Only 484 (29.2%) of the respondents were aware of the benefits of HPV immunization (Q4), whereas 543 (32.7%) were aware of the adverse effects (Q5).

Willingness to Consider Children's Vaccination

Only 352 (21.2%) of the respondents displayed a favorable attitude toward HPV immunization for their daughters (Q6). However, an additional 40 (4.8%) participants responded affirmatively in the intervention group (aOR 1.32, 95% CI 1.04-1.69) compared to those in the control group (Table 2). For Q7, there were additional 33 (3.9%) satisfied respondents willing to consider vaccination for their sons in the intervention group (aOR 1.38, 95% CI 1.05-1.80) compared to those in the control group (Table 2).

The number of respondents expressing a desire to receive a screening test (Q8) (*P*=.37) or to communicate HPV-related issues (Q10) did not increase (*P*=.17) (Table 2).

Table 2. Comparison of attitudes toward human papillomavirus vaccination and screening tests according to intervention.

Variable	All (N=1660), n (%)	Intervention (n=830), n (%)	Control (n=830), n (%)	Yes response vs other			
				OR[a] (95% CI)	*P* value[b]	Adjusted OR[c] (95% CI)	*P* value[b]
Q6. If you have/had a daughter, do/would you consider getting her vaccinated against HPV[d]?							
Yes	352 (21.2)	196 (23.6)	156 (18.8)	1.34 (1.05-1.69)	.02	1.32 (1.04-1.69)	.02
No	240 (14.5)	115 (13.9)	125 (15.1)				
I'm not sure	1068 (64.3)	519 (62.5)	549 (66.1)				
Q7. If you have/had a son, do/would you consider getting him vaccinated against HPV?							
Yes	273 (16.4)	153 (18.4)	120 (14.5)	1.34 (1.03-1.74)	.03	1.38 (1.05-1.80)	.02
No	254 (15.3)	127 (15.3)	127 (15.3)				
I'm not sure	1133 (68.3)	550 (66.3)	583 (70.2)				
Q8. Would you consider undergoing a pap[e] smear? If male, will you want your family or partner to have a smear?							
Yes	1186 (71.4)	601 (72.4)	585 (70.5)	1.10 (0.89-1.36)	.39	1.10 (0.89-1.37)	.37
No	474 (28.6)	229 (27.6)	245 (29.5)				
Q9. Would you undergo the cancer screening tests recommended by the government?							
Yes	1128 (68.0)	574 (69.2)	554 (66.7)	1.12 (0.91-1.37)	.29	1.12 (0.91-1.38)	.30
No	532 (32.0)	256 (30.8)	276 (33.3)				
Q10. Do you plan informing family members, friends, or others about cancer prevention and screening (through Facebook, LINE, Twitter, etc)?							
Yes	760 (45.8)	396 (47.7)	364 (43.9)	1.17 (0.96-1.42)	.12	1.15 (0.94-1.40)	.17
No	900 (54.2)	434 (52.3)	466 (56.1)				

[a]OR: odds ratio.

[b]*P* value estimated using binomial logistic regression analysis.

[c]Q2, Q4, and Q5 were included as covariates in the adjusted OR.

[d]HPV: human papillomavirus.

[e]Pap: Papanicolaou test.

Sex-Wise Attitudes Toward HPV Vaccination

Table 3 presents the subanalysis results according to sex.

Differences were identified in Q6 (men: aOR 1.46, 95% CI 1.05-2.02 vs women: aOR 1.20, 95% CI 0.83-1.73) and Q7 (men: aOR 1.53, 95% CI 1.08-2.18 vs women: aOR 1.21, 95% CI 0.80-1.83). The willingness to consider vaccination for daughters in men was significantly higher in the intervention group (by 8.2%, *P*=.02; Multimedia Appendix 1); however, such a difference was not observed in women (*P*=.22; Table 3, Multimedia Appendix 2).

In an overall comparison between men and women irrespective of intervention, the willingness to consider vaccination for daughters in men was significantly higher than that in women (25.1% vs 17.3%, *P*<.001), and the willingness to consider vaccination for sons was also higher in men than in women (20.1% vs 12.8%, *P*<.001).

In the intervention group, higher rates were identified in men than in women for Q6 (29.2% vs 18.1%, *P*<.001) and Q7 (23.9% vs 13.0%, *P*<.001). While in the control group, differences were not identified between men and women for Q6 (21.0% vs 16.6%, *P*=.11) and Q7 (16.4% vs 12.5%, *P*=.11).

Table 3. Comparison of attitudes toward human papillomavirus vaccination and screening tests according to sex.

Variable	Yes response vs other			
	OR[a] (95% CI)	*P* value[b]	Adjusted OR[c] (95% CI)	*P* value[b]
Q6. If you have/had a daughter, do/would you consider getting her vaccinated against HPV[d]?				
Men	1.55 (1.13-2.13)	.01	1.46 (1.05-2.02)	.03
Women	1.11 (0.77-1.59)	.58	1.20 (0.83-1.73)	.33
Q7. If you have/had a son, do/would you consider getting him vaccinated against HPV?				
Men	1.60 (1.13-2.26)	.01	1.53 (1.08-2.18)	.02
Women	1.04 (0.70-1.57)	.84	1.21 (0.80-1.83)	.38
Q8. Would you consider undergoing a pap[e] smear? If male, will you want your family or partner to have a smear?				
Men	1.05 (0.78-1.42)	.76	1.04 (0.77-1.42)	.78
Women	1.15 (0.85-1.56)	.36	1.18 (0.87-1.61)	.30
Q9. Would you undergo the cancer screening tests recommended by the government?				
Men	1.01 (0.76-1.34)	.94	1.02 (0.76-1.36)	.92
Women	1.25 (0.93-1.69)	.15	1.31 (0.96-1.79)	.09
Q10. Do you plan informing family members, friends, or others about cancer prevention and screening (through Facebook, LINE, Twitter, etc)?				
Men	1.27 (0.96-1.67)	.09	1.24 (0.94-1.64)	.13
Women	1.08 (0.82-1.42)	.58	1.11 (0.84-1.47)	.47

[a]OR: odds ratio.

[b]*P* value estimated using binomial logistic regression analysis.

[c]Q2 and Q4 were included as covariates in the adjusted OR in the analysis involving men. Q1, Q2, and Q5 were included as covariates in the adjusted OR in the analysis involving women.

[d]HPV: human papillomavirus.

[e]Pap: Papanicolaou test.

Discussion

Principal Findings

We conducted a web-based randomized controlled trial (RCT) to assess the benefits of BI-material employing brief scientific information and its ability to motivate individuals to consider the HPV vaccine for their children. Our results showed that providing brief scientific information could increase the willingness to consider HPV vaccination for daughters and sons. This effect was observed typically among male participants. Similar minor interventions may potentially modify mindsets favorably. However, such brief digital information failed to affect the mindset in women. A possible reason why the intervention was more effective among men than women is that women had a more negative image toward HPV vaccination. There was a significant difference between men and women in the awareness level. Overall, 45.7% of women responded that they know about the adverse events of HPV vaccination, while this rate was 19.8% in men (data not shown). Such awareness might have influenced the difference in the intervention effect.

In terms of the COVID-19 pandemic, a difference in awareness of prevention strategies was observed between men and women [20]; therefore, it is essential to build a method appropriate for sex subgroups to transform general behavior via the internet and social networking services.

Comparison With Prior Work

According to a systematic review regarding the effect of the pedagogical approach on social awareness or action, more reliable and validated studies are required to change the perception or mindset of the target population [21]. This paves the way for mitigating hesitancy toward vaccination [21]. Therefore, this RCT may be valuable to ensure a change in public attitudes toward vaccination.

A systematic review revealed that general communication about childhood vaccination resulted in a positive change by 20%; however, this excluded the HPV vaccine [21]. In contrast, the brief educational intervention was observed to improve the willingness to consider the HPV vaccine for daughters and sons, with aORs of 1.32 and 1.38, respectively (Table 2). Therefore, when compared to the review data of the in-person approach [21], the aORs were not small. Furthermore, the BI-material was designed by a specialist of behavioral insight, resulting in a moderate cost; therefore, this intervention is more reliable and cost-effective than those used in previous studies.

A US RCT reported that education through social media is effective for improving general awareness regarding vaccination. However, the study was limited to a sample of 58 participants [22]. A phase 2 Japanese trial evaluated how the extent of intervention affected HPV vaccination acceptance and reported that providing appropriate medical information resulted in

beneficial effects [23]. Thus, this study proved the utility of effective and adequate guidance in improving vaccine acceptance. There is a paucity of literature on the efficacy of pragmatic educational materials for HPV vaccination promotion [22]. Therefore, more extensive trials focusing on the mode of endorsing HPV vaccination like our trial are required for evidence-based promotion.

Limitations

We recognize several limitations in this study. First, the sustainability of effective change was not evaluated. Typically, with respect to health issues, taking action requires time. Therefore, a study should assess not only a change in mindset, but also the appropriate course of action. We have already performed an RCT (UMIN000039273) assessing the sustainability of general acceptance and concrete behavior for HPV vaccination. Second, the impact of the study on the behavioral outcome was unclear; specific vaccination functioning needs to be tracked. Third, selection bias was present as the respondents were Japanese individuals enrolled by an internet survey company.

Based on the sex-wise subanalysis, improved information in male participants may be the key to improve the rate of HPV vaccination in Japan. Additionally, video-based content with patient feedback is expected to result in a broader impact. Thus, a varied approach for men and women may be required.

Conclusions

Inference of the Study Findings

Our study revealed a positive outlook toward HPV vaccination following a brief web-based educational intervention, especially among men. Such an approach is extremely effective to overcome challenges related to communication and information overload.

Impact of the Findings

A brief web-based educational intervention based on the theory of behavioral insights increases the willingness of Japanese adults to consider the HPV vaccine for their daughters and sons.

Acknowledgments

The authors would like to thank the members of the research group affiliated to Niigata University, Osaka University, and Yokohama City University for their cooperation. We would like to thank Editage (http://www.editage.jp) for English language editing. We also deeply thank Macromill Inc (Tokyo, Japan) for their inputs in this research study, and Yokohama Behavioral Insights and Design Team for offering suggestions on behavioral insights.

Authors' Contributions

All authors have made substantial contributions to the study and manuscript preparation. YS was responsible for the study design, data collection, analysis, and interpretation. AS, YU, MS, TE, and EM assisted in data interpretation. All authors have read and approved the final manuscript.

Conflicts of Interest

EM received grants and lecture fees from Merck Sharp & Dohme Corp and Chugai Pharmaceutical Co, Ltd. EM also received honoraria from Takeda Pharmaceutical Co, Ltd and AstraZeneca plc.

Multimedia Appendix 1
Comparison of attitudes toward human papillomavirus vaccination and screening tests in men.
[DOCX File , 23 KB - jmir_v23i9e28355_app1.docx]

Multimedia Appendix 2
Comparison of attitudes toward human papillomavirus vaccination and screening tests in women.
[DOCX File , 22 KB - jmir_v23i9e28355_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 97 KB - jmir_v23i9e28355_app3.pdf]

References

1. New drugs approved in May 2020. Pharmaceuticals and Medical Devices Agency. 2020. URL: https://www.pmda.go.jp/PmdaSearch/iyakuDetail/GeneralList/631341C [accessed 2021-09-03]
2. Suzuki Y, Sukegawa A, Nishikawa A, Kubota K, Motoki Y, Asai-Sato M, et al. Current knowledge of and attitudes toward human papillomavirus-related disease prevention among Japanese: A large-scale questionnaire study. J Obstet Gynaecol Res 2019 May;45(5):994-1005. [doi: 10.1111/jog.13929] [Medline: 30723974]

3. Hanley SJB, Yoshioka E, Ito Y, Konno R, Sasaki Y, Kishi R, et al. An exploratory study of Japanese fathers' knowledge of and attitudes towards HPV and HPV vaccination: does marital status matter? Asian Pac J Cancer Prev 2014;15(4):1837-1843 [FREE Full text] [doi: 10.7314/apjcp.2014.15.4.1837] [Medline: 24641418]
4. Sekine M, Kudo R, Yamaguchi M, Hanley SJB, Hara M, Adachi S, et al. Japan's Ongoing Crisis on HPV Vaccination. Vaccines (Basel) 2020 Jul 06;8(3):362 [FREE Full text] [doi: 10.3390/vaccines8030362] [Medline: 32640691]
5. Ministry of Health, Labour and Welfare Committee on vaccine adverse reactions. Ministry of Health, Labour and Welfare. URL: https://www.mhlw.go.jp/stf/shingi/shingi-kousei_284075.html [accessed 2021-09-03]
6. Hanley SJB, Yoshioka E, Ito Y, Kishi R. HPV vaccination crisis in Japan. The Lancet 2015 Jun 20;385(9987):2571-2537 [FREE Full text] [doi: 10.1016/S0140-6736(15)61152-7] [Medline: 26122513]
7. Global strategy to accelerate the elimination of cervical cancer as a public health problem. World Health Organization. 2019. URL: https://www.who.int/publications/i/item/9789240014107 [accessed 2021-09-03]
8. Hall MT, Simms KT, Lew J, Smith MA, Brotherton JM, Saville M, et al. The projected timeframe until cervical cancer elimination in Australia: a modelling study. The Lancet Public Health 2019 Jan;4(1):e19-e27 [FREE Full text] [doi: 10.1016/S2468-2667(18)30183-X] [Medline: 30291040]
9. Luostarinen T, Apter D, Dillner J, Eriksson T, Harjula K, Natunen K, et al. Vaccination protects against invasive HPV-associated cancers. Int J Cancer 2018 May 15;142(10):2186-2187 [FREE Full text] [doi: 10.1002/ijc.31231] [Medline: 29280138]
10. Lei J, Ploner A, Elfström KM, Wang J, Roth A, Fang F, et al. HPV Vaccination and the Risk of Invasive Cervical Cancer. N Engl J Med 2020 Oct 01;383(14):1340-1348. [doi: 10.1056/NEJMoa1917338] [Medline: 32997908]
11. Kudo R, Yamaguchi M, Sekine M, Adachi S, Ueda Y, Miyagi E, et al. Bivalent Human Papillomavirus Vaccine Effectiveness in a Japanese Population: High Vaccine-Type-Specific Effectiveness and Evidence of Cross-Protection. J Infect Dis 2019 Jan 09;219(3):382-390 [FREE Full text] [doi: 10.1093/infdis/jiy516] [Medline: 30299519]
12. Soroya SH, Farooq A, Mahmood K, Isoaho J, Zara S. From information seeking to information avoidance: Understanding the health information behavior during a global health crisis. Inf Process Manag 2021 Mar;58(2):102440 [FREE Full text] [doi: 10.1016/j.ipm.2020.102440] [Medline: 33281273]
13. Behavioural insights. Organisation for Economic Co-operation and Development. URL: http://www.oecd.org/gov/regulatory-policy/behavioural-insights.htm [accessed 2021-02-21]
14. Thaler R, Sunstein C. Nudge: Improving Decisions About Health, Wealth, and Happiness. New York, NY: Penguin Books; 2009.
15. Kahneman D, Tversky A. Prospect Theory: An Analysis of Decision under Risk. Econometrica 1979 Mar;47(2):263. [doi: 10.2307/1914185]
16. Gong J, Zhang Y, Yang Z, Huang Y, Feng J, Zhang W. The framing effect in medical decision-making: a review of the literature. Psychol Health Med 2013;18(6):645-653. [doi: 10.1080/13548506.2013.766352] [Medline: 23387993]
17. EAST: Four Simple Ways to Apply Behavioural Insights. Behavioural Insights Team. URL: https://www.bi.team/publications/east-four-simple-ways-to-apply-behavioural-insights/ [accessed 2021-02-21]
18. Coussens S. Behaving Discretely: Heuristic Thinking in the Emergency Department. SSRN Journal 2018 Aug:1-51. [doi: 10.2139/ssrn.3743423]
19. Yokohama Behavioral Insights and Design Team (YBiT). URL: https://ybit.jp/top-en [accessed 2021-09-03]
20. Li S, Feng B, Liao W, Pan W. Internet Use, Risk Awareness, and Demographic Characteristics Associated With Engagement in Preventive Behaviors and Testing: Cross-Sectional Survey on COVID-19 in the United States. J Med Internet Res 2020 Jun 16;22(6):e19782 [FREE Full text] [doi: 10.2196/19782] [Medline: 32501801]
21. Kaufman J, Ryan R, Walsh L, Horey D, Leask J, Robinson P, et al. Face-to-face interventions for informing or educating parents about early childhood vaccination. Cochrane Database Syst Rev 2018 May 08;5:CD010038 [FREE Full text] [doi: 10.1002/14651858.CD010038.pub3] [Medline: 29736980]
22. Brandt HM, Sundstrom B, Monroe CM, Turner-McGrievy G, Larsen C, Stansbury M, et al. Evaluating a Technology-Mediated HPV Vaccination Awareness Intervention: A Controlled, Quasi-Experimental, Mixed Methods Study. Vaccines (Basel) 2020 Dec 10;8(4):749 [FREE Full text] [doi: 10.3390/vaccines8040749] [Medline: 33321975]
23. Mizumachi K, Aoki H, Kitano T, Onishi T, Takeyama M, Shima M. How to recover lost vaccine acceptance? A multi-center survey on HPV vaccine acceptance in Japan. J Infect Chemother 2021 Mar;27(3):445-449. [doi: 10.1016/j.jiac.2020.10.012] [Medline: 33127289]

Abbreviations

aOR: adjusted odds ratio
BI: behavioral insights
CC: cervical cancer
EAST: Easy, Attractive, Social, and Timely
HPV: human papillomavirus
OR: odds ratio

RCT: randomized controlled trial

Edited by R Kukafka; submitted 03.03.21; peer-reviewed by L Walsh, J Kaewkungwal; comments to author 19.05.21; revised version received 20.05.21; accepted 10.08.21; published 27.09.21.

Please cite as:
Suzuki Y, Sukegawa A, Ueda Y, Sekine M, Enomoto T, Miyagi E
Effect of a Brief Web-Based Educational Intervention on Willingness to Consider Human Papillomavirus Vaccination for Children in Japan: Randomized Controlled Trial
J Med Internet Res 2021;23(9):e28355
URL: https://www.jmir.org/2021/9/e28355
doi:10.2196/28355
PMID:34569941

Original Paper

Use of a Mobile Lifestyle Intervention App as an Early Intervention for Adolescents With Obesity: Single-Cohort Study

Chu Shan Elaine Chew[1], MBBS, MRCPCH, MCI; Courtney Davis[1], MPH, MD; Jie Kai Ethel Lim[1], BSc (Hons); Chee Meng Micheal Lim[2], BSc (Hons), GDHM; Yi Zhen Henny Tan[1], MClinPsych; Jean Yin Oh[1], MB BCh BAO, MRCPCH; Kumudhini Rajasegaran[1], MB BCh BAO, MRCPCH; Yong Hwa Michael Chia[3], PhD; Eric Andrew Finkelstein[4], MHA, PhD

[1]KK Women's and Children's Hospital, Singapore, Singapore

[2]ReMark Group, Singapore, Singapore

[3]National Institute of Education, Singapore, Singapore

[4]Duke-NUS Medical School, Health Services and Systems Research, Singapore, Singapore

Corresponding Author:
Chu Shan Elaine Chew, MBBS, MRCPCH, MCI
KK Women's and Children's Hospital
100 Bukit Timah Road
Singapore, 229899
Singapore
Phone: 65 62255554
Email: elaine.chew.c.s@singhealth.com.sg

Abstract

Background: Effective, resource-efficient treatment is urgently needed to address the high rates of pediatric and adolescent obesity. This need has been accelerated by the COVID-19 pandemic. The use of a mobile health tool as an early intervention before a clinic-based multidisciplinary weight management program could be an effective treatment strategy that is appropriate during a pandemic.

Objective: This study aims to assess the effectiveness of and adolescent engagement with a mobile app–based lifestyle intervention program as an early intervention before enrollment in a clinic-based multidisciplinary weight management program.

Methods: This prospective single-cohort study involved adolescents, aged 10-16 years, who were overweight and obese (defined as BMI percentile above the 85th percentile). Participants used the mobile Kurbo app as an early intervention before enrolling in a clinic-based multidisciplinary weight management program. Kurbo's health coaches provided weekly individual coaching informed by a model of supportive accountability via video chat, and participants self-monitored their health behavior. The implementation of Kurbo as an early intervention was evaluated using the reach, effectiveness, adoption, implementation, and maintenance framework by reach (number who consented to participate out of all patients approached), implementation (Kurbo engagement and evaluation), and effectiveness as measured by the primary outcome of the BMI z-score at 3 months. Secondary outcome measures included changes in body fat percentage, nutrition and physical activity levels, and quality of life at 3 months. Maintenance was defined as the outcome measures at 6-month follow-up.

Results: Of the 73 adolescents who were approached for enrollment, 40 (55%) of adolescents were recruited. The mean age was 13.8 (SD 1.7) years, and the mean BMI z-score was 2.07 (SD 0.30). In the multiethnic Asian sample, 83% (33/40) of the participants had household incomes below the national median. Kurbo engagement was high, with 83% (33/40) of participants completing at least 7 coaching sessions. In total, 78% (18/23) of participants rated the app as good to excellent and 70% (16/23) stated that they would recommend it to others. There were no statistically significant changes in BMI z-scores at 3 months (*P*=.19) or 6 months *(P*=.27). Participants showed statistically significant improvements in measured body fat percentage, self-reported quality of life, and self-reported caloric intake from the 3-day food diaries at 3 and 6 months.

Conclusions: The use of Kurbo before enrollment in an outpatient multidisciplinary clinical care intervention is a feasible strategy to expand the reach of adolescent obesity management services to a low-income and racially diverse population. Although there was no significant change in BMI z-scores, the use of Kurbo as an early intervention could help to improve quality of life and reduce body fat percentage and total caloric intake.

(J Med Internet Res 2021;23(9):e20520) doi:10.2196/20520

KEYWORDS

pediatric obesity; mobile health; apps; health behavior; mHealth; obesity; adolescent; lifestyle; well-being; mobile phone

Introduction

Background

Mirroring global trends, the prevalence of overweight among Singaporean adolescents increased from 2.2% in 1975 to 15.9% in 2016. Overweight adolescents are at a higher risk for adult obesity as well as short- and long-term medical and psychosocial complications [1]. To address this concern in a resource-efficient manner, experts have proposed a staged care approach for the management of adolescents who are overweight and obese. Those who are less successful in reaching a healthy weight through less-intensive interventions, such as mobile health (mHealth) apps [2], are then recommended for higher stage interventions. mHealth apps have the potential for broad reach and adoption among Singaporean youth because of their low cost, widespread internet availability, and high levels of smartphone ownership [3-5]. Given the current COVID-19 pandemic, mHealth is a timely and essential tool to engage adolescents when in-person services are unavailable and when COVID-19 response strategies contribute to decreased physical activity and increase other unhealthy lifestyle behaviors [6,7,8].

The commercially available mHealth weight management program, Kurbo, has provided individualized health coaching, educational videos on nutrition and physical activity, and self-monitoring to improve diet and physical activity behaviors [9]. Prior research among Kurbo users showed that increased engagement with the Kurbo features, including the web-based coaching sessions, was associated with greater weight loss, suggesting that Kurbo may be appropriate as an early-stage intervention among adolescents with obesity [10].

Objectives

Therefore, the primary objective of this study is to examine the implementation of Kurbo as an early intervention for adolescents with obesity before enrollment in a clinic-based multidisciplinary weight management program in Singapore. The evaluation used the relevant dimensions of the reach, effectiveness, adoption, implementation, and maintenance evaluation framework [11].

Methods

Study Design

This was a prospective, single-arm study that conducted evaluations at three time points: baseline, 3 months, and 6 months. A total of 40 participants, with informed parental consent and child assent, were enrolled in the study at the point of referral to the KK Women's and Children's Hospital (KKH) weight management clinic (WMC). Participants were enrolled between October 2018 and March 2019. All study procedures were approved by the Singhealth Centralized Institutional Review Board. The study was registered at ClinicalTrials.gov (NCT03561597).

Multidisciplinary WMC Clinic

KKH is an 830-bed tertiary pediatric teaching hospital that provides two-third of the government-subsidized pediatric care in Singapore. The adolescent WMC is a physician-led multidisciplinary clinic where adolescents with overweight and their families engage with a multidisciplinary care team consisting of physicians, dietitians, exercise physiologists, and psychologists to set and monitor behavioral goals to manage obesity-related comorbidities. The KKH WMC protocols and outcomes have been previously published, with a historical dropout rate of 58% [12].

The usual waiting time for a first visit to the WMC clinic varies from 4 to 8 weeks after the initial referral. For this study, WMC providers had access to information about participants' Kurbo progress through an administrator site during this period. This allowed for monitoring safety concerns and guiding discussions during clinic visits. After the first WMC visit, dietary recommendations and counseling were provided by the dietitian according to the recommended nutritional guidelines. Physical activity counseling was also performed by exercise trainers based on the World Health Organization guidelines on physical activity and sedentary behavior.

Study Participants

Adolescents, referred to the WMC, aged 10-17 years with a BMI percentile above the 85th percentile [13], were eligible for enrollment. Adolescents with secondary causes of obesity, such as Cushing syndrome, those whose parents were non-English speakers, and those without smartphone access were not eligible for enrollment. Study participants were asked to download the Kurbo app onto their mobile phones and enroll in the app at the point of study enrollment. The participants had free access to the Kurbo program.

Kurbo Program

Kurbo is a mobile app developed to aid adolescents and their families with weight management through dietary self-monitoring and weekly coaching sessions (Figure 1). The details of the Kurbo mobile app and use of supportive accountability in the coaching program have been described previously [10,14]. In short, the Kurbo program consists of a mobile app for self-monitoring of eating, physical activity behaviors, and weight and individualized coaching sessions by Kurbo-certified behavioral coaches. The mobile app's data-driven platform provides users with feedback via push notifications, text messages, and emails. Using the traffic light diet to categorize foods [15], Kurbo promotes the gradual reduction of high-calorie (red) foods over time. Study participants were asked to use the mobile app to log their daily food intake and were encouraged to gradually reduce their *red* food consumption and increase consumption of *green* food. They were also recommended to monitor their daily physical activity levels and to work toward the recommended 60 minutes of moderate-to-vigorous physical activity (MVPA) each day.

Weight tracking and uploading into the app were recommended at least weekly. Individualized coaching included a weekly check-in with the coach for 15 minutes via video, phone, or text over a 12-week period. Participants were paired with the same coach for the duration of the program. Coaches monitored participants' self-report of weight, physical activity, and dietary behaviors and provided individualized feedback based on the information provided. After each coaching session, coaches emailed a copy of the session summary and a tailored plan for the coming week.

Figure 1. Screenshot of Kurbo app.

Measures

Demographic and parental characteristics were obtained at the baseline. Questionnaires and food diaries were obtained through self-reports using physical forms during the study visits.

Evaluation

The implementation of the Kurbo program as an early intervention was evaluated using the relevant dimensions of reach, effectiveness, adoption, implementation, and maintenance framework [11] as follows: (1) reach, number who consented to participate out of all patients approached; (2) implementation, program engagement as detailed for the 12-week program and program evaluation; (3) effectiveness, measured by 3-month BMI *z*-score changes from baseline and changes in blood pressure, nutritional intake physical activity levels, quality of life, and disordered eating behaviors as defined below; and (4) maintenance, measured using the identical outcomes at 6 months and 3 months after the program conclusion. Throughout the 12-week program, we also recorded the weekly frequency of information entered for meal consumption, body weight, physical activity, and coaching sessions. We evaluated satisfaction with the Kurbo program through solicitation of participants' feedback on ease of program navigation and overall use as well as their likelihood to recommend Kurbo to others.

Anthropometric and Blood Pressure Measurements

Height, weight, waist circumference [16], body fat percentage [17], and blood pressure were measured by trained staff at each study visit in the clinic. Height was measured to the nearest 0.1 cm via a stadiometer (Seca, Model 220). Weight was measured to the nearest 0.1 kg using a medical weighing scale without shoes and in light clothing. Body fat percentage was assessed

using bioimpedance analysis (Impedimed, DF50 Body Composition Analyser). Blood pressure was measured using an electronic sphygmomanometer (Dinamap model 8101, Critikon Inc). Anthropometric measurements were taken at baseline and at 3 and 6 months.

Nutritional Intake

Adolescents' daily total caloric intake and fruit and vegetable consumption were assessed using a 3-day food diary that has been previously validated for use with Singaporean adolescents [18,19]. The food diary was administered at baseline and at 3 and 6 months.

Physical Activity Levels

Physical activity was assessed using a wGT3X+ ActiGraph accelerometer. Participants wore the accelerometer for a 7-day period at baseline, 3 months, and 6 months. The ActiGraph data were processed using ActiLife 6 software. The Puyau cutoff point of 3200 counts per minute was used to estimate the time spent in MVPA. When 20 minutes of consecutive zeros were present in the accelerometry data, it was assumed that the monitor was not being worn at that time. All days with >500 minutes of valid data were included in the analysis [20-22].

Psychosocial Outcomes

Self-reported questionnaires were administered to adolescents at baseline, 3 months, and 6 months. The Pediatric Quality of Life Inventory (PedsQoL; UK version 4) was administered to evaluate physical, emotional, school, and social functioning. The Eating Pattern Inventory for Children (EPI-C) [23] was used to assess four dimensions of psychological eating behavior: external eating, emotional eating, dietary restraint, and parental pressure to eat. Higher scores were indicative of concerning behavior in each dimension. Both the PedsQoL and EPI-C have been validated in adolescent cohorts, including overweight adolescents [24,25].

Statistical Analysis

Data were analyzed using SPSS version 9.3 for Windows (IBM Corp). BMI was computed as kg/m^2, and the BMI z-score was calculated using the L, M, and S parameters published by the Centers for Disease Control and Prevention [13]. Baseline demographic, anthropometric measurements, and parent characteristics were compared between study completers and those who dropped out of the study using the Wilcoxon rank-sum test and two-sample, 2-tailed *t* tests for nonnormal and normal continuous variables, respectively, and the Fisher exact test for categorical variables. Changes in anthropometric measurements, quality of life, disordered eating behaviors, MVPA, total caloric intake, and blood pressure measurements between the first visit and months 3 and 6 were compared using paired-sample *t* tests. Linear regression was used to determine the association between BMI z-scores at 3 months (primary outcome) and the number of WMC visits attended during the Kurbo program. The level of statistical significance was set at $P<.05$.

Results

Implementation of Kurbo Program as an Early Intervention

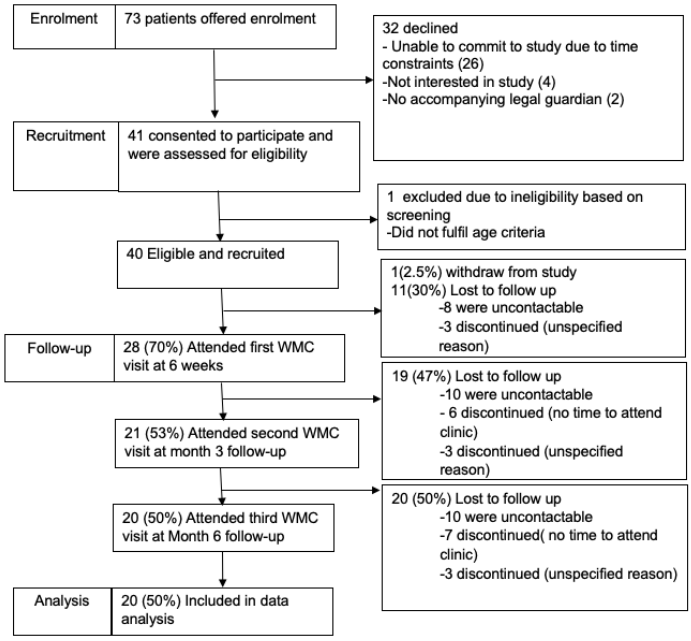
Reach

Of the 73 eligible participants, 40 (55%) were consented and enrolled in the study. The mean age of the participants was 13.8 (SD 1.7) years. Among this, 58% (23/40) of the enrolled adolescents were male, 45% (18/40) were Chinese, 33% (13/40) were Malay, and 13% (5/40) were Indian. Of 40 participants, 32 (80%) were referred to WMC as a result of opportunistic screening during medical visits for nonobesity-related conditions and 8 (20%) were referred specifically for obesity-related comorbidities. Moreover, 65% (26/40) of participants had a family history of metabolic diseases, and 83% (33/40) of accompanying parents were overweight or obese. Overall, 83% (33/40) of participants had household income less than the national median monthly household income of SGD 9500 (US $7125) [26], and approximately one-third had a monthly household income below SGD 2000 (US $1500). Half of the accompanying parents had a secondary school–level education or less. Other baseline characteristics of the participants are summarized in Table 1. There were no significant differences in baseline characteristics between those who completed the study and those who dropped out (Table 1). In total, 53% (21/40) of adolescents completed the 3-month assessment, and 50% (20/40) completed the 6-month assessment (Figure 2).

Table 1. Baseline characteristics of participants in the study (N=40).

Characteristics	Total (N=40)	Completers (n=20)	Drop-outs (n=20)	*P* value
Baseline characteristics				
Gender (male), n (%)	23 (58)	12 (60)	11 (55)	.79
Ethnicity, n (%)				.25
Chinese	18 (45)	12 (60)	6 (30)	
Malay	13 (33)	3 (15)	11 (55)	
Indian	5 (13)	2 (10)	3 (15)	
Other	4 (10)	4 (35)	0 (0)	
Weight category, n (%)				.32
Overweight	8 (20)	5 (25)	3 (15)	
Obesity	32 (80)	15 (75)	17 (85)	
Family history of metabolic diseases, n (%)	26 (65)	12 (60)	14 (74)	.37
Age (years), mean (SD)	13.8 (1.7)	14.6 (1.5)	13.4 (1.8)	.62
Body mass (kg), mean (SD)	81.2 (17.2)	83.8 (19.1)	78.6 (15.1)	.34
Height (cm), mean (SD)	161.9 (11.4)	163.3 (12.2)	160.4 (10.7)	.43
BMI (kg/m^2), mean (SD)	30.7 (3.9)	30.6 (4.3)	30.3 (3.5)	.79
BMI z-scores, mean (SD)	2.07 (0.30)	2.05 (0.34)	2.09 (0.25)	.68
Waist-to-height ratio, mean (SD)	0.61 (0.06)	0.60 (0.06)	0.61 (0.05)	.80
Body fat percentage (%), mean (SD)	43.3 (5.9)	42.8 (5.8)	44 (6)	.53
Blood pressure (mm Hg), mean (SD)				
Systolic	121 (13)	120 (13)	117 (10)	.26
Diastolic	69 (9)	68 (10)	67 (5)	.60
Accompanying parental baseline characteristics				
Age (years), mean (SD)	43.6 (5.3)	44.7 (5.2)	42.5 (5.2)	.19
Gender (female), n (%)	33 (82)	17 (85)	16 (80)	.68
Weight status, n (%)				
Overweight	11 (28)	7 (35)	4 (20)	.57
Obesity	22 (55)	10 (50)	12 (60)	.57
Marital status, n (%)				.79
Married parent	33 (82)	17 (85)	16 (80)	
Single parent	7 (18)	3 (15)	4 (20)	
Highest education level, n (%)				.11
Secondary school (equivalent to 10 years of education)	21 (53)	8 (40)	13 (65)	
Diploma	11 (27)	8 (40)	3 (15)	
Bachelor’s degree and above	8 (20)	4 (20)	4 (20)	
Monthly household income (SGD; US $), n (%)				.27
Below 1500 (1125)	13 (33)	4 (20)	9 (45)	
1500-4499 (1125-3374.25)	12 (31)	9 (45)	3 (15)	
4500-7500 (3375-5625)	11 (26)	6 (30)	5 (25)	
Not reported	4 (10)	1 (5)	3 (15)	

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) diagram showing the flow of participants through each stage of the trial. WMC: weight management clinic.



Engagement and Evaluation

Overall, 83% (33/40) of participants completed at least 1 health coaching session. Participants completed a median of 7 (IQR 2-10) weekly sessions. Initial participant engagement across all Kurbo components was initially high but decreased over time (Figure 3). During the 12-week program, participants logged mean of 4.1 (SD 2.1) weight measurements. On average, participants tracked mean *green foods* 3.8 (SD 3.9) times and *red foods* 4.6 (SD 3.7) times and recorded physical activity 3.2 (SD 2.6) times each week. In addition, 18% (7/40) of participants did not engage in any health coaching session, weight, meals, or physical activity tracking.

A total of 23 participants completed the evaluation of Kurbo and individual components, as shown in Figures 4 and 5. Participants rated communication with the health coach as the most user-friendly and useful component. A total of 21 participants reported following the health coaches' advice to a moderate-to-large extent. The participants did not report difficulties in using the app. Moreover, 18 participants rated the app as good to excellent, with 16 participants stating that they would recommend it to others.

Figure 3. Number of participants who logged in at least one weight, food, or physical activity by week.

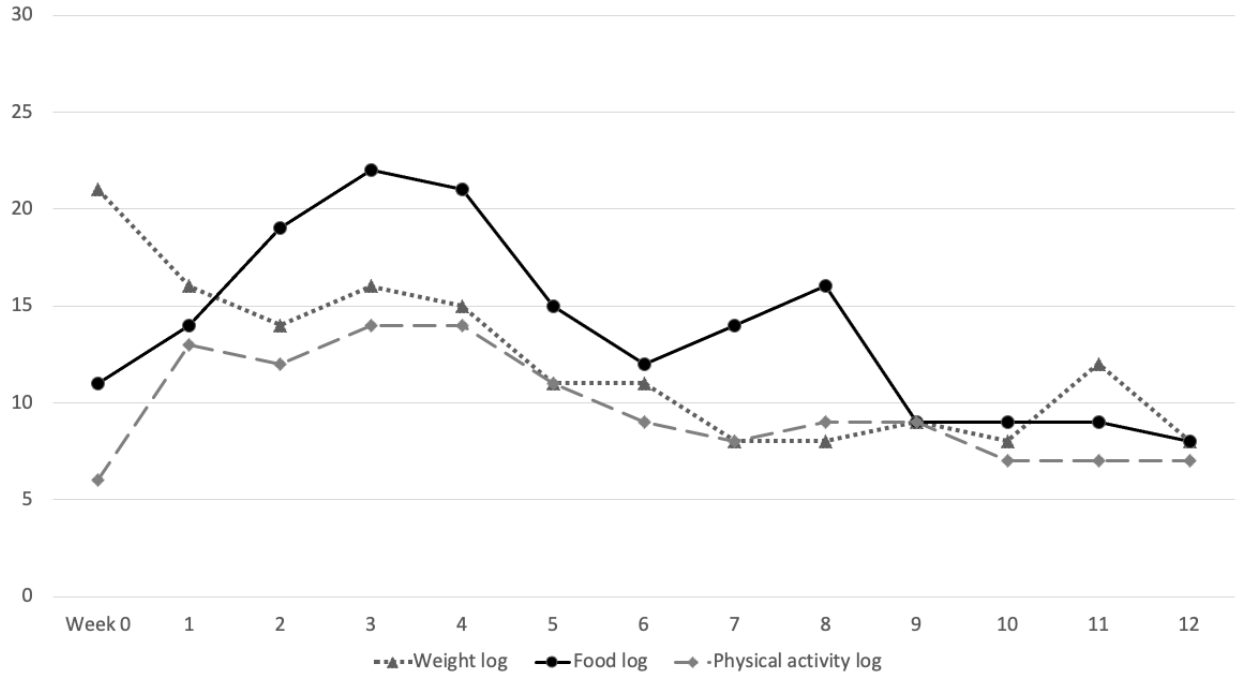


Figure 4. User-friendliness of the various components of Kurbo where 1=very difficult and 5=very easy.

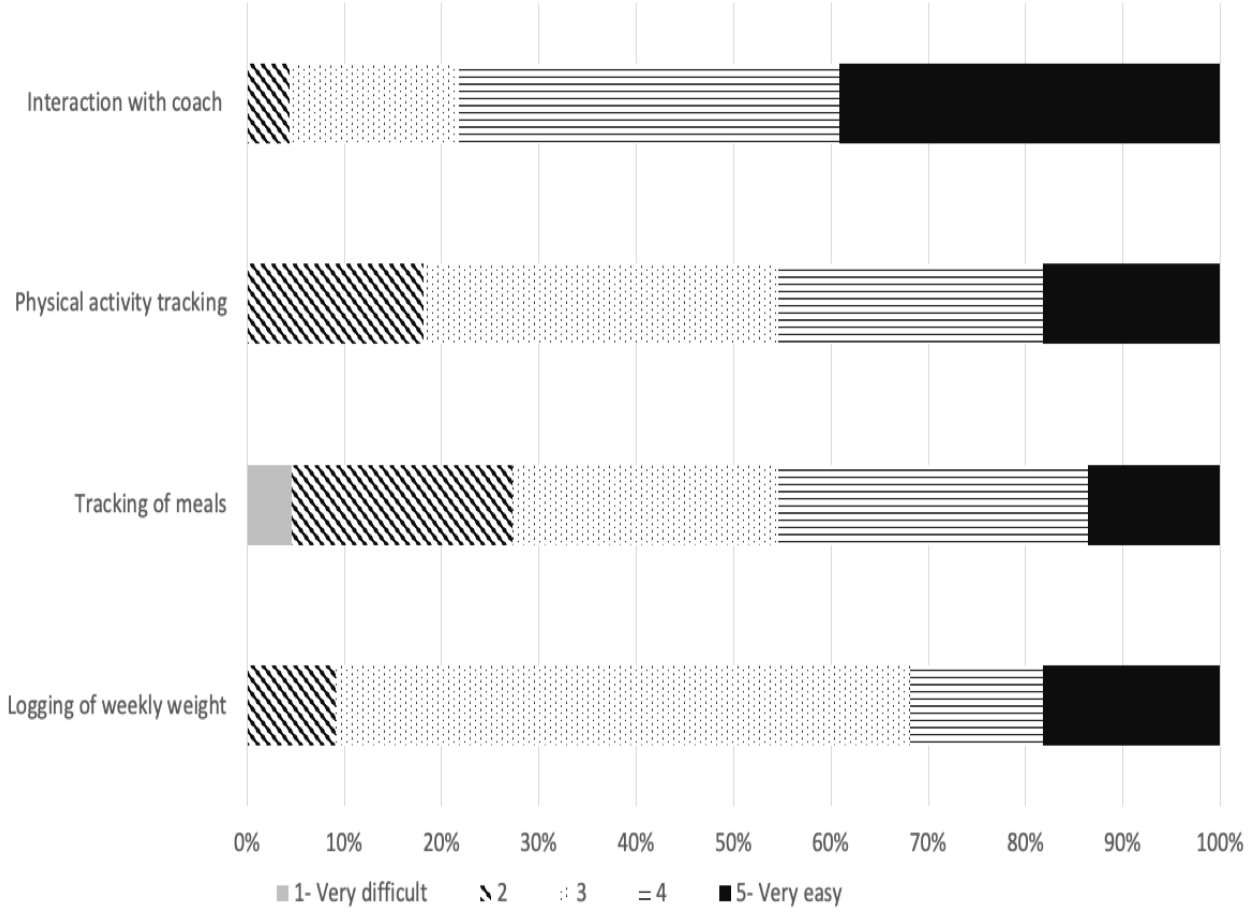
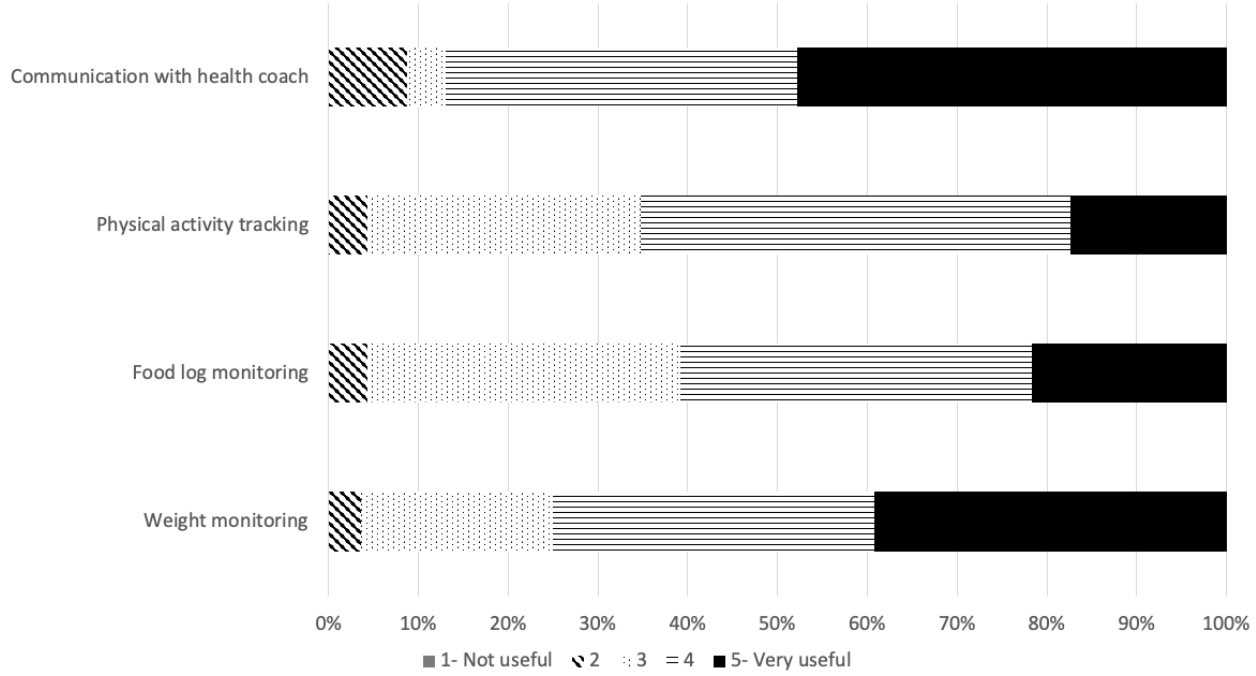


Figure 5. Usefulness of the various component of Kurbo where 1=not useful and 5=very useful.



Effectiveness (3-Month Outcome) and Maintenance (6-Month Outcome)

Anthropometric and Blood Pressure Outcomes

There were no significant changes in BMI z-score (primary outcome) at either 3 or 6 months. However, there was a significant reduction in body fat percentage at both 3 months (−1.3%; 95% CI −2.5% to −0.2%; *P*=.03) and 6 months (−2%; 95% CI −3.6% to −0.4%; *P*=.02; Table 2). After adjusting for multidisciplinary weight management visits during engagement with Kurbo *P*, there was no significant difference in BMI z-scores at 3 months (F_{27}=0.406; *P*=.53). No significant differences were detected at 6 months for blood pressure, waist circumference, or waist-to-height ratio (Table 2).

Table 2. Changes in adolescents' anthropometric and blood pressure (N=40).

Variable	Baseline to 3 months (n=21)		Baseline to 6 months (n=20)	
	Value, mean (SD; 95% CI)	*P* value	Value, mean (SD; 95% CI)	*P* value
Body mass (kg)	2.7 (4.74; 0.5 to 4.8)	.02	3.59 (4.55; 1.45 to 5.71)	.002
BMI z-score	0.045 (0.15; −0.024 to 0.114)	.19	0.035 (0.14; −0.028 to 0.098)	.27
Waist circumference (cm)	1.1 (6.74; −1.9 to 4.2)	.45	0.3 (6.21; −2.6 to 3.2)	.84
Waist-to-height ratio	−0.003 (0.040; −0.20 to 0.015)	.75	0.004 (0.038; −0.014 to 0.022)	.67
Body fat (%)	−1.31 (2.54; −2.47 to −0.15)	.03	−2.0 (3.46; −3.6 to −0.38)	.02
Systolic BP[a] (mm Hg)	−5.5 (9.21; −9.8 to −1.2)	.02	−2.1 (9.62; −6.8 to 2.5)	.35
Diastolic BP (mm Hg)	−4.2 (8.50; −8.1 to −0.2)	.04	−3.2 (8.47; −7.3 to 0.85)	.11

[a]BP: blood pressure.

Eating and Physical Activity Behaviors

The 3-day food diary revealed significant reductions in caloric intake at 3 months (mean −300, SD 456; 95% CI −576 to −24; *P*=.04) and 6 months (mean −332, SD 517; 95% CI 598 to −66; *P*=.02; Table 3) [5]. Moreover, the time spent in MVPA (minutes) at 6 months increased (mean 5.3, SD 4.8; 95% CI 0.88-9.75; *P*=.03; Table 3).

Table 3. Changes in adolescents' health behavior and psychosocial parameters (N=40).

Variable	Baseline to 3 months (n=21)		Baseline to 6 months (n=20)	
	Value, mean (SD; 95% CI)	*P* value	Value, mean (SD; 95% CI)	*P* value
Total (kcal/day)	−300 (457; −576 to −24)	.04	−332 (518; −598 to −66)	.02
Servings of vegetables per day	−0.17 (0.50; −0.5 to 0.1)	.24	−0.0 (0.79; −0.4 to 0.4)	.99
Average moderate-to-vigorous physical activity per day (minutes)	1.47 (10.03; −4.3 to 7.2)	.59	5.3 (4.78; 0.88 to 9.75)	.03
Adolescents Pediatric Quality of Life Inventory				
Total	2.4 (13.11; −3.1 to 7.9)	.38	1.4 (11.67; −4.2 to 7.0)	.61
Physical	4.0 (13.56; −1.7 to 9.8)	.16	1.0 (11.60; −4.6 to 6.6)	.72
Emotional	6.5 (23.34; −3.4 to 16.3)	.19	9.7 (20.58; −0.2 to 19.7)	.05
School	8.3 (18.28; 0.6 to 16.1)	.04	6.9 (13.52; 0.2 to 13.7)	.04
Psychosocial	6.3 (14.52; 0.2 to 12.5)	.04	6.5 (12.98; 0.2 to 12.7)	.04
Eating Pattern Inventory for Children				
Dietary restraint	0.033 (0.42; −0.15 to 0.21)	.72	0.00 (0.48; 0.23 to −0.23)	.99
External eating	−0.087 (0.82; −0.44 to 0.27)	.62	−0.19 (0.65; −0.50 to 0.12)	.22
Parental pressure to eat	0.116 (0.54; −0.12 to 0.35)	.31	0.018 (0.66; −0.30 to 0.33)	.91
Emotional eating	0.00 (0.79; −0.34 to 0.34)	.99	0.171 (0.67; −0.15 to 0.49)	.28

Psychosocial Outcomes

At 3 months, adolescents' self-reported quality of life improved in the school (mean 8.3; 95% CI 0.6-16.1; *P*=.04) and psychosocial (mean 6.3; 95% CI 0.19-12.5; *P*=.04) domains (Table 2). These improvements persisted for 6 months. There were no significant changes in eating patterns, including dietary restraint, external eating, parental pressure to eat, and emotional eating subscales.

Discussion

Principal Findings

This pilot study is one of the few studies to evaluate the implementation of a multicomponent mobile app as an early intervention before enrollment in an adolescent WMC. The Kurbo pilot was successful in reaching a low-income and racially diverse population. Although there was no significant reduction in BMI z-scores, there were significant improvements in fat percentage, total caloric intake, and quality of life, suggesting potential benefits of enrolment and the need for a more formal randomized trial.

Obtaining a reach of 58% is comparable with that of other pediatric obesity studies [27]. Moreover, the high percentage of minority, lower education, and low-income enrollees suggests that Kurbo is more likely to reach disadvantaged populations than traditional in-person programs. Given the high rates of obesity among children of lower socioeconomic status, this is a significant advantage of Kurbo.

Kurbo participants completed a median of 7 (IQR 2-10) coaching sessions. This level of engagement is considered very low (<10 hours of intervention time) based on the US Preventive Special Task Force criteria [28]. This may account for the lack of BMI changes among our participants. However, this lack of engagement is not unique to Kurbo, as other programs, including clinic-based programs, have shown similar levels of engagement [29-31].

Our results show that the maximal period of engagement with Kurbo occurred in the first 7 weeks, which corresponded to the period between the initial WMC referral and the first WMC clinic visit. This suggests that Kurbo may be helpful in engaging participants as an early intervention before the first WMC visit. Early engagement may account for the increase in WMC attendance at 6 months (20/40, 50% attendance) compared with our historical rate of 42.1% (51/121). The use of a mobile app as an early intervention also provided mutual benefits to both the health care provider team and Kurbo health coaches. The administrator platform allowed the multidisciplinary team to gain a better understanding of patient progress in health behaviors and weight before presenting to the clinic. This allowed for more targeted discussions about barriers that adolescents faced in the management of obesity and more efficient care. Kurbo health coaches were able to highlight any concerns that they faced during the health coaching sessions of the health care team.

A challenge with Kurbo as an early intervention was the high attrition rate. Although the study's dropout rate of 50% (20/40) is less than the in-clinic rate of 57.8% (70/121), it suggests that strategies need to be crafted to reduce attrition if Kurbo is to be successfully used as an early-stage intervention. Further research on the reasons for attrition is recommended.

Despite the lack of changes in BMI z-scores, a significant improvement in quality of life is an important finding. Quality of life among adolescents with overweight is lower than that among normal-weight peers [25] and youth with other chronic conditions [32]. Improvement in quality of life has been reported

with participation in other obesity treatment programs, even in children who do not achieve significant weight loss [33]. Given the prevalence of weight-based victimization experienced by adolescents with obesity in school [34], the improvement in the school dimension of quality of life is promising. Improvement in quality of life is critical, as it has been associated with improved long-term health indicators, with lower use of health care resources and greater long-term weight reduction [35]. To the best of our knowledge, improvement in self-reported quality of life has not previously been reported with the use of a mobile app to address adolescent obesity, which should be validated in future research.

Reassuringly, our study found no measured increase in disordered eating behaviors, as measured by the EPI-C. On the basis of these results, the integrated model of a mobile app with multidisciplinary adolescent obesity management is unlikely to increase disordered eating behaviors despite concerns for the development of disordered eating habits with the use of mobile apps [36].

Limitations

This study had several limitations. The study had a small sample size and a high attrition rate. Second, as this was a feasibility study, the study did not include a control group.

Conclusions

In this pilot study, the use of the Kurbo mobile app as an early intervention before a multidisciplinary clinical care for adolescent obesity treatment is feasible in a low-income and ethnically diverse Asian population. Although there was no significant change in the BMI z-score, Kurbo showed promise in improving quality of life and reducing body fat percentage and total caloric intake. Given the promising outcomes in several dimensions, further research using more rigorous trial designs should be conducted to evaluate the effects of Kurbo as part of an early, stepped care intervention for adolescents with obesity.

Acknowledgments

The authors thank Kurbo for providing the data for this study. Kurbo was not involved in the design of the study, the analysis and interpretation of the data, or the preparation and submission of this manuscript. The authors wish to thank the patients and their families for their participation in the study and members of the pediatric weight management team for their contribution to the study. The study was funded by the Pediatrics Academic Clinical Programme/Tan Cheng Lim Fund grant PAEDACP-TCL/2017/CLIN/006.

Conflicts of Interest

EAF is on the scientific advisory board for WW Int (formerly Weight Watchers), the company that now owns Kurbo.

References

1. Singh AS, Mulder C, Twisk JW, van MW, Chinapaw MJ. Tracking of childhood overweight into adulthood: a systematic review of the literature. Obes Rev 2008 Sep;9(5):474-488. [doi: 10.1111/j.1467-789X.2008.00475.x] [Medline: 18331423]
2. Barlow SE, Expert C. Expert committee recommendations regarding the prevention, assessment, and treatment of child and adolescent overweight and obesity: summary report. Pediatrics 2007 Dec;120 Suppl 4:S164-S192. [doi: 10.1542/peds.2007-2329C] [Medline: 18055651]
3. Teo TS. Demographic and motivation variables associated with internet usage activities. Internet Res 2001 May;11(2):125-137. [doi: 10.1108/10662240110695089]
4. Langlet B, Maramis C, Diou C, Maglaveras N, Fagerberg P, Heimeier R, et al. Formative evaluation of a smartphone app for monitoring daily meal distribution and food selection in adolescents: acceptability and usability study. JMIR Mhealth Uhealth 2020 Jul 21;8(7):e14778 [FREE Full text] [doi: 10.2196/14778] [Medline: 32706684]
5. Nikolaou CK, Tay Z, Leu J, Rebello SA, Te Morenga L, Van Dam RM, et al. Young people's attitudes and motivations toward social media and mobile apps for weight control: mixed methods study. JMIR Mhealth Uhealth 2019 Oct 10;7(10):e11205 [FREE Full text] [doi: 10.2196/11205] [Medline: 31603431]
6. Almandoz JP, Xie L, Schellinger JN, Mathew MS, Gazda C, Ofori A, et al. Impact of COVID-19 stay-at-home orders on weight-related behaviours among patients with obesity. Clin Obes 2020 Oct;10(5):e12386 [FREE Full text] [doi: 10.1111/cob.12386] [Medline: 32515555]
7. Elnaggar RK, Alqahtani BA, Mahmoud WS, Elfakharany MS. Physical activity in adolescents during the social distancing policies of the COVID-19 pandemic. Asia Pac J Public Health 2020 Nov;32(8):491-494. [doi: 10.1177/1010539520963564] [Medline: 33016086]
8. Pietrobelli A, Pecoraro L, Ferruzzi A, Heo M, Faith M, Zoller T, et al. Effects of COVID-19 lockdown on lifestyle behaviors in children with obesity living in Verona, Italy: a longitudinal study. Obesity (Silver Spring) 2020 Aug;28(8):1382-1385 [FREE Full text] [doi: 10.1002/oby.22861] [Medline: 32352652]
9. Kurbo homepage. Kurbo. URL: https://kurbo.com [accessed 2021-09-03]
10. Cueto V, Wang CJ, Sanders LM. Impact of a mobile app-based health coaching and behavior change program on participant engagement and weight status of overweight and obese children: retrospective cohort study. JMIR Mhealth Uhealth 2019 Nov 15;7(11):e14458 [FREE Full text] [doi: 10.2196/14458] [Medline: 31730041]

11. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
12. Nicholas Hong WJ, Huang HL, Rajasegaran K, Oh JY, Kelly S, Saffari SE, et al. Presence of obesity related comorbidities associated with lower attrition rate in pediatric weight management program. J Child Obes 2017 May 29;2(2):2-10. [doi: 10.21767/2572-5394.100031]
13. Kuczmarski RJ, Ogden CL, Guo SS, Grummer-Strawn LM, Flegal KM, Mei Z, et al. 2000 CDC Growth Charts for the United States: methods and development. Vital Health Stat 11 2002 May(246):1-190 [FREE Full text] [Medline: 12043359]
14. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
15. Epstein LH, Gordy CC, Raynor HA, Beddome M, Kilanowski CK, Paluch R. Increasing fruit and vegetable intake and decreasing fat and sugar intake in families at risk for childhood obesity. Obes Res 2001 Mar;9(3):171-178. [doi: 10.1038/oby.2001.18] [Medline: 11323442]
16. Mukherjee S, Leong HF, Wong XX. Waist circumference percentiles for Singaporean children and adolescents aged 6-17 years. Obes Res Clin Pract 2016 Sep;10 Suppl 1:S17-S25. [doi: 10.1016/j.orcp.2016.04.004] [Medline: 27180165]
17. Talma H, Chinapaw MJ, Bakker B, HiraSing RA, Terwee CB, Altenburg TM. Bioelectrical impedance analysis to estimate body composition in children and adolescents: a systematic review and evidence appraisal of validity, responsiveness, reliability and measurement error. Obes Rev 2013 Nov;14(11):895-905. [doi: 10.1111/obr.12061] [Medline: 23848977]
18. Chew CS, Oh JY, Rajasegaran K, Saffari SE, Lim CM, Lim SC, et al. Evaluation of a group family-based intervention programme for adolescent obesity: the LITE randomised controlled pilot trial. Singapore Med J 2019 Oct 08;62(1):39-47 [FREE Full text] [doi: 10.11622/smedj.2019122] [Medline: 31598732]
19. Ling Ang K, Foo S. An exploratory study of eating patterns of Singapore children and teenagers. Health Educ 2002 Oct 1;102(5):239-248. [doi: 10.1108/09654280210444119]
20. Trost SG, Loprinzi PD, Moore R, Pfeiffer KA. Comparison of accelerometer cut points for predicting activity intensity in youth. Med Sci Sports Exerc 2011 Jul;43(7):1360-1368. [doi: 10.1249/MSS.0b013e318206476e] [Medline: 21131873]
21. Puyau MR, Adolph AL, Vohra FA, Zakeri I, Butte NF. Prediction of activity energy expenditure using accelerometers in children. Med Sci Sports Exerc 2004 Sep;36(9):1625-1631. [Medline: 15354047]
22. Ching Ting JL, Mukherjee S, Yong Hwa MC. Physical activity and sedentary behavior patterns of Singaporean adolescents. J Phys Act Health 2015 Sep;12(9):1213-1220. [doi: 10.1123/jpah.2014-0207] [Medline: 25393419]
23. Schacht M, Richter-Appelt H, Schulte-Markwort M, Hebebrand J, Schimmelmann BG. Eating pattern inventory for children: a new self-rating questionnaire for preadolescents. J Clin Psychol 2006 Oct;62(10):1259-1273. [doi: 10.1002/jclp.20300] [Medline: 16897691]
24. Gow ML, Baur LA, Ho M, Chisholm K, Noakes M, Cowell CT, et al. Can early weight loss, eating behaviors and socioeconomic factors predict successful weight loss at 12- and 24-months in adolescents with obesity and insulin resistance participating in a randomised controlled trial? Int J Behav Nutr Phys Act 2016 Apr 01;13(1):43 [FREE Full text] [doi: 10.1186/s12966-016-0367-9] [Medline: 27036113]
25. Østbye T, Malhotra R, Wong H, Tan S, Saw S. The effect of body mass on health-related quality of life among Singaporean adolescents: results from the SCORM study. Qual Life Res 2010 Mar;19(2):167-176. [doi: 10.1007/s11136-009-9578-8] [Medline: 20066564]
26. NG PJ. Foreword. Int J Water Resour Dev 2020 Nov 05;36(6):871-873. [doi: 10.1080/07900627.2020.1820692]
27. Lew MS, L'Allemand D, Meli D, Frey P, Maire M, Isenschmid B, et al. Evaluating a childhood obesity program with the Reach, Effectiveness, Adoption, Implementation, Maintenance (RE-AIM) framework. Prev Med Rep 2019 Jan 24;13:321-326 [FREE Full text] [doi: 10.1016/j.pmedr.2019.01.017] [Medline: 30792947]
28. Whitlock EP, O'Connor EA, Williams SB, Beil TL, Lutz KW. Effectiveness of weight management interventions in children: a targeted systematic review for the USPSTF. Pediatrics 2010 Feb;125(2):e396-e418. [doi: 10.1542/peds.2009-1955] [Medline: 20083531]
29. Andrews M, Sawyer C, Frerichs L, Skinner AC, Hoffman J, Gaskin K, et al. Feasibility of a clinic-community partnership to treat childhood obesity. Clin Pediatr (Phila) 2018 Jun;57(7):783-791. [doi: 10.1177/0009922817734359] [Medline: 28969465]
30. Hoffman J, Frerichs L, Story M, Jones J, Gaskin K, Apple A, et al. An integrated clinic-community partnership for child obesity treatment: a randomized pilot trial. Pediatrics 2018 Jan 01;141(1):783-791 [FREE Full text] [doi: 10.1542/peds.2017-1444] [Medline: 29237800]
31. Serra-Paya N, Ensenyat A, Castro-Viñuales I, Real J, Sinfreu-Bergués X, Zapata A, et al. Effectiveness of a multi-component intervention for overweight and obese children (Nereu Program): a randomized controlled trial. PLoS One 2015 Dec 14;10(12):e0144502 [FREE Full text] [doi: 10.1371/journal.pone.0144502] [Medline: 26658988]
32. Ingerski LM, Modi AC, Hood KK, Pai AL, Zeller M, Piazza-Waggoner C, et al. Health-related quality of life across pediatric chronic conditions. J Pediatr 2010 Apr;156(4):639-644 [FREE Full text] [doi: 10.1016/j.jpeds.2009.11.008] [Medline: 20117793]

33. Mollerup PM, Nielsen TR, Bøjsøe C, Kloppenborg JT, Baker JL, Holm J. Quality of life improves in children and adolescents during a community-based overweight and obesity treatment. Qual Life Res 2017 Jun;26(6):1597-1608. [doi: 10.1007/s11136-017-1504-x] [Medline: 28213684]
34. Puhl RM, Peterson JL, Luedicke J. Weight-based victimization: bullying experiences of weight loss treatment-seeking youth. Pediatrics 2013 Jan;131(1):e1-e9. [doi: 10.1542/peds.2012-1106] [Medline: 23266918]
35. Griffiths LJ, Parsons TJ, Hill AJ. Self-esteem and quality of life in obese children and adolescents: a systematic review. Int J Pediatr Obes 2010 Aug;5(4):282-304. [doi: 10.3109/17477160903473697] [Medline: 20210677]
36. Weight loss for kids? Thanks to WW, there's an app for that. Time. 2019 Aug. URL: https://time.com/5649964/kurbo-by-ww-app/ [accessed 2021-09-08]

Abbreviations

EPI-C: Eating Pattern Inventory for Children
KKH: KK Women's and Children's Hospital
mHealth: mobile health
MVPA: moderate-to-vigorous physical activity
WMC: weight management clinic

Edited by R Kukafka; submitted 28.12.20; peer-reviewed by J Alvarez Pitti, A Videira-Silva; comments to author 16.03.21; revised version received 07.05.21; accepted 12.08.21; published 28.09.21.

Please cite as:
Chew CSE, Davis C, Lim JKE, Lim CMM, Tan YZH, Oh JY, Rajasegaran K, Chia YHM, Finkelstein EA
Use of a Mobile Lifestyle Intervention App as an Early Intervention for Adolescents With Obesity: Single-Cohort Study
J Med Internet Res 2021;23(9):e20520
URL: https://www.jmir.org/2021/9/e20520
doi:10.2196/20520
PMID:34581672

Original Paper

Psychoeducational Messaging to Reduce Alcohol Use for College Students With Type 1 Diabetes: Internet-Delivered Pilot Trial

Lauren E Wisk[1,2,3], PhD; Kara M Magane[2], MS; Eliza B Nelson[2], MS; Rebecca K Tsevat[1], MS, MD; Sharon Levy[3,4], MPH, MD; Elissa R Weitzman[2,3,5], MSc, ScD

[1]Division of General Internal Medicine and Health Services Research, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, United States

[2]Division of Adolescent/Young Adult Medicine, Boston Children's Hospital, Boston, MA, United States

[3]Department of Pediatrics, Harvard Medical School, Boston, MA, United States

[4]Division of Developmental Medicine, Boston Children's Hospital, Boston, MA, United States

[5]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, United States

Corresponding Author:
Lauren E Wisk, PhD
Division of General Internal Medicine and Health Services Research
David Geffen School of Medicine
University of California Los Angeles
1100 Glendon Ave
Suite 850
Los Angeles, CA, 90024
United States
Phone: 1 3102675308
Email: lwisk@mednet.ucla.edu

Abstract

Background: College environments promote high-volume or binge alcohol consumption among youth, which may be especially harmful to those with type 1 diabetes (T1D). Little is known about the acceptability and effectiveness of interventions targeting reduced alcohol use by college students with T1D, and it is unclear whether intervention framing (specifically, the narrator of intervention messages) matters with respect to affecting behavior change. Interventions promoted by peer educators may be highly relatable and socially persuasive, whereas those delivered by clinical providers may be highly credible and motivating.

Objective: The aim of this study is to determine the acceptability and impacts of an alcohol use psychoeducational intervention delivered asynchronously through web-based channels to college students with T1D. The secondary aim is to compare the impacts of two competing versions of the intervention that differed by narrator (peer vs clinician).

Methods: We recruited 138 college students (aged 17-25 years) with T1D through web-based channels and delivered a brief intervention to participants randomly assigned to 1 of 2 versions that differed only with respect to the audiovisually recorded narrator. We assessed the impacts of the exposure to the intervention overall and by group, comparing the levels of alcohol- and diabetes-related knowledge, perceptions, and use among baseline, immediately after the intervention, and 2 weeks after intervention delivery.

Results: Of the 138 enrolled participants, 122 (88.4%) completed all follow-up assessments; the participants were predominantly women (98/122, 80.3%), were White non-Hispanic (102/122, 83.6%), and had consumed alcohol in the past year (101/122, 82.8%). Both arms saw significant postintervention gains in the knowledge of alcohol's impacts on diabetes-related factors, health-protecting attitudes toward drinking, and concerns about drinking. All participants reported significant decreases in binge drinking 2 weeks after the intervention (21.3%; odds ratio 0.48, 95% CI 0.31-0.75) compared with the 2 weeks before the intervention (43/122, 35.2%). Changes in binge drinking after the intervention were affected by changes in concerns about alcohol use and T1D. Those who viewed the provider narrator were significantly more likely to rate their narrator as knowledgeable and trustworthy; there were no other significant differences in intervention effects by the narrator.

Conclusions: The intervention model was highly acceptable and effective at reducing self-reported binge drinking at follow-up, offering the potential for broad dissemination and reach given the web-based format and contactless, on-demand content. Both

intervention narrators increased knowledge, improved health-protecting attitudes, and increased concerns regarding alcohol use. The participants' perceptions of expertise and credibility differed by narrator.

Trial Registration: ClinicalTrials.gov NCT02883829; https://clinicaltrials.gov/ct2/show/NCT02883829

International Registered Report Identifier (IRRID): RR2-10.1177/1932296819839503

(J Med Internet Res 2021;23(9):e26418) doi:10.2196/26418

KEYWORDS

adolescent; young adult; diabetes mellitus; type 1; binge drinking; alcohol drinking; self care; risk-taking; universities; students; attitude; mobile phone

Introduction

Background

Effective management of chronic conditions such as type 1 diabetes (T1D) is affected by individual-, family-, community-, and health care system–level features [1]. For adolescents and young adults with T1D who may struggle to meet glycemic targets [2], disease self-management (eg, blood glucose monitoring) and risk behaviors (eg, alcohol consumption and poor diet) are some of the many factors that influence the achievement of glycemic targets [3,4]. The transition to college may be a particularly vulnerable period, as self-management processes are disrupted, risk behaviors rise, and access to support and supervision from family and health care providers diminishes [5].

Clinical studies in the United States have found that 1 in 3 youths with T1D may drink alcohol by the time they reach high school [6]. Postsecondary education (eg, college or university) students with T1D may be especially vulnerable to initiating or intensifying alcohol use, as many reside in settings that are ill-prepared to identify and support their health needs [7]. *Alcogenic* college environments [8-10] create social pressures that foster alcohol use [11,12]. Despite the risks associated with alcohol consumption and T1D [13,14], college-aged youths with T1D have poor alcohol-related health literacy [15,16] and engage in risky alcohol use [8,17]. There is a dearth of evidence-based interventions to minimize alcohol-related risks in youths with T1D in general [3,16,18] and fewer targeting those in college [19].

Given the high potential for alcohol-related health harm, it is critically important to identify how best to engage college students with T1D around harm reduction [20]. Consistent with the students' developmental status, successful interventions may require a patient-centered focus and peer or provider directives, reinforcing the value of health-protecting behavior [21-24]. A digital approach to delivering specialized information may be especially advantageous in situations where college students are distanced from their specialty care or where college health services may be underresourced for the provision of chronic disease–specific guidance [25]. The digitally oriented nature of this age group suggests that the students will be highly receptive to such delivery [26]; however, even digital materials may benefit from delivery by a compelling spokesperson who can appropriately frame the educational materials. Communication sciences and behavioral economics posit different ways in which the presentation of information can affect decision-making; thus, understanding the acceptability and impact of different narrators for delivering harm reduction messages may optimize their effects. Peer delivery could be especially meaningful if college students with T1D personally identify with the narrator and feel supported regarding the distress arising from contravening social norms that encourage drinking. Conversely, provider delivery may be highly salient—clinician expertise, gravitas, and the grounding of guidance in diabetes science may be impactful for shaping health beliefs and behaviors.

Objective

The aims of this study are to understand the acceptability of a digital health intervention for addressing alcohol use risk among college students with T1D and to test the relative salience of two types of narrators for its delivery and framing. Consistent with a stepped approach for designing behavioral interventions [27,28], we undertook a pilot trial to evaluate the acceptability and impacts of the intervention overall and measure the relative salience of competing versions of the intervention, with one version delivered by a peer educator and another delivered by a specialty care provider (ie, a medicine-pediatric endocrinologist). The intervention coupled diabetes-related medical science and social-emotional content, building on formative work indicating that both elements are considered high priorities for substance use–related interventions targeting adolescents and young adults with chronic conditions [22]. Given prior evidence about the effectiveness of peer support for diabetes self-management [29-34], we hypothesized that exposure to peer delivery of this information would show higher acceptability and impacts on knowledge, attitudes, and behaviors than provider delivery.

Methods

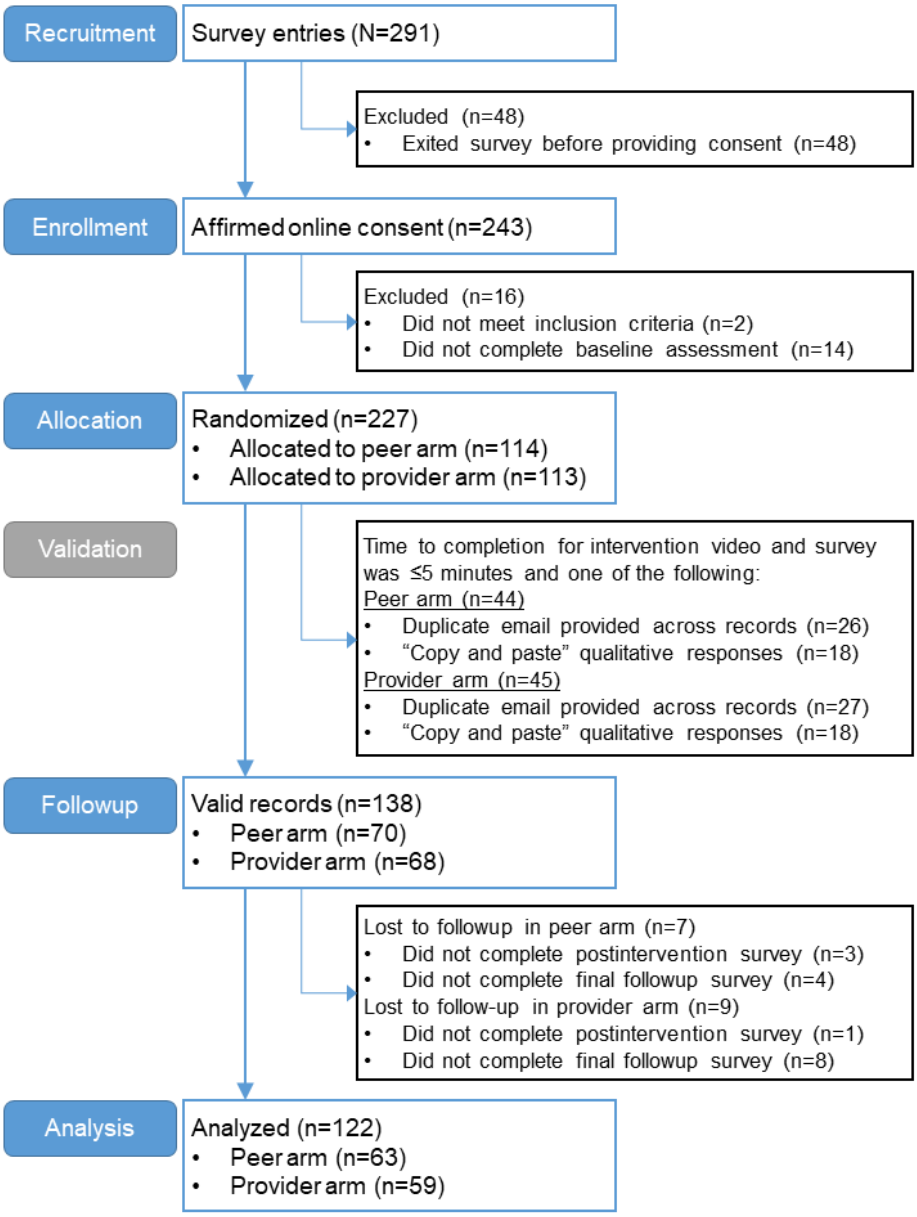
Overview

We implemented a pilot trial to test two competing versions of a novel psychoeducational intervention targeting alcohol-related knowledge, attitudes, and behaviors of college students with T1D from the United States and Canada. The approach to engaging participants has been previously reported [35]. For the pilot, we randomized the narrators delivering and framing the educational materials. Both structured and open-ended survey data were collected. This study was approved by the institutional review board.

Recruitment and Consent

The participants were recruited from two diabetes advocacy groups through social media, direct email newsletters, and a website banner. The recruitment messages were linked to a website with additional study details and safety resources. The website, in turn, transferred interested individuals to a REDCap (Research Electronic Data Capture) survey hosted on secure servers; the website and survey were accessible on a smartphone, tablet, or computer. After entering the survey, the individuals completed the web-based informed consent (or assented with a waiver of parental consent for those aged 17 years). Additional details of recruitment procedures and study implementation, including an evaluation of sampling procedures and generalizability, have been previously published [35] and are summarized in the CONSORT (Consolidated Standards of Reporting Trials) diagram (Figure 1).

Figure 1. Trial CONSORT (Consolidated Standards of Reporting Trials) diagram showing participant flow through each stage of the randomized trial from recruitment through analyses.



After providing consent, the respondents answered screening questions to ascertain eligibility (aged 17-25 years, received a diagnosis of T1D, and currently attending, enrolled at a college, or university), and those who met the inclusion criteria (hereafter *participants*) were directed to complete the remainder of the baseline survey.

Intervention and Assessments

Upon completion of the baseline survey, the participants were automatically randomized to receive one of two brief psychoeducational videos narrated by either an endocrinologist (provider) or a college student with T1D (peer educator). A REDCap module executed stratified randomization based on sex, year in college, and alcohol use in the past year. The intervention video included educational content about the effects of alcohol on physical and psychological functioning in persons

with T1D and addressed the social aspects of alcohol use through visually rich, annotated graphics that described how and why youths with T1D choose to limit alcohol use in social settings; overall, the video had a harm reduction rather than abstinence focus that did not specifically call for abstinence or suggest safe limits in use.

To develop the psychoeducational content for the intervention, we conducted qualitative interviews using a semistructured interview guide grounded in social cognitive theories of risk-taking among adolescents with chronic conditions [22]. The topics covered daily experiences of the patient's medical condition; information about personal, social, situational, and clinical factors affecting decision-making regarding alcohol use; and preferences for clinical communication about substance use. The results of the qualitative interviews, in combination with prior epidemiological investigations to ascertain salient concerns, motivations, and consequences of alcohol use [6,9,22,36-38], informed the development of an intervention containing patient-centered, accessible information regarding disease-specific concerns and goals while also addressing the complexity of navigating chronic disease management and alcohol use in the setting of adolescent-typical peer pressures and impulses. Clinically accurate content to address disease-specific concerns related to alcohol use within the context of T1D (eg, the effects of alcohol on glucagon efficacy and identification of hypoglycemic symptoms) was developed with input from pediatric endocrinologists, chronic disease epidemiologists, and a developmental behavioral pediatrician specializing in addiction medicine and treating adolescent substance use. Experiential or *slice of life* observations about alcohol use and living with T1D were also incorporated into the intervention using illustrative quotes taken from formative qualitative research [22], with content designed to address the main social and psychological concerns of the target group. Drawings created by an artist embedded in the study team aided in the explanation of both clinical and social-behavioral concepts and were included in the intervention.

Before testing, the intervention underwent formative evaluation through iterative rounds of piloting (for accessibility and acceptability) with patients representing the target group (ie, adolescents and young adults with T1D presenting for routine clinical care to an outpatient hospital clinic). Feedback on the intervention was collected in a structured session with trained members of the research team. The formative evaluation also included a review by pediatric endocrinologists to ensure the accuracy of clinical and health-related content before moving to video production and with stakeholders representing the advocacy organizations through which the study samples were to be recruited. The visual intervention content was then overlaid with the corresponding audio explanation, narrated by either a peer or provider narrator. The narration scripts were identical for both the peer and provider versions of the intervention, with the exception of the initial narrator introductions containing narrator-specific identification (ie, the narrator's name and role or credentials). The peer and provider narrators were purposefully selected to be of the same gender and same race, and they were similarly styled to reduce the influence on audience perceptions related to factors other than the narrator's role. Other than the narrator, all video content was identical. The final versions of the intervention (approximately 7.5 minutes in duration) were uploaded to a private channel on a video hosting site and embedded within the REDCap survey so that the participants could automatically view the video without having to exit the survey.

The participants completed a brief survey immediately after viewing the video, and 2 weeks after completion of the *initial* session (baseline survey, intervention video, and immediate postintervention survey), they were sent a follow-up survey through email (automated through REDCap, with up to two additional reminders). The participants received a US $20 gift card for completion of each of the two sessions (initial and follow-up). Of the 138 valid responses at baseline, 134 (97.1%) completed the immediate postintervention survey, and 122 (88.4%) subsequently completed the follow-up assessment. Compared with the participants with complete follow-up, those lost to follow-up were older at diagnosis (13.3 years vs 10.6 years; P=.04; data not shown); no other significant differences were observed.

Covariates

The participants provided sociodemographic data (eg, age, sex, race or ethnicity, parental education, health insurance, enrollment status, school year, living arrangement, and region of college or university attended) and clinical information (eg, age during diagnosis or disease duration, last glycated hemoglobin [HbA_{1c}] value, current insulin pump use, current continuous glucose monitoring [CGM] device use, average blood glucose tests per day, average HbA_{1c} tests per year, and self-rated health) at baseline.

Self-reported Alcohol Use (Primary Outcome)

The participants self-reported alcohol use behaviors at baseline and follow-up assessments, including validated measures of binge use during the past 2 weeks (defined as having 3, 4, 5, or more drinks containing alcohol on one occasion, depending on age and sex threshold [39]). The participants identified any recent events or breaks that were atypical for their schedule in response to novel questions. The events were grouped into those potentially associated with a greater likelihood of drinking (eg, spring break spent with friends) or those with a reduced likelihood of drinking (eg, midterms or final examinations).

Knowledge, Attitudes, and Concerns Toward Alcohol Use (Secondary Outcomes)

The secondary outcomes relied on novel test batteries that were developed for a series of related studies and informed by clinical reviews and epidemiologic insights [6,22,40,41]. During all three assessments, the participants were asked about their knowledge of the potential impacts of alcohol use on diabetes (10 items with responses of "true," "false," or "don't know"), attitudes toward drinking with diabetes (six items rated from "strongly disagree" [score=1] to "strongly agree" [score=10], with two items only asked among drinkers), and concerns about the potential impact of alcohol on diabetes (seven items rated from "not concerned" [score=0] to "very concerned" [score=6]). To summarize knowledge, the number of correct items was determined. To summarize attitudes and concerns, the mean

score across all items was determined for each; higher scores reflected health-protecting attitudes toward drinking or greater concern about the impact of alcohol. Although the measures were novel and not previously validated, they demonstrated reasonable internal consistency (Cronbach α=.9079 for concerns and .6462 for attitudes).

Intentions and Impressions (Secondary Outcomes)

At postintervention and follow-up assessments, the participants were asked about their impressions of the video (13 items rated from "strongly disagree" [score=1] to "strongly agree" [score=10]) and their intentions (10 items rated from "definitely will not" [score=1] to "definitely will" [score=8], with two items only asked among drinkers). To summarize impressions and intentions, the mean score across all items was determined for each; higher scores reflected more favorable impressions of the intervention (ie, acceptability) or greater intention to engage in lower-risk behaviors. The measures demonstrated reasonable internal consistency (Cronbach α=.8581 for impressions and .8294 for intentions). For aligning with intentions, the participants were asked at the final follow-up to specify actual behaviors, including (1) if they had talked in the past 2 weeks with friends, health care providers, or someone else about how alcohol affects diabetes and (2) the frequency with which they tested their blood sugar before, during, or after drinking during the past 2 weeks.

The participants were also prompted with open-response questions on their diabetes care or alcohol use, first impressions of the video (after the intervention), and final thoughts about the video (final follow-up). The participants who endorsed learning something new or having unanswered questions were prompted for further clarification. These open-ended questions were intended to complement the quantitative impression measures by providing a more subjective and nuanced perspective on intervention acceptability.

Analytic Approach

The analyses followed a mixed methods protocol, investigating structured survey data and thematically coding optional comment data. Quantitative analyses were performed using SAS version 9.4 (SAS Institute Inc). Statistical significance was set at $P<.05$. All individuals who were randomized into the study and completed the follow-up were included in the analyses (N=122). Differences in characteristics between the arms were assessed using appropriate bivariate tests (Table 1). Differences in outcomes between the arms at baseline and follow-up were evaluated using bivariate tests (Table 2). Multivariate mixed effects models estimated the intervention effects on the outcomes overall (main effects models) and by arm (interaction models; Table 3); adjusted predictive margins were output from these models to visualize the intervention effects (Figure 2). Adjusted regression analysis models controlled for age at survey, age at diagnosis, sex, race or ethnicity, parent education, last HbA_{1c} value, and CGM device use, while accounting for repeated measures; the models for alcohol use outcomes were additionally adjusted for atypical events affecting the likelihood of drinking. We selected measures for covariate adjustment based on their known association with alcohol use behaviors (eg, age and sex), known association with diabetes knowledge (eg, HbA_{1c} value), or imbalance across treatment arm (eg, parental education). We additionally evaluated the secondary outcomes (change in knowledge, attitudes, and concerns from baseline to after the intervention) as predictors of the overall change in binge drinking (at final follow-up vs baseline). The participants' responses to the open-response questions were analyzed using an iterative process, including open coding, to identify emergent themes [22] and refinement of the coding scheme over several rounds of joint review with 3 coders (RKT, LEW, and ERW). Disagreements were resolved through discussion and consensus. Themes that reflected consistent commonality across the responses or important distinctive motifs were specified; illustrative quotes for each theme were selected and summarized as context to understand the impacts of the intervention (Multimedia Appendix 1). All 122 participants provided a textual response for their impression of the video, and 91 provided a textual response for the remaining optional prompts; the response rates did not differ by arm (P=.40).

Table 1. Sociodemographic and clinical characteristics by intervention narrator (N=122).

Characteristics	Total	Peer	Provider	*P* value
Total, n (%)	122 (100)	63 (51.6)	59 (48.4)	N/A[a]
Sociodemographic characteristics				
Age at survey (years), mean (SD)	20.46 (1.44)	20.35 (1.48)	20.58 (1.40)	.35
Sex, n (%)				.86
Male	24 (19.7)	12 (19)	12 (20.3)	
Female	98 (80.3)	51 (81)	47 (79.7)	
Race and ethnicity, n (%)				.25
White or non-Hispanic	102 (83.6)	55 (87.3)	47 (79.7)	
Person of color or Hispanic	20 (16.4)	8 (12.7)	12 (20.3)	
Parental education, n (%)				.02
High school degree or less	8 (6.6)	2 (3.2)	6 (10.2)	
Some college, no degree	16 (13.1)	6 (9.5)	10 (16.9)	
Associate's degree	13 (10.7)	7 (11.1)	6 (10.2)	
Bachelor's degree	43 (35.2)	18 (28.6)	25 (42.4)	
Graduate degree	41 (33.6)	30 (47.6)	11 (18.6)	
Unsure	1 (0.8)	0 (0)	1 (1.7)	
Health insurance, n (%)				.78
Parent's plan	108 (88.5)	57 (90.5)	51 (86.4)	
Own plan (including school plan)	7 (5.7)	3 (4.8)	4 (6.8)	
Public or uninsured	7 (5.7)	3 (4.8)	4 (6.8)	
Enrollment status, n (%)				.68
Full time	109 (89.3)	57 (90.5)	52 (88.1)	
Part time or other	13 (10.7)	6 (9.5)	7 (11.9)	
Year in school, n (%)				.94
Freshman	14 (11.5)	8 (12.7)	6 (10.2)	
Sophomore	38 (31.1)	20 (31.7)	18 (30.5)	
Junior	38 (31.1)	19 (30.2)	19 (32.2)	
Senior	20 (16.4)	9 (14.3)	11 (18.6)	
Fifth year or graduate student	12 (9.8)	7 (11.1)	5 (8.5)	
Living arrangement during school, n (%)				.18
At home, with parent, or guardian	23 (18.9)	9 (14.3)	14 (23.7)	
On or off campus housing	99 (81.1)	54 (85.7)	45 (76.3)	
Region of college or university, n (%)				0.28
Northeast	34 (27.9)	17 (27)	17 (28.8)	
Midwest	29 (23.8)	14 (22.2)	15 (25.4)	
South	46 (37.7)	27 (42.9)	19 (32.2)	
West	9 (7.4)	2 (3.2)	7 (11.9)	
Outside the United States	4 (3.3)	3 (4.8)	1 (1.7)	
Clinical characteristics				
Age at diagnosis (years), mean (SD)	10.56 (5.14)	9.89 (5.28)	11.27 (4.93)	.08
Disease duration (years), mean (SD)	9.90 (5.32)	10.46 (5.48)	9.31 (5.12)	.18

Characteristics	Total	Peer	Provider	*P* value
Last HbA_{1c}[b] value, mean (SD)	7.65 (1.20)	7.47 (1.10)	7.84 (1.27)	.06
Insulin pump use, n (%)				.69
Not using	21 (17.2)	10 (15.9)	11 (18.6)	
Currently using	101 (82.8)	53 (84.1)	48 (81.4)	
Continuous glucose monitor use, n (%)				.08
Not using	48 (39.3)	20 (31.7)	28 (47.5)	
Currently using	74 (60.7)	43 (68.3)	31 (52.5)	
Average blood glucose tests (times per day), n (%)				.22
0-2	25 (20.5)	16 (25.4)	9 (15.3)	
3-4	45 (36.9)	25 (39.7)	20 (33.9)	
5-6	32 (26.2)	12 (19)	20 (33.9)	
≥7	20 (16.4)	10 (15.9)	10 (16.9)	
Average HbA_{1c} tests per year, n (%)				.06
0-1	12 (9.8)	5 (7.9)	7 (11.9)	
2	31 (25.4)	19 (30.2)	12 (20.3)	
3	36 (29.5)	23 (36.5)	13 (22)	
≥4	43 (35.2)	16 (25.4)	27 (45.8)	
Self-rated health, n (%)				.77
Fair or poor	13 (10.7)	5 (7.9)	8 (13.6)	
Good	56 (45.9)	29 (46)	27 (45.8)	
Very good	46 (37.7)	25 (39.7)	21 (35.6)	
Excellent	7 (5.7)	4 (6.3)	3 (5.1)	
Past-year alcohol use, n (%)				.94
None	21 (17.2)	11 (17.5)	10 (16.9)	
Any	101 (82.8)	52 (82.5)	49 (83.1)	

[a]N/A: not applicable.

[b]HbA_{1c}: glycated hemoglobin.

Table 2. Unadjusted outcomes at baseline and follow-up by narrator (N=122)[a].

Characteristics	Total	Peer	Provider	*P* value
Past 2-week binge use (self-reported), n (%)				
Baseline (any)	43 (35.2)	21 (33.3)	22 (37.3)	.65
Follow-up (any)	26 (21.3)	14 (22.2)	12 (20.3)	.80
Difference at follow-up versus baseline	–13.9	–11.1	–16.9	.48
Knowledge summary[b]				
Baseline, mean (SD)	8.11 (1.58)	8.35 (1.30)	7.86 (1.81)	.20
Postintervention, mean (SD)	8.84 (1.28)	8.97 (1.15)	8.69 (1.41)	.27
Follow-up, mean (SD)	8.78 (1.29)	8.95 (0.92)	8.59 (1.58)	.32
Difference at postintervention versus baseline	+0.72	+0.62	+0.83	.75
Difference at follow-up versus baseline	+0.66	+0.60	+0.73	.97
Attitude summary[c]				
Baseline, mean (SD)	7.66 (1.38)	7.71 (1.27)	7.61 (1.50)	.81
Postintervention, mean (SD)	7.87 (1.25)	7.84 (1.21)	7.91 (1.29)	.63
Follow-up, mean (SD)	7.72 (1.24)	7.68 (1.24)	7.76 (1.24)	.77
Difference at postintervention versus baseline	+0.22	+0.13	+0.31	.20
Difference at follow-up versus baseline	+0.06	–0.03	+0.15	.44
Concern summary[d]				
Baseline, mean (SD)	3.10 (1.62)	2.92 (1.67)	3.29 (1.55)	.23
Postintervention, mean (SD)	3.33 (1.67)	3.09 (1.76)	3.58 (1.55)	.15
Follow-up, mean (SD)	3.13 (1.70)	2.80 (1.72)	3.48 (1.62)	.02
Difference at postintervention versus baseline	+0.23	+0.17	+0.29	.84
Difference at follow-up versus baseline	+0.03	–0.12	+0.19	.14
Intention summary[e]				
Postintervention, mean (SD)	5.13 (1.39)	4.94 (1.43)	5.33 (1.33)	.10
Follow-up, mean (SD)	4.80 (1.18)	4.76 (1.25)	4.84 (1.10)	.67
Difference at follow-up versus postintervention	–0.33	–0.18	–0.49	.08
Impression summary[f]				
Postintervention, mean (SD)	7.28 (1.48)	6.95 (1.67)	7.64 (1.14)	.03
Follow-up, mean (SD)	7.07 (1.53)	6.94 (1.68)	7.22 (1.35)	.38
Difference at follow-up versus postintervention	–0.21	–0.01	–0.42	.07
In past 2 weeks...(at follow-up), n (%)				
Talked with anyone about alcohol and type 1 diabetes	68 (55.7)	39 (61.9)	29 (49.2)	.16
Always tested before, during, or after drinking	54 (44.3)	23 (36.5)	31 (52.5)	.07

[a]*P* values compare the peer versus provider arms.

[b]The knowledge summary indicates the number of correct items out of a total of 10 items.

[c]The attitude summary reflects the mean (SD) score across six items for drinkers and four items for nondrinkers; the response options ranged from "strongly disagree" (score=1) to "strongly agree" (score=10).

[d]The concern summary reflects the mean (SD) score across seven items; the response options ranged from "not concerned" (score=0) to "very concerned" (score=6).

[e]The intention summary reflects the mean (SD) score across 10 items for drinkers and eight items for nondrinkers; the response options ranged from "definitely will not" (score=1) to "definitely will" (score=8).

[f]The impression summary reflects the mean (SD) score across 12 items; the response options ranged from "strongly disagree" (score=1) to "strongly agree" (score=10).

Table 3. Adjusted regression analysis results[a].

Characteristics	Main effects models		Interaction models	
	Estimate (95% CI)	*P* value	Estimate (95% CI)	*P* value
Past 2-week binge[b]				
Peer versus provider	0.72 (0.33 to 1.60)	.42	0.69 (0.29 to 1.67)	.41
Follow-up versus baseline	0.48 (0.31 to 0.75)	<.001	0.45 (0.25 to 0.83)	.01
Peer×follow-up	—[c]	—	1.14 (0.46 to 2.79)	.78
Knowledge[d]				
Peer versus provider	0.25 (–0.21 to 0.72)	.28	0.37 (–0.22 to 0.96)	.22
Postintervention versus baseline	0.72 (0.50 to 0.95)	<.001	0.83 (0.46 to 1.20)	<.001
Follow-up versus baseline	0.66 (0.42 to 0.91)	<.001	0.73 (0.32 to 1.14)	<.001
Peer×postintervention	—	—	–0.21 (–0.67 to 0.24)	.36
Peer×follow-up	—	—	–0.13 (–0.61 to 0.36)	.61
Attitudes[e]				
Peer versus provider	–0.08 (–0.48 to 0.33)	.71	0.13 (–0.37 to 0.63)	.62
Postintervention versus baseline	0.22 (0.06 to 0.37)	.007	0.31 (0.06 to 0.56)	.02
Follow-up versus baseline	0.06 (–0.10 to 0.22)	.47	0.15 (–0.08 to 0.38)	.20
Peer×postintervention	—	—	–0.18 (–0.49 to 0.13)	.26
Peer×follow-up	—	—	–0.18 (–0.50 to 0.13)	.26
Concerns[f]				
Peer versus provider	–0.59 (–1.19 to 0.01)	.05	–0.44 (–1.06 to 0.18)	.16
Postintervention versus baseline	0.23 (0.06 to 0.40)	.009	0.29 (–0.01 to 0.58)	.06
Follow-up versus baseline	0.03 (–0.19 to 0.25)	.81	0.19 (–0.15 to 0.52)	.28
Peer×postintervention	—	—	–0.12 (–0.46 to 0.23)	.51
Peer×follow-up	—	—	–0.31 (–0.75 to 0.13)	.17
Intentions[g]				
Peer versus provider	–0.13 (–0.57 to 0.31)	.56	–0.28 (–0.78 to 0.22)	.27
Follow-up versus postintervention	–0.33 (–0.51 to –0.15)	<.001	–0.49 (–0.78 to –0.20)	.01
Peer×follow-up	—	—	0.30 (–0.06 to 0.67)	.10
Impressions[h]				
Peer versus provider	–0.50 (–1.01 to 0.01)	.05	–0.71 (–1.25 to –0.17)	.01
Follow-up versus postintervention	–0.21 (–0.40 to –0.01)	.04	–0.42 (–0.74 to –0.10)	.01
Peer×follow-up	—	—	0.41 (0.02 to 0.80)	.04

[a]All models were adjusted for age at survey, age at diagnosis, sex, race or ethnicity, parent education, last glycated hemoglobin value, and continuous glucose monitoring device use. Models for binge use additionally adjust for atypical events that affect the likelihood of drinking. The *peertime interaction* coefficient represents the difference in the change over time for the peer arm versus the provider arm (only included in *Interaction* models); this indicates if the effect of the intervention differed by arm. When not statistically significant at *P*<.05, results from the *Main Effects* column are preferred.

[b]Self-reported past 2-week binge alcohol use was modeled with a binomial distribution. Odds ratios are shown instead of β coefficients.

[c]These are models in which this term was not included.

[d]The knowledge summary score indicates the number of correct items out of a total of 10 items, modeled with a normal distribution.

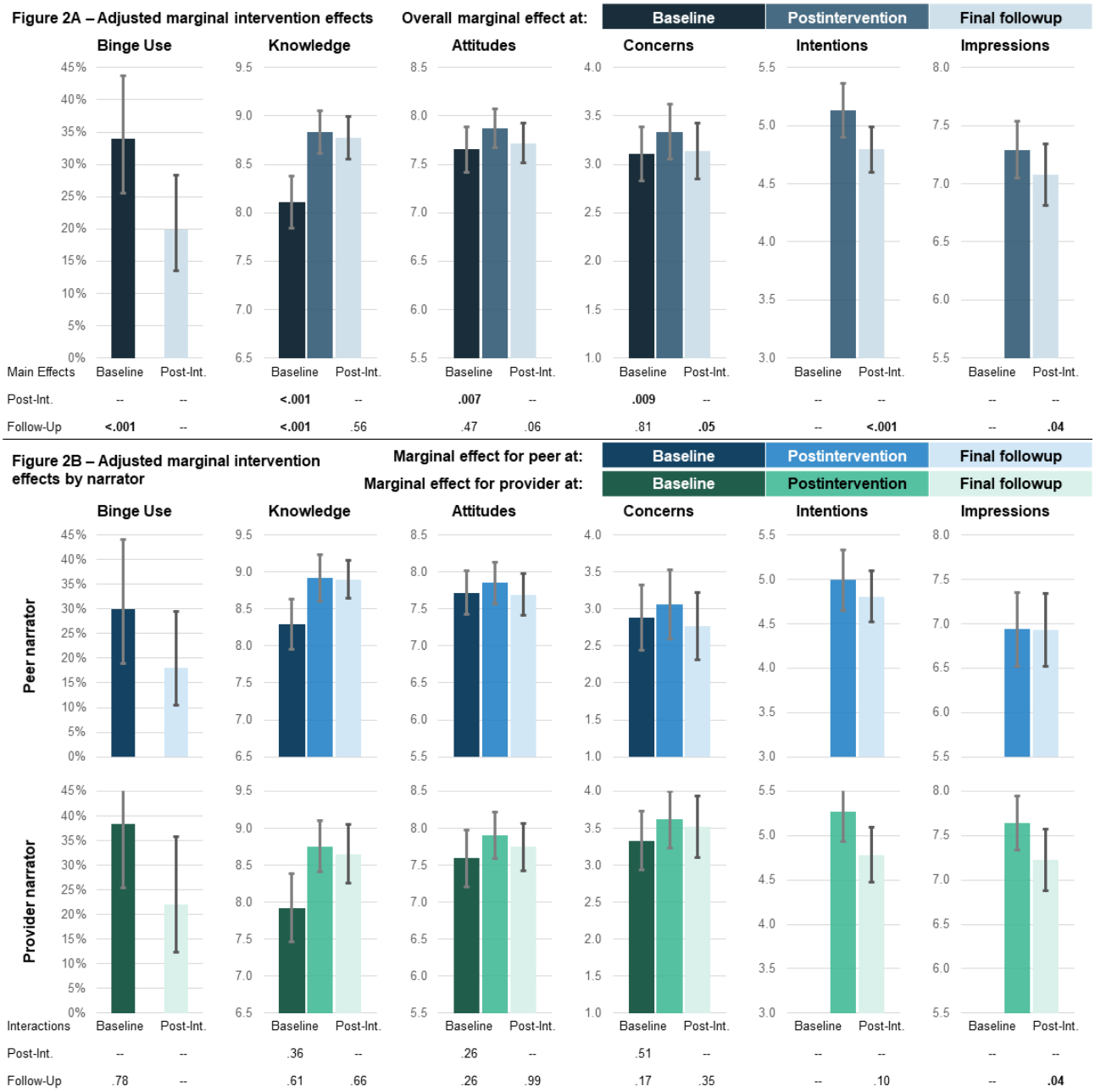
[e]The attitude summary score (average of six items for drinkers and four items for nondrinkers, scored from "strongly disagree" [score=1] to "strongly agree" [score=10]) was modeled with a normal distribution.

[f]The concern summary score (average of seven items, scored from "not concerned" [score=0] to "very concerned" [score=6]) was modeled with a normal distribution.

[g]The intention summary score (average of 10 items for drinkers and eight items for nondrinkers, scored from "definitely will not" [score=1] to "definitely will" [score=8]) was modeled with a normal distribution.

[h]The impression summary score (average of 12 items, scored from "strongly disagree" [score=1] to "strongly agree" [score=10]) was modeled with a normal distribution.

Figure 2. Adjusted marginal intervention effects (y axis) both overall (A, main effects models) and by treatment arm (B, interaction models) across time points (x axis). Adjusted marginal means or probability and 95% CIs were output from multivariate mixed effects models as described in the Methods section. In panel A, the *P* values indicate the statistical significance for the overall intervention effect across time; for example, the intervention reduced the prevalence of binge drinking at final follow-up compared with baseline at *P*<.001. In panel B, the *P* values indicate the statistical significance for the difference in the intervention effect across time by arm (interaction); for example, the reduction in the prevalence of binge drinking at final follow-up compared with baseline was not significantly different by treatment arm at *P*=.78.



Results

Sample Characteristics

The participants were aged 20.46 (SD 1.44) years; 80.3% (98/122) were women, 83.6% (102/122) were White and non-Hispanic, 68.9% (84/122) had a parent with a bachelor's degree or higher, 82.8% (101/122) used insulin pumps, 60.7% (74/122) used CGM devices, their last HbA_{1c} value was 7.65 (SD 1.20), and 82.8% (101/122) self-reported past-year alcohol use (Table 1). The characteristics were largely balanced across the peer (63/122, 51.6%) and provider (59/122, 48.4%) arms, except for the fact that the peer arm participants reported higher parental education (*P*=.02) than their counterparts.

Knowledge, Attitudes, and Concerns

Despite the participants' high baseline knowledge (participants averaged 8.11, SD 1.58 correct items out of 10; Table 2), there was a significant adjusted improvement in correct knowledge items of +0.72 after the intervention (95% CI 0.50-0.95; *P*<.001;

Table 3; Figure 2) and +0.66 at follow-up (95% CI 0.42-0.91; *P*<.001); the improvements after the intervention and at follow-up were not significantly different by arm (*P*=.36 and *P*=.61, respectively). The improvements were driven by an increased understanding of the impact of alcohol on the liver, the distinction between being drunk and low blood glucose, the interaction between alcohol and glucagon, and alcohol-related dehydration (Multimedia Appendix 2).

The participants saw adjusted improvements in health-protecting attitudes of +0.22 after the intervention compared with baseline (95% CI 0.06-0.37; *P*=.007; Table 3; Figure 2); overall improvements after the intervention were similar by arm (*P*=.26). Attitudes did not differ between the follow-up and baseline (+0.06, 95% CI –0.10 to 0.22; *P*=.47). Postintervention changes were driven by attitudes related to telling friends that they cannot drink because of diabetes, and at follow-up, fewer college students agreed that college students with diabetes could drink if they are careful (Multimedia Appendix 2).

The participants reported adjusted increases in concerns about alcohol of +0.23 after the intervention versus baseline (95% CI 0.06-0.40; *P*=.009; Table 3; Figure 2); improvements after the intervention were similar by arm (*P*=.51). Concerns were not different at follow-up versus baseline (+0.03, 95% CI –0.19 to 0.25; *P*=.81). Postintervention changes were driven by increased concern for the effectiveness of glucagon while drinking, whereas increased concerns about the impact of alcohol on HbA_{1c} tests were greater for those in the provider arm than for those in the peer arm (Multimedia Appendix 2).

Intentions and Impressions

Adjusted overall intentions to engage in lower-risk behaviors decreased from after the intervention to final follow-up (–0.33, 95% CI –0.51 to –0.15; *P*<.001; Table 3). Positive impressions of the intervention decreased for those in the provider arm (–0.42, 95% CI –0.74 to –0.10; *P*=.01), whereas those in the peer arm had lower impressions than those in the provider arm (–0.71, 95% CI –1.25 to –0.17; *P*=.01), but their impressions were stable over time (interaction +0.41, 95% CI 0.02-0.80; *P*=.04). The adjusted ratings of narrator knowledge and trust were significantly higher for the provider arm than for the peer arm (Multimedia Appendix 2). After the intervention, the provider arm participants also rated their trust in the information presented and willingness to share this video more highly than the peer arm participants. The provider arm was significantly more likely to decrease their rating of whether they learned something new from after the intervention to follow-up than the peer arm. Intentions to tell their friends about the effect of alcohol on diabetes and to ask their provider about diabetes and alcohol use were stable for the peer arm but more likely to decrease at follow-up for the provider arm.

Alcohol Use Outcomes

Overall, the intervention was associated with a 52% adjusted relative reduction in self-reported binge drinking (odds ratio [OR] 0.48, 95% CI 0.31-0.75; *P*=.001; Table 3; Figure 2) from baseline to follow-up, with no significant difference across narrators (*P*=.78). The change in concerns from baseline to after the intervention was significantly associated with a reduction in binge drinking at follow-up (OR 0.71, 95% CI 0.58-0.88; data not shown), whereas the changes in knowledge and attitudes were not (knowledge OR 0.93, 95% CI 0.75-1.15; attitudes OR 0.90, 95% CI 0.70-1.15).

Qualitative Outcomes

Overall, the qualitative responses suggested that the participants seemed to derive value from the intervention, with four primary themes emerging (Multimedia Appendix 1). First, the youth want disease-specific information about alcohol and have various sources of knowledge acquisition, including through the intervention. Some participants were disappointed that they had not previously been exposed to information about the impact of alcohol on diabetes from other sources (such as their care providers), whereas others expressed appreciation for receiving this new information during the study. However, the information was not novel to all participants, and some described prior acquisition of knowledge through sources such as the internet or peers, as well as their health care teams. Second, the youth noted the influence of personal safety on alcohol use decision-making and the desire for harm reduction messages, as well as new safety practices motivated by the intervention. The participants recognized the prevalence of alcohol use among college students with T1D, expressing a preference for video content encouraging safe and responsible use rather than complete abstinence. Some participants commented upon current safety practices such as drinking with friends and checking their blood sugar while consuming alcohol, whereas others discussed new behaviors inspired by the video, such as recognition of their need to monitor and adjust for their own alcohol intake limits. The third theme encompasses the appeal of intervention content that is not only developmentally appropriate and affirming of personal experiences but also contains an authoritative voice and is professionally produced. Many participants remarked upon their ability to relate to the video because of the relevance of the information and trustworthiness of the speakers, whereas others yearned for more personal stories, easily digestible content, and improved production quality. Finally, the youth addressed factors complicating diabetes and alcohol use management at college, including the need for and role of adequate systems of support. The participants elucidated ongoing challenges faced by youth with T1D, such as social isolation from peers without T1D, inadequate access to specialists on campus, and difficulty balancing their health with other competing priorities during the college years.

Discussion

Principal Findings

This pilot trial demonstrates that a novel, web-based psychoeducational intervention reduces self-reported binge drinking 2 weeks after the intervention among college students with T1D, irrespective of the narrator. Both narrators improved knowledge of the effects of alcohol and temporarily affected attitudes toward and concerns about drinking, although the provider narrator was more highly rated regarding subject knowledge and trustworthiness. Given measurable reductions in binge drinking at follow-up, this brief video intervention

demonstrates promise for mitigating risky alcohol consumption among college students with T1D, and the web-based format is suitable for delivery across numerous settings.

The substantial reduction in binge alcohol use observed for the participants in both arms suggests that both peer and provider narrators can motivate behavioral change. However, the mechanisms by which the narrator affects change may be different. The provider narrator appeared to quantitatively *win out* over the peer narrator with respect to perceived knowledge and trustworthiness, suggesting that delivery of factual content may be enhanced by the aura of clinical authority or that some youths strongly trust and respond to perceived medical experts. Prior qualitative work reported that youths living with chronic conditions conveyed an emphatic preference for the inclusion of information about alcohol use in subspecialty care and reported thinking that their specialty care team know them and their disease best [22]; current qualitative findings echo this and indicate high respect for providers' authority on this topic. The peer narrator appeared to better motivate the intention of the participants to speak with others about the effects of alcohol on their diabetes, suggesting that emotional processing and sharing of experiences may be important for motivating behavioral change and reinforcing knowledge gains. The qualitative findings also supported the participants' desire for narrator and content relatability, as the participants responded positively to hearing their own experiences reflected by their peers. Subtle differences in secondary outcomes by narrator underscore the need for purposeful, well-designed future work to tease out the components that produce cognitive, affective, and behavioral changes, including over a longer time horizon, using insights from pilot work to design and refine a robust intervention suitable for testing in a traditional, controlled trial, consistent with best practice recommendations for behavioral trials [42].

Future work should evaluate delivery on a longer timescale and determine if greater personalization might solidify and sustain the effects to extend the promising results from this pilot. As changes in concerns about alcohol and diabetes (an intermediate outcome predicting a reduction in binge use) were not sustained at the final follow-up, the single-dose intervention appeared to alert the participants to possible dangers, generating a near-term response of increased concern. The lack of sustained change may reflect recalibration of concern, given time to absorb new information (knowledge and experience), growth in perceived self-efficacy for navigating drinking, or retrenchment from the initial state of alarm. A model that delivers multiple doses (boosters) may help to reinforce and sustain the response; hence, future studies should investigate changes over longer periods for a single and possibly multiple-dose intervention. Both narrators in this trial were unknown to the participants; intervention delivery from someone with whom participants have an established relationship, such as a trusted provider [43] or close friend, may also boost intervention effects; however, these designs would limit the scalability of a broadly diffusible digital intervention. Alternatively, the effectiveness of a one-time interaction may be enhanced by providing youths with the opportunity to subsequently discuss the intervention's educational content with a trusted provider or peer.

Although in-person anticipatory conversations between adolescents or young adults and their diabetes care teams about the effects of alcohol on their health and strategies for avoiding risk may be the gold standard for delivering a preventive intervention, this model may not be practical. Providers are increasingly tasked with covering a myriad of topics during short clinical encounters [44], and health guidance is increasingly delivered through web-based channels. Therefore, understanding the acceptability and effects of electronic interventions is vital [45]. Prior work has demonstrated that web-based information and diabetes support can augment clinical care [46] and improve quality of life [47]. This study demonstrates that digitally delivered interventions can effectively engage college students with T1D and provide them with vital health information, a promising model for larger-scale dissemination [25,48]. Future research and implementation studies are needed to evaluate the long-term effectiveness of digital tools and build evidence around best practices for content and delivery [49,50].

Finally, many of the participants noted that they would have liked to receive this intervention earlier (ie, before they began drinking or matriculated to college). As the intervention materials were not specific only to college, the current intervention might be generalizable for use with younger individuals or young adults not attending college, and such extensions could be tested. However, as this intervention was designed to focus on harm reduction, further investigation, including with respect to timing of or age at delivery, is needed to determine if these materials are suitable for the goal of preventing drinking initiation.

Limitations

This study includes several limitations. First, as with all internet-based recruitment, the underlying population (*denominator*) [51] is unclear, making it difficult to characterize sample representativeness and imposing constraints on generalizability. Although a prior methodological evaluation identified that this approach yields a sample that is representative of the sampling frame [35], this sample likely does not represent all college students with T1D (eg, this cohort included predominantly women, and most of the participants were privately insured, had higher pump and CGM use, and also had better self-reported glycemic control than other similarly aged cohorts [52]). This sample may be socioeconomically advantaged, with well-managed diabetes; hence, the findings may be indicative of a *best case scenario*. Second, this study was designed as a pilot trial to gauge the acceptability and relative impacts of the narrator and lacked a nonintervention control group. The observed effects could stem from test-retest bias, carryover effects, or regression to the mean; future controlled evaluations might mitigate these concerns. Third, all outcomes (and diabetes status) were self-reported, and only the alcohol use questions relied on validated items. To our knowledge, there are no pre-existing validated tools to assess knowledge, attitudes, and concerns regarding the intersection of alcohol use and diabetes. The survey items were extensively pretested and have been used in other surveys [6,37,38]; however, they may be subject to measurement errors and other biases stemming from self-reporting. Despite the use of validated

items for alcohol use, self-reported behaviors are subject to social desirability bias, which may result in misclassification. Finally, although the intervention videos were made to be maximally comparable (eg, same-gendered, same-race, and similarly styled narrators), other differences may have existed that differentially affected the response to the intervention, including among certain participant subgroups (eg, men respond differently to female narrators). Relying on an actor or more professional video editing, qualitatively noted by some participants, may diminish the potential differences.

Conclusions

We demonstrate the acceptability and impacts of a psychoeducational intervention for mitigating binge alcohol use among college students with T1D. We further determine that peer and provider narrators have similar short-term impacts as framing devices for this educational content. As this pilot demonstrates the feasibility of delivering this intervention to a hard-to-reach group through web-based channels, similar outreach methods could be used to deliver this content more broadly. In light of these promising findings, future work should further test these materials against the control of *care as usual* and should determine what further modifications are needed to enhance intervention effects.

Acknowledgments

The authors wish to acknowledge the generous funding support for this project and the investigator efforts provided by the Boston Children's Hospital Research Faculty Council Awards Committee Pilot Research Project Funding FP01017994 (coprimary investigators: LEW and ERW), the Agency for Healthcare Research and Quality K12HS022986 (primary investigator: Jon Finkelstein), Conrad N Hilton Foundation Clinical Research 20140273 (coprimary investigators: SL and ERW), and the National Institute of Diabetes and Digestive and Kidney Diseases K01DK116932 (primary investigator: LEW). The funders had no involvement in the design, data collection, interpretation, or reporting of the results. The authors wish to acknowledge Dr Courtney N Sandler, MD, MPH, and Hannah Hast, PT, DPT, for their assistance with the intervention video.

Authors' Contributions

LEW and ERW conceptualized and designed the study. LEW, ERW, and SL designed the data collection instruments and interventions. KMM and EBN collected data; LEW and ERW coordinated and supervised the data collection. LEW carried out all analyses; RKT carried out qualitative analyses; KMM and EBN assisted with data management and interpretation. LEW drafted the initial manuscript, and all authors critically reviewed and revised the manuscript. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Qualitative themes regarding intervention impressions and impact.
[DOCX File , 19 KB - jmir_v23i9e26418_app1.docx]

Multimedia Appendix 2
Adjusted results for specific items.
[DOCX File , 32 KB - jmir_v23i9e26418_app2.docx]

References

1. Modi AC, Pai AL, Hommel KA, Hood KK, Cortina S, Hilliard ME, et al. Pediatric self-management: a framework for research, practice, and policy. Pediatrics 2012 Feb;129(2):473-485 [FREE Full text] [doi: 10.1542/peds.2011-1635] [Medline: 22218838]
2. McCarthy MM, Grey M. Type 1 diabetes self-management from emerging adulthood through older adulthood. Diabetes Care 2018 Aug;41(8):1608-1614. [doi: 10.2337/dc17-2597] [Medline: 29802144]
3. Barnard K, Sinclair JM, Lawton J, Young AJ, Holt RI. Alcohol-associated risks for young adults with Type 1 diabetes: a narrative review. Diabet Med 2012 Apr;29(4):434-440. [doi: 10.1111/j.1464-5491.2012.03579.x] [Medline: 22248115]
4. Jaser SS, Yates H, Dumser S, Whittemore R. Risky business: risk behaviors in adolescents with type 1 diabetes. Diabetes Educ 2011;37(6):756-764 [FREE Full text] [doi: 10.1177/0145721711422610] [Medline: 22002971]
5. Hanna KM, Woodward J. The transition from pediatric to adult diabetes care services. Clin Nurse Spec 2013;27(3):132-145 [FREE Full text] [doi: 10.1097/NUR.0b013e31828c8372] [Medline: 23575170]
6. Weitzman ER, Ziemnik RE, Huang Q, Levy S. Alcohol and marijuana use and treatment nonadherence among medically vulnerable youth. Pediatrics 2015 Sep;136(3):450-457 [FREE Full text] [Medline: 26668849]

7. Lemly DC, Lawlor K, Scherer EA, Kelemen S, Weitzman ER. College health service capacity to support youth with chronic medical conditions. Pediatrics 2014 Nov;134(5):885-891 [FREE Full text] [doi: 10.1542/peds.2014-1304] [Medline: 25349315]
8. Hanna KM, Stupiansky NW, Weaver MT, Slaven JE, Stump TE. Alcohol use trajectories after high school graduation among emerging adults with type 1 diabetes. J Adolesc Health 2014 Aug;55(2):201-208 [FREE Full text] [doi: 10.1016/j.jadohealth.2014.02.002] [Medline: 24661736]
9. Wisk LE, Weitzman ER. Substance use patterns through early adulthood: results for youth with and without chronic conditions. Am J Prev Med 2016 Jul;51(1):33-45 [FREE Full text] [doi: 10.1016/j.amepre.2016.01.029] [Medline: 27039116]
10. Knight JR, Wechsler H, Kuo M, Seibring M, Weitzman ER, Schuckit MA. Alcohol abuse and dependence among U.S. college students. J Stud Alcohol 2002 May;63(3):263-270. [doi: 10.15288/jsa.2002.63.263] [Medline: 12086126]
11. Hingson RW, Zha W, Weitzman ER. Magnitude of and trends in alcohol-related mortality and morbidity among U.S. college students ages 18-24, 1998-2005. J Stud Alcohol Drugs Suppl 2009 Jul(16):12-20 [FREE Full text] [doi: 10.15288/jsads.2009.s16.12] [Medline: 19538908]
12. Weitzman ER, Nelson TF, Wechsler H. Taking up binge drinking in college: the influences of person, social group, and environment. J Adolesc Health 2003 Jan;32(1):26-35. [doi: 10.1016/s1054-139x(02)00457-3] [Medline: 12507798]
13. Harjutsalo V, Feodoroff M, Forsblom C, Groop P, FinnDiane Study Group. Patients with Type 1 diabetes consuming alcoholic spirits have an increased risk of microvascular complications. Diabet Med 2014 Feb;31(2):156-164. [doi: 10.1111/dme.12307] [Medline: 23952587]
14. MacNaught N, Holt P. Type 1 diabetes and alcohol consumption. Nurs Stand 2015 Aug 12;29(50):41-47. [doi: 10.7748/ns.29.50.41.e9812] [Medline: 26264336]
15. Barnard KD, Dyson P, Sinclair JM, Lawton J, Anthony D, Cranston M, et al. Alcohol health literacy in young adults with type 1 diabetes and its impact on diabetes management. Diabet Med 2014 Dec;31(12):1625-1630. [doi: 10.1111/dme.12491] [Medline: 24823681]
16. Tamony P, Holt R, Barnard K. The role of mobile applications in improving alcohol health literacy in young adults with type 1 diabetes: help or hindrance? J Diabetes Sci Technol 2015 Aug 06;9(6):1313-1320 [FREE Full text] [doi: 10.1177/1932296815588559] [Medline: 26251369]
17. Knychala MA, Jorge ML, Muniz CK, Faria PN, Jorge PT. High-risk alcohol use and anxiety and depression symptoms in adolescents and adults with type 1 diabetes mellitus: a cross-sectional study. Diabetol Metab Syndr 2015;7:24 [FREE Full text] [doi: 10.1186/s13098-015-0020-9] [Medline: 25821523]
18. Ramsey SE, Engler PA, Harrington M, Smith RJ, Fagan MJ, Stein MD, et al. Brief alcohol intervention among at-risk drinkers with diabetes. Subst Abuse 2010 Jan 01;4:1-8 [FREE Full text] [doi: 10.4137/sart.s4071] [Medline: 20305798]
19. Lakey WC, Barnard K, Batch BC, Chiswell K, Tasneem A, Green JB. Are current clinical trials in diabetes addressing important issues in diabetes care? Diabetologia 2013 Jun;56(6):1226-1235. [doi: 10.1007/s00125-013-2890-4] [Medline: 23564296]
20. Engler PA, Ramsey SE, Smith RJ. Alcohol use of diabetes patients: the need for assessment and intervention. Acta Diabetol 2013 Apr;50(2):93-99 [FREE Full text] [doi: 10.1007/s00592-010-0200-x] [Medline: 20532803]
21. Powell PW, Corathers SD, Raymond J, Streisand R. New approaches to providing individualized diabetes care in the 21st century. Curr Diabetes Rev 2015;11(4):222-230 [FREE Full text] [doi: 10.2174/1573399811666150421110316] [Medline: 25901504]
22. Weitzman ER, Salimian PK, Rabinow L, Levy S. Perspectives on substance use among youth with chronic medical conditions and implications for clinical guidance and prevention: A qualitative study. PLoS One 2019;14(1):e0209963 [FREE Full text] [doi: 10.1371/journal.pone.0209963] [Medline: 30673730]
23. Spijkerman R, Roek MA, Vermulst A, Lemmers L, Huiberts A, Engels RC. Effectiveness of a web-based brief alcohol intervention and added value of normative feedback in reducing underage drinking: a randomized controlled trial. J Med Internet Res 2010;12(5):e65 [FREE Full text] [doi: 10.2196/jmir.1465] [Medline: 21169172]
24. Norman P, Cameron D, Epton T, Webb TL, Harris PR, Millings A, et al. A randomized controlled trial of a brief online intervention to reduce alcohol consumption in new university students: Combining self-affirmation, theory of planned behaviour messages, and implementation intentions. Br J Health Psychol 2018 Feb;23(1):108-127. [doi: 10.1111/bjhp.12277] [Medline: 28941040]
25. Kerr D, King F, Klonoff DC. Digital Health Interventions for Diabetes: Everything to Gain and Nothing to Lose. Diabetes Spectr 2019 Aug;32(3):226-230 [FREE Full text] [doi: 10.2337/ds18-0085] [Medline: 31462878]
26. Jones E, Sinclair JM, Holt RI, Barnard KD. Social networking and understanding alcohol-associated risk for people with type 1 diabetes: friend or foe? Diabetes Technol Ther 2013 Apr;15(4):308-314. [doi: 10.1089/dia.2012.0327] [Medline: 23421853]
27. Kistin C, Silverstein M. Pilot Studies: A critical but potentially misused component of interventional research. J Am Med Assoc 2015 Oct 20;314(15):1561-1562 [FREE Full text] [doi: 10.1001/jama.2015.10962] [Medline: 26501530]
28. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]

29. Piette JD, Resnicow K, Choi H, Heisler M. A diabetes peer support intervention that improved glycemic control: mediators and moderators of intervention effectiveness. Chronic Illn 2013 Dec;9(4):258-267 [FREE Full text] [doi: 10.1177/1742395313476522] [Medline: 23585636]
30. Heisler M. Different models to mobilize peer support to improve diabetes self-management and clinical outcomes: evidence, logistics, evaluation considerations and needs for future research. Fam Pract 2010 Jun;27 Suppl 1:23-32 [FREE Full text] [doi: 10.1093/fampra/cmp003] [Medline: 19293400]
31. Brownson CA, Heisler M. The role of peer support in diabetes care and self-management. Patient 2009 Mar 01;2(1):5-17. [doi: 10.2165/01312067-200902010-00002] [Medline: 22273055]
32. Fisher EB, Boothroyd RI, Elstad EA, Hays L, Henes A, Maslow GR, et al. Peer support of complex health behaviors in prevention and disease management with special reference to diabetes: systematic reviews. Clin Diabetes Endocrinol 2017;3:4 [FREE Full text] [doi: 10.1186/s40842-017-0042-3] [Medline: 28702258]
33. Heisler M, Vijan S, Makki F, Piette JD. Diabetes control with reciprocal peer support versus nurse care management: a randomized trial. Ann Intern Med 2010 Oct 19;153(8):507-515 [FREE Full text] [doi: 10.7326/0003-4819-153-8-201010190-00007] [Medline: 20956707]
34. Kaselitz E, Shah M, Choi H, Heisler M. Peer characteristics associated with improved glycemic control in a randomized controlled trial of a reciprocal peer support program for diabetes. Chronic Illn 2019 Jun;15(2):149-156 [FREE Full text] [doi: 10.1177/1742395317753884] [Medline: 29343089]
35. Wisk LE, Nelson EB, Magane KM, Weitzman ER. Clinical trial recruitment and retention of college students with type 1 diabetes via social media: an implementation case study. J Diabetes Sci Technol 2019 May;13(3):445-456 [FREE Full text] [doi: 10.1177/1932296819839503] [Medline: 31010315]
36. Levy S, Dedeoglu F, Gaffin JM, Garvey KC, Harstad E, MacGinnitie A, et al. A screening tool for assessing alcohol use risk among medically vulnerable youth. PLoS One 2016;11(5):e0156240 [FREE Full text] [doi: 10.1371/journal.pone.0156240] [Medline: 27227975]
37. Weitzman ER, Magane KM, Wisk LE, Allario J, Harstad E, Levy S. Alcohol use and alcohol-interactive medications among medically vulnerable youth. Pediatrics 2018 Oct;142(4):e20174026 [FREE Full text] [doi: 10.1542/peds.2017-4026] [Medline: 30228168]
38. Wisk LE, Magane KM, Levy S, Weitzman ER. Alcohol use behaviors and reasons to abstain from or limit drinking among medically vulnerable youth. J Addict Med 2020;14(4):311-318. [doi: 10.1097/ADM.0000000000000603] [Medline: 31985512]
39. Alcohol screening and brief intervention for youth: A practitioner's guide. National Institute on Alcohol Abuse and Alcoholism, American Academy of Pediatrics. 2015. URL: https://www.niaaa.nih.gov/alcohols-effects-health/professional-education-materials/alcohol-screening-and-brief-intervention-youth-practitioners-guide [accessed 2021-08-03]
40. Johnston L, O'Malley P, Miech R, Bachman J, Schulenberg J, Patrick M. Monitoring the future national survey results on drug use, 1975-2018: overview, key findings on adolescent drug use. Institute for Social Research Report, The University of Michigan, MI. 2019. URL: https://eric.ed.gov/?id=ED594190 [accessed 2021-08-25]
41. Mallett KA, Varvil-Weld L, Turrisi R, Read A. An examination of college students' willingness to experience consequences as a unique predictor of alcohol problems. Psychol Addict Behav 2011 Mar;25(1):41-47 [FREE Full text] [doi: 10.1037/a0021494] [Medline: 21219039]
42. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):112-118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
43. Datye KA, Moore DJ, Russell WE, Jaser SS. A review of adolescent adherence in type 1 diabetes and the untapped potential of diabetes providers to improve outcomes. Curr Diab Rep 2015 Aug;15(8):51 [FREE Full text] [doi: 10.1007/s11892-015-0621-6] [Medline: 26084580]
44. Hilliard ME, De Wit M, Wasserman RM, Butler AM, Evans M, Weissberg-Benchell J, et al. Screening and support for emotional burdens of youth with type 1 diabetes: strategies for diabetes care providers. Pediatr Diabetes 2018 May;19(3):534-543 [FREE Full text] [doi: 10.1111/pedi.12575] [Medline: 28940936]
45. Harris SK, Aalsma MC, Weitzman ER, Garcia-Huidobro D, Wong C, Hadland SE, et al. Research on clinical preventive services for adolescents and young adults: where are we and where do we need to go? J Adolesc Health 2017 Mar;60(3):249-260 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.10.005] [Medline: 28011064]
46. Nicholas DB, Fellner KD, Frank M, Small M, Hetherington R, Slater R, et al. Evaluation of an online education and support intervention for adolescents with diabetes. Soc Work Health Care 2012;51(9):815-827. [doi: 10.1080/00981389.2012.699507] [Medline: 23078013]
47. Jaser SS, Whittemore R, Chao A, Jeon S, Faulkner MS, Grey M. Mediators of 12-month outcomes of two Internet interventions for youth with type 1 diabetes. J Pediatr Psychol 2014 Apr;39(3):306-315 [FREE Full text] [doi: 10.1093/jpepsy/jst081] [Medline: 24163439]
48. Kerr D, Gabbay RA, Klonoff DC. Finding real value from digital diabetes health: Is digital health dead or in need of resuscitation? J Diabetes Sci Technol 2018 Sep;12(5):911-913 [FREE Full text] [doi: 10.1177/1932296818771200] [Medline: 29719978]

49. Klonoff DC. Behavioral theory: the missing ingredient for digital health tools to change behavior and increase adherence. J Diabetes Sci Technol 2019 Mar;13(2):276-281 [FREE Full text] [doi: 10.1177/1932296818820303] [Medline: 30678472]
50. Reidy C, Klonoff DC, Barnard-Kelly KD. Supporting good intentions with good evidence: how to increase the benefits of diabetes social media. J Diabetes Sci Technol 2019 Sep;13(5):974-978 [FREE Full text] [doi: 10.1177/1932296819850187] [Medline: 31096768]
51. Chunara R, Wisk LE, Weitzman ER. Denominator issues for personally generated data in population health monitoring. Am J Prev Med 2017 Apr;52(4):549-553 [FREE Full text] [doi: 10.1016/j.amepre.2016.10.038] [Medline: 28012811]
52. Foster NC, Beck RW, Miller KM, Clements MA, Rickels MR, DiMeglio LA, et al. State of type 1 diabetes management and outcomes from the T1D exchange in 2016-2018. Diabetes Technol Ther 2019 Feb;21(2):66-72. [doi: 10.1089/dia.2018.0384] [Medline: 30657336]

Abbreviations

CGM: continuous glucose monitoring
CONSORT: Consolidated Standards of Reporting Trials
HbA_{1c}: glycated hemoglobin
OR: odds ratio
REDCap: Research Electronic Data Capture
T1D: type 1 diabetes

Edited by R Kukafka; submitted 10.12.20; peer-reviewed by R Wasserman, C Reidy, J Santos; comments to author 16.03.21; revised version received 15.04.21; accepted 06.05.21; published 30.09.21.

Please cite as:
Wisk LE, Magane KM, Nelson EB, Tsevat RK, Levy S, Weitzman ER
Psychoeducational Messaging to Reduce Alcohol Use for College Students With Type 1 Diabetes: Internet-Delivered Pilot Trial
J Med Internet Res 2021;23(9):e26418
URL: https://www.jmir.org/2021/9/e26418
doi:10.2196/26418
PMID:34591022

Original Paper

Consumer Perspectives on Maternal and Infant Health Apps: Qualitative Content Analysis

Rizwana Biviji[1], MS, PhD; Karmen S Williams[2], CPH, MBA, DrPH; Joshua R Vest[3,4], MPH, PhD; Brian E Dixon[4,5], FHIMSS, FACMI, MPA, PhD; Theresa Cullen[4,6], MS, MD; Christopher A Harle[7], PhD

[1]Science of Healthcare Delivery, College of Health Solutions, Arizona State University, Phoenix, AZ, United States

[2]Graduate School of Public Health and Health Policy, City University of New York, New York, NY, United States

[3]Department of Health Policy and Management, Richard M Fairbanks School of Public Health, Indiana University, Indianapolis, IN, United States

[4]Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN, United States

[5]Department of Epidemiology, Richard M. Fairbanks School of Public Health,, Indiana University, Indianapolis, IN, United States

[6]Department of Family Medicine, Indiana University School of Medicine, Indianapolis, IN, United States

[7]Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville, FL, United States

Corresponding Author:
Rizwana Biviji, MS, PhD
Science of Healthcare Delivery
College of Health Solutions
Arizona State University
550 North 3rd Street
Phoenix, AZ, 85004
United States
Phone: 1 602 496 3300
Email: Rizwana.Biviji@asu.edu

Abstract

Background: Despite the popularity of maternal and infant health mobile apps, ongoing consumer engagement and sustained app use remain barriers. Few studies have examined user experiences or perceived benefits of maternal and infant health app use from consumer perspectives.

Objective: This study aims to assess users' self-reported experiences with maternal and infant health apps, perceived benefits, and general feedback by analyzing publicly available user reviews on two popular app stores—Apple App Store and Google Play Store.

Methods: We conducted a qualitative assessment of publicly available user reviews (N=2422) sampled from 75 maternal and infant health apps designed to provide health education or decision-making support to pregnant women or parents and caregivers of infants. The reviews were coded and analyzed using a general inductive qualitative content analysis approach.

Results: The three major themes included the following: app functionality, where users discussed app features and functions; technical aspects, where users talked about technology-based aspects of an app; and app content, where users specifically focused on the app content and the information it provides. The six minor themes included the following: patterns of use, where users highlighted the frequency and type of use; social support, where users talked about receiving social support from friends, family and community of other users; app cost, where users talked about the cost of an app within the context of being cost-effective or a potential waste of money; app comparisons, where users compared one app with others available in app stores; assistance in health care, where users specifically highlighted the role of an app in offering clinical assistance; and customer care support, where users specifically talked about their interaction with the app customer care support team.

Conclusions: Users generally tend to value apps that are of low cost and preferably free, with high-quality content, superior features, enhanced technical aspects, and user-friendly interfaces. Users also find app developer responsiveness to be integral, as it offers them an opportunity to engage in the app development and delivery process. These findings may be beneficial for app developers in designing better apps, as no best practice guidelines currently exist for the app environment.

(J Med Internet Res 2021;23(9):e27403) doi:10.2196/27403

KEYWORDS

mHealth; mobile applications; maternal and infant health; smartphones; mobile phone

Introduction

Background

Globally, there is a growing demand for, and use of, mobile smartphone apps to disseminate maternal and infant health information and self-management tools effectively to pregnant and postpartum women [1-5]. This increase in acceptance of digital technologies is attributed to the fact that women find it extremely convenient to seek answers to their questions, with easy access to health-related information throughout the day [6]. Furthermore, these technologies may provide important social support, especially when pregnant women feel isolated, time constrained, or need reassurance [6], and apps have shown the potential to produce positive health behavior changes [7]. Maternal and infant health apps frequently appear on the Apple App and Google Play stores' list of most downloaded apps, and some have been downloaded over 5 million times [8].

Although popular, consumers often rate maternal and infant health apps lower (ie, fewer star ratings) than other categories of apps [9]. Moreover, the percentage of users who discontinue (ie, uninstall) the use of health-focused apps in general approaches 50% [10]. This disconnect between popularity, ratings, and use may be driven by several factors. One reason could be price—maternal and infant health apps are priced higher than apps targeting other populations and health conditions [9]. Alternatively, app efficacy may be a challenge. Apps with unclear or inaccurate consumer decision support could lead to poor end user choices and potentially undesirable outcomes [11]. Most maternal and infant health-focused apps are developed by non–health care organizations [12]. Furthermore, popular maternal and infant health apps typically include only a minority of behavior change techniques that are found to be useful in health promotion [13]. Apps often fail to provide sources of their information and lack warnings pertaining to the use of this information [5], or the reasons could be related to design and usability. Overall, end users have discontinued apps because of the time-consuming data entry process, hidden costs, use difficulty, lack of data privacy, and lack of maintained interest [10]. User commentaries from women's health apps indicate that women are in favor of apps that are easy to use, contain new information, and are motivational, whereas areas that need improvements are associated with the quality of graphics, download speed, compatibility with other devices, ability to transfer data onto newer versions, and certification or affiliation with credible organizations [9].

Objectives

Mobile technology can be an effective platform for delivering resources and interventions, but consumer engagement remains a barrier for uptake and continued use [14]. Therefore, it is imperative to consider the consumer's perspective on these apps to improve the utility of such resources [14]. The objective of this study is to assess users' self-reported experiences with maternal and infant health apps, perceived benefits, and general feedback by analyzing publicly available user reviews on two popular app stores—Apple App Store and Google Play Store.

Methods

Study Design

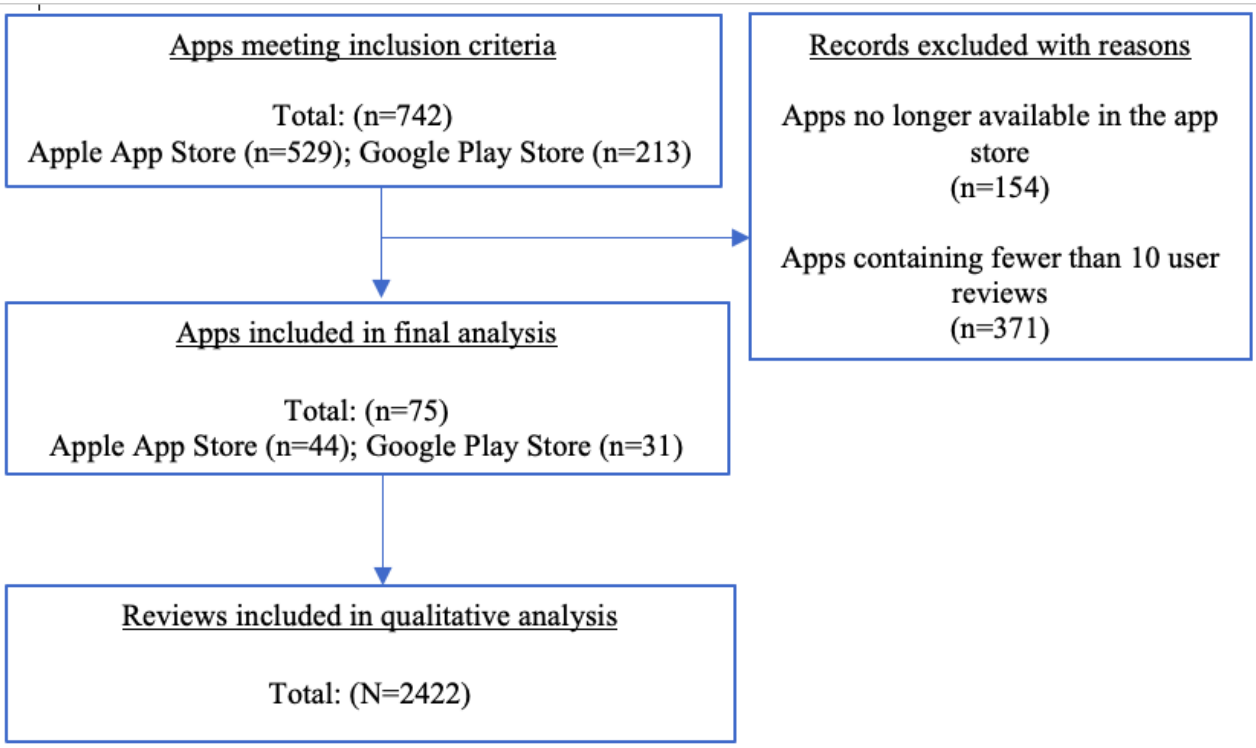
We conducted a qualitative assessment of publicly available user reviews from a sample of maternal and infant health apps available in the Apple App Store and Google Play Store. We followed a four-step process to create a data set of user reviews from a set of maternal and infant health apps that offer health education or decision-making support to pregnant women or parents or caregivers of infants. First, we populated a database with the set of available maternal and infant health apps by systematically fetching app details from the Apple App Store [15] and Google Play Store [16] using a Java-based scraper program called Node.js (OpenJS Foundation) [17]. Through this automated process, data are retrieved from the web using http in a central database [18]. Scraping programs automate the extraction of information displayed to users in the same way that chart reviewers record details from a medical chart. Second, we identified an inclusive list of relevant keywords that users might search for when locating apps related to maternal and infant health. Third, we scraped the two app stores for candidate apps and merged and deduplicated the resultant apps first within stores and then across stores. We then applied inclusion and exclusion criteria to identify apps that were eligible for the study. Inclusion criteria were as follows: (1) app description written in English; (2) target users judged as to-be pregnant women, to-be parents, and other caregivers of infant children as primary users; (3) listed in the medical, health and fitness, books and references, and education categories in the Apple App Store or listed in the medical, health and fitness, books and references, education, and parenting categories in the Google Play Store; and (4) described as intending to provide health education and user decision-making support. Exclusion criteria were as follows: (1) app description written in any language other than English; (2) inadequate or no app description provided; (3) target users judged as health professionals, providers, and students in health professions as primary users; (4) listed in all other categories; (5) described solely for purposes such as gestational age or due date calculators, identifying baby names, and shopping for baby products; and (6) apps meant to be used by members and patients associated with special programs (eg, fitness centers, clubs, or other paid memberships) or health care facilities (eg, a clinic or hospital). Fourth, we sampled a set of user reviews from the candidate maternal and infant health apps. The data reflected the app store content as of February 2018. Steps 1 to 3 are described in detail elsewhere [12]. The Indiana University Institutional Review Board approved this study as nonhuman subjects' research.

Search Strategy

The three steps listed earlier resulted in a total of 742 eligible maternal and infant health apps (Figure 1). Of these, 20.8% (154/742) were no longer available in the app stores at the time

of data collection (and therefore excluded). Apps with fewer than 10 (371/742, 50%) reviews were also excluded to obtain a robust set of reviews for analyses. From the 217 resulting apps, a simple random sample of 75 (34.6%; 44 Apple App Store and 31 Google Play Store) apps were selected. The sample size was sufficient to achieve data saturation. Each app name was entered into an individual scraper search. Each scrape resulted in maximum allowed user reviews of 50 from the Apple App Store and 40 user reviews from the Google Play Store (ordered by the most recent review). The upper limit per app was set at 40 user reviews for a total sample of 2422 user reviews (Figure 1) to maintain consistency across platforms. In addition to the text of the user reviews, the scraper program returned the star rating, price, and review date (see Multimedia Appendix 1 for a list of apps included in this study). Finally, using a general inductive content analysis approach, we qualitatively analyzed user reviews, where the underlying themes were identified from the data.

Figure 1. Schematic representation of the user review selection process for qualitative analysis.



Data Analysis

Descriptive Summary

Descriptive statistics were used to summarize the app characteristics such as app price, user star ratings, app platform, and review characteristics such as publication date and length of reviews.

Qualitative Analysis

User reviews were analyzed using a general inductive content analysis approach [19,20]. In this approach, themes were derived from data, as opposed to using preconceived categories [21]. First, 2 coders (RB and KSW) undertook a joint reading of a random sample of 30 user reviews to establish consistency in the textual unit of analysis, identification of categories, and formation of themes [20]. Next, the same 2 coders independently read and identified the preliminary codes from a new sample of 200 user reviews. Through joint reading sessions, these preliminary codes were refined, collapsed, assigned descriptive labels, and arranged into a coding framework to be applied to the sample. To assess coding consistency, both the coders independently analyzed a new sample of 100 user reviews. Agreement between the coders was high (κ=0.89). The 2 coders independently coded the remaining sample and met regularly to resolve any coding discrepancies and discuss the themes that were detected in the data. All data management and analyses were conducted in Dedoose 8.0.39 (SocioCultural Research Consultant, LLC) [22].

Results

Description of Apps and Reviews

Of the apps that were included, 75% (56/75) were free. Of the paid apps, the prices ranged from US $0.99 to US $8.99, with an average price of US $3.14 and a median of US $2.99. These prices reflected one-time payments to download the app, not monthly or annual subscription fees. More than half (44/75, 59%) of the apps were from the Apple App Store, and the remaining were from the Google Play Store. The average star rating for all apps was 4.1, with a range of 2.5-5.0. The oldest review in the data was written on July 5, 2009, whereas the latest review was published on February 13, 2018. The longest user review was 403 words long, and the shortest was only 1 word. Three major and six minor themes (present in less than 10% of the reviews) were identified in the data (Figure 2). The definitions of the themes, along with the frequency of occurrence, are presented in Table 1.

Figure 2. Summary of themes identified from qualitative content analysis.

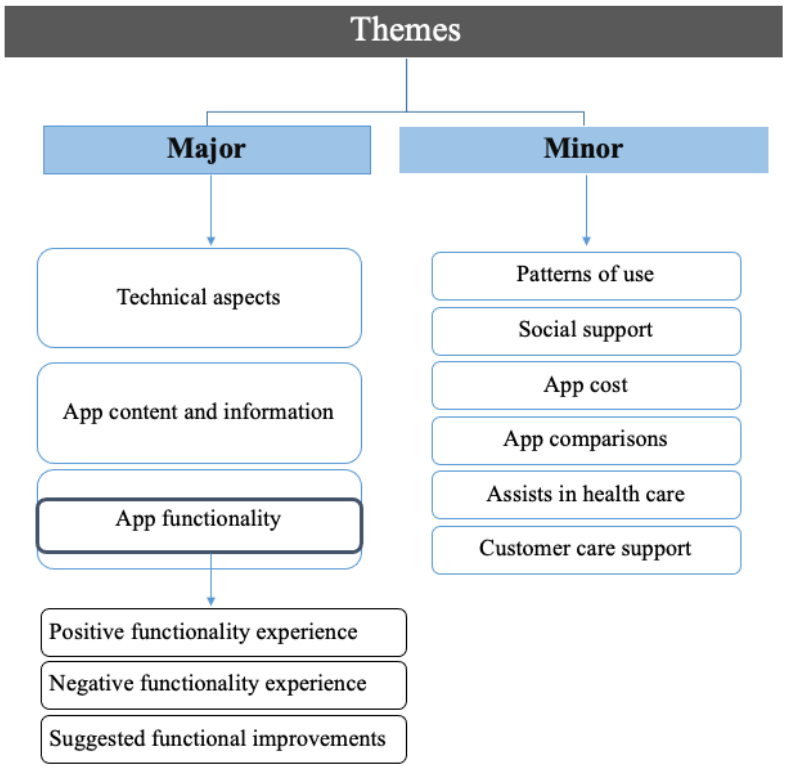


Table 1. Prevalence of major and minor themes identified in the data (N=2422).

Themes[a]	Definition	Reviews, n (%)
Major		
App functionality	The user specifically talks about app functions and features and what it does either positively or negatively.	2119 (87.49)
Technical aspects	The user specifically talks about aspects pertaining to how an app operates either positively or negatively, that is, privacy, security of data, or other technology-based aspects.	510 (21.05)
App content	The user specifically talks about the app content and information it provides.	349 (14.41)
Minor		
Patterns of use	The user highlights the frequency and type of use—whether the app is used for the first pregnancy or the app is being used for a long time.	219 (9.04)
Social support	The user specifically talks about receiving support from friends and family or offering support to other women while using the app.	207 (8.54)
App cost	The user specifically talks about the cost of an app, ie, cost-effective, or a waste of money.	201 (8.3)
App comparisons	The user compares the app with other apps in the market.	104 (4.29)
Assists in health care	The user specifically highlights the role of the app in offering clinical assistance.	77 (3.18)
Customer care support	The user specifically talks about their interaction with the app customer care support either positively or negatively.	44 (1.82)

[a]These themes were not exclusive.

Major Themes

App Functionality

Around 87.49% (2119/2422) of user reviews focused on commenting on the functionality of apps, such as app features, indicating that users frequently discuss the functionality of an app while using them. The discussion about app functionality featured three main subthemes: (1) positive functionality experience, (2) negative functionality experience, and (3) suggested functional improvements.

Positive Functionality Experience

A majority of the reviews focusing on app functionality indicated an overall positive experience with the functionality of the app. Many of the positive reviews were geared toward an app's overall design and app features describing these apps as useful, helpful, and easy to use. Users mainly elaborated on

app features that assisted in data recording and tracking of data pertaining to infant needs such as feeds, milk pumps, diaper changes, and sleep, akin to a *one-stop shop.* Some of the other tracking features focused on recording the height and weight of the infant, monitoring the trends over time, and keeping a log of immunization and other health histories:

> *Best tracking all app ever! ibabylog is the best app I have! It helps me track everything from nap times, to meal times, to dirty diapers. Its best feature is that it helps you keep intact with breast feedings. You can track everything in between like medication, doctor visits, growth, anything you can think of. I love this app and I recommend it big time!*

Users have also expressed a general preference toward the feature of data sharing and synchronizing of data between two or more devices:

> *Awesome App! My wife has been using MammBaby on her iPhone and now I am using it on my Android phone. It is definitely the easiest one to use and sharing feature is awesome!*

Some of the other features that have received favorable attention include baby heartbeat self-reported monitors, contraction timers, week-by-week fetal development stages, and diet or exercise tips and recommendations:

> *With countless pregnancy and newborn resources on the market, it can be hard for parents to know where to turn for information. This app is created by experts, for parents. As a mother, I appreciate the easy-to-understand and reassuring information found in the app. Like most pregnancy apps, the week-by-week updates are perfect. And the contraction timer came in handy on the big day! Thank you, Lamaze, for cutting the fuss and giving parents what we really need: power to make the best decisions for our family!*

Negative Functionality Experience

A frequent complaint pertained to app features that did not meet user expectations. Reviewers expressed concerns over apps that presented inaccurate calculations to users, that is, wrong due dates or height, weight, or percentile calculations:

> *Inaccurate Percentile Calculation, I don't know how they're trying to calculate the percentiles by age, but it's clearly not working. None of the percentiles in the app have matched up with either the hospital's or the pediatrician's data for my son--by a long shot! Even without the doctor's records telling me he's somewhere near the 75th percentile, it's obvious that an 8 lbs. boy is NOT in the 30th percentile for weight for his age at 13 days old, as the app reports.*

Users also highlighted issues pertaining to loss of data or inability to synchronize data:

> *Have to leave app running or else you lose your contractions data. Isn't that kinda the point in having a contraction app? I will be deleting this one and finding something that actually will track my contractions.*

Similarly, there were complaints about apps with no updates or poor updates:

> *Terrible update! Before this last update on 7/30, the app was fantastic! But now I HATE IT! It's too cutesy and it's hard to tell anything anymore! The photo section is far too large now when that is not the information that I need. The daily schedule section is MUCH more confusing, and I can't see any patterns. I will be deleting and finding a new app now! It is so confusing now. They went with trying to make it look cute instead of functional. Horrible app!!!!*

Finally, users reported that they discontinued the use of an app with inadequate features, with limited ability to track data, modify an inaccurate entry, or lacking flexibility in using conversion metrics, that is, mg to oz or kg to lbs:

> *I deleted it after one time, this app is terrible. It just tells you how much weight you should gain. It doesn't track your weight or monitor your progress. It doesn't even save any information. You can find this same information on several sites or easily figure it out on your own. And either way, if you want to keep track, you would have to do it yourself anyway. I deleted the app after the first time I tried it and realized it was useless.*

Suggested Functional Improvements

Along with critiquing an app's functionality and features, users often provided additional recommendations pertaining to features that could be added to improve their overall experience with the app. Reviewers often had a bucket list of requests to customize an app and tailor it to their specific needs and requirements. Textbox 1 provides a summary of suggestions for functional improvements. Some of the more common recommendations focused on data visualization or trends data (for height, weight, feedings, or immunizations), ability to export or print data, additional tracking facilities, or capabilities to synchronize data between two or more devices.

Textbox 1. Suggested functional improvements for maternal and infant health apps.

Requested Features and Direct Quotes

- Export data
  - "Helpful, but...I would love to be able to export or print data. Otherwise helps me keep track of my preemie between Dr visits."
- Data visualization and trends data
  - "The only thing that would make it better in my opinion would be to have a graph of each activity to show any types of trends. Knowing that my baby has slept 14% more in this 7-day period vs the last 7 days is maybe interesting, but not helpful when I'm trying to see any patterns of when he's sleeping and for how long. Having a graph that shows when he sleeps each day for a week would be so helpful for getting into more of a natural routine with the baby."
- Additional data tracking: symptoms, weight, illness, pumping, feeding, and medications
  - "Wish it had a medicine tracker and reminder, Wish the app helped with keeping track of when a dose of the medication was given and help with reminders for the next dose."
- Alerts, alarms, and reminders
  - "A Great App!! I use this app all the time! I pretty much rely on it! The only thing I've noticed that I'd like them to add is an alarm feature for medicines."
- Data synchronization between devices and data integration across apps
  - "I wish they could allow for other apps to feed info into it. For example, I use a digital thermometer that connects to my phone. If it could sync with this app it would be perfect!"
- Edit or delete incorrect data entries
  - "I wish I could edit time entries in case I look to see which side but forgot to click a new entry. But overall, great and easy to use."

Technical Aspects

The second most common theme (510/2422, 21.05%) that users discussed in their reviews were associated with the technical aspects of an app, such as how an app operates. Users preferred apps with a user-friendly app interface, free from advertisements or forced ratings, and those that did not occupy a large space in the phone memory.

However, most of the comments were geared toward technology failures, such as apps that crashed or froze on users, thereby rendering them useless for further use:

> *Unfortunately, it crashes frequently (usually in the middle of the night when I'm trying to time nursing sessions) and multiples times in a row at that (I think my record is 7). Incredibly irritating when you're sleep deprived, it's 2 am, and all you want to do is get back to sleep not continuously open an app all, so it can hang for 30 seconds then crash.*

Users have also criticized certain apps that were extremely slow to load the information or content or stream videos:

> *Generally good, a little slow, switched to this app once the written log from our lactation consultant ran out of space. It's been very helpful for tracking feedings and diapers. I find, however, that it runs really slowly sometimes in terms of entering information and the nursing/pumping timer.*

Other reasons cited as causes for discontinuing app use were inability to register oneself, not available offline, and other glitches associated with synchronizing data or loss of data:

> *It no longer lets me register or log in, and every time I open it I get the message that its opening for the first time, with a long wait. I probably won't use it when I'm in labor because of this.*

A few users also recommended certain technical improvements that would help enhance their overall experience with the app. These recommendations pertained to app availability on multiple devices and app platforms, or making an app esthetically pleasing (improved user interface, fonts, or image quality):

> *LOVE, I absolutely love this app, I just wish they would make it Apple Watch compatible as well!*

App Content and Information

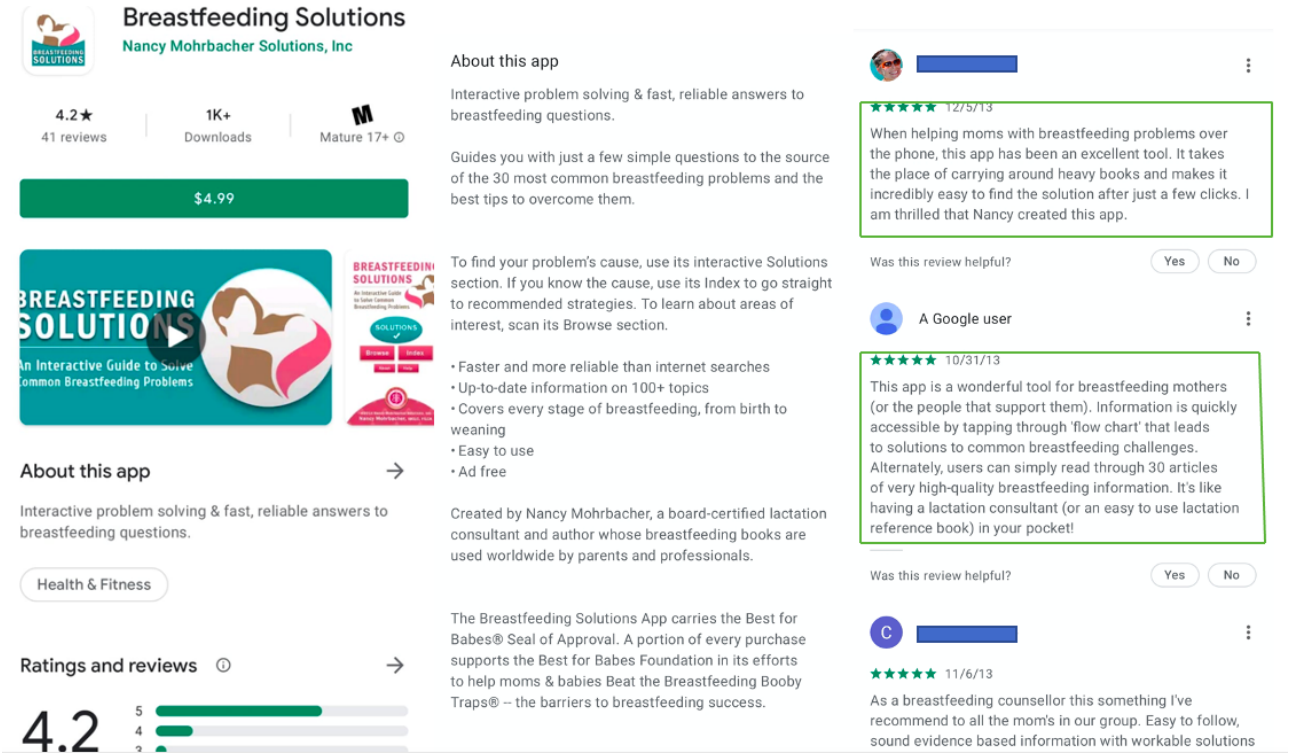
Reviews related to app content or information typically referred to the *wealth of information* these apps had to offer. Users appreciated apps providing detailed information, daily or weekly tips and reminders, latest and updated content, and evidence-based articles:

> *An evidence-based app you can trust, the information provided by Lamaze comes from evidence-based research, so you know you're not getting the "fluff" that you read in other pregnancy or parenting apps.*

New parents expressed their satisfaction with apps that met their increased information needs. A few users also suggested that certain apps were better organized and offered greater support as opposed to seeking information on the internet. Other users indicated greater satisfaction with using an app for information seeking over the traditional use of books (Figure 3):

> *Yes! Love love love this app, used it for my first pregnancy and it put me at such ease being a first time mom, I love how they have something to say everyday because you get so anxious being pregnant and just want the baby and this app helps you calm down and understand every stage. I love that you can put pictures at the bottom and see your progress before the birth of your child.*

Figure 3. User reviews centered around app content and information.



However, not all users found the content of these apps to be beneficial. Some users articulated a general dislike toward apps that had very brief content, contained basic information that could be found on the internet, or comprised a number of typos or grammatical errors, making it incomprehensible. This was especially true for paid apps, where users had higher expectations in terms of content quality:

> *Underwhelming application., I cannot speak to the integrity of the information this application gives you, A friend of mine told me to purchase it as I was expecting. For the price I expected much more. The Videos section simply opens their YouTube page in a web browser. And the very limited information the application gives you could fit on one sheet of paper. I was expecting for suggestions on how to deal with or enrich my child's experience during her leaps, but instead I only found a very brief description of what was happening developmentally. Certain aspects of the application just don't work at all, you can tell this application was outsourced to international developers for whom English is a second language. I would perhaps give this application three stars if it were free, but considering I paid for it I was expecting more. I'm sure all the information in this app can be found online with little effort. Save your money on this one.*

A few users were also disapproving of apps that contained unrelated information or were very narrow in scope; for example, they covered a selective range of topics that were not applicable to many:

> *I didn't put five stars because I disapprove the zodiac stuff, I don't find it useful, fun, or educational (on the contrary).*

Furthermore, there were reviews that offered suggestions to improve the app content. Most of the recommendations were geared toward offering content in a different language and adding additional content:

> *I've got a son with one of the rarest of rare genetic diseases that I thought would be awesome for this app it's called Spinal Muscular Atrophy Respiratory Distress also known as (Smard 1).*

Some recommendations were also for simplifying the information to account for the lay audiences:

> *Hard to understand, the information is correct. However, it would be extremely helpful to nursing mommies if they understood what the medical mumbo jumbo meant!*

Minor Themes

Patterns of Use

Over 9.04% (219/2422) of the reviewers described their patterns of app use by identifying the length, duration, and frequency of use. Some of them were long-term users, using these apps for multiple pregnancies. Users have also highlighted their frequency of use, with some users using these apps almost daily.

In addition, most app users tend to be first-time parents who use maternal and infant health apps for information seeking, tracking their pregnancy or infant needs, or connecting with other parents for moral support:

> *Great app! I don't normally write reviews for apps but this app is amazing. I am a new mom and I use it throughout the day, every day. I use it to keep track of feeding and sleeping trends.*

Social Support

Over 8.54% (207/2422) of all reviews referred to apps that acted as a safe haven for users, as these apps connected pregnant women and mothers to a larger community of app users, thereby offering social support that is typically needed by many during this period. First-time mothers especially found these communities encouraging and useful to lessen their anxieties over issues where they had minimal experience:

> *Awesome video blogs, never seen video blogs from other moms before and it helped me feel like I wasn't the only one. Same goes for the forum. Really good to connect to other moms.*

However, a few reviewers were not satisfied with their interactions within such communities. A handful of women faced some level of cyberbullying by others in these communities, especially those who had differing perspectives or values pertaining to childcare. Such negative experiences were geared toward the strong opinions, biases, and immaturity of other users.

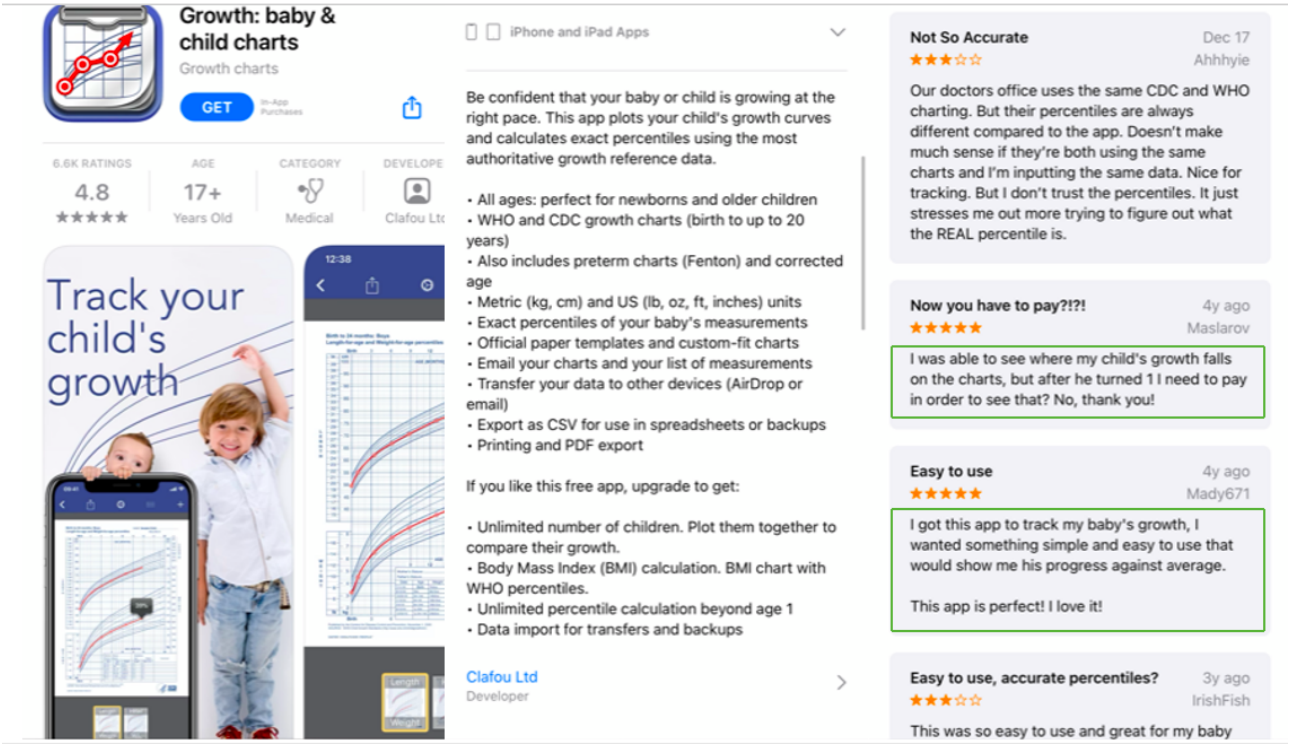
Users also expressed an overall appreciation for apps that had the capability to synchronize their data with other caregivers, especially their spouses, family members, or significant others. These features helped parents stay abreast with their infants' activities at all times:

> *Great app, I didn't know a lot about pregnancy and it has helped a lot with the different stages. It has even given my wife some comfort because I know things when she asks me questions.*

App Cost

Generally speaking, app users preferred apps that were free or had a very low cost associated with it (Figure 4). Users were willing to purchase apps with enhanced capabilities, such as data synchronization on multiple devices, additional tracking features, or data visualization functions. Users were also willing to pay for a *pro* version to eliminate advertisements and pop-ups and appreciated apps that were esthetically appealing.

Figure 4. User reviews centered around app cost and ease of use.



However, users were disapproving of apps that were labeled as free on the app store but offered very limited functionalities and required an upgrade (pro version) to enjoy additional features. Users called such apps *misleading*, *unfair*, or *false advertising*. Similarly, users were very critical of apps that were paid and did not meet their expectations. Users associated price with quality, thereby voicing extreme distress if the app faced any technical issues, provided basic information, or contained mediocre features:

> *Lot of glitches, poor info, no matter what article I click on- every other page starts with the same header: "Article good fats from bad ones" and then an incomplete sentence about using safe household cleaners...I tried to leave a review ON one of these pages- because EVERY page asks you for a review, and it said "sorry we couldn't complete your review right now." I cannot imagine who would pay 6.99 for this. (Or that they raised the price from 4.99???) I*

downloaded it for exercises but they don't have videos to show the exercises so I'm better off just going back to YouTube.

App Comparisons

Approximately 4.29% (104/2422) of users compared certain app features, either in a positive or negative context, with other apps that they used in the past or were currently using. Users also compared paid apps with apps that were offered for free to establish whether the app was worth the purchase or not:

Meh, Really...I have a ton of other baby apps that give more details on my baby's development and what I should be eating. This app is lacking a lot of details I wouldn't waste my money on going "premium." I'll stick with the bump and nurture. Thanks.

Assists in Health Care

Approximately 3.18% (77/2422) of app users established the role of maternal and infant health apps in health care management. Consumers especially found these apps useful for keeping track of infants who were underweight, short for age, premature at birth, or with congenital birth defects. They confirmed the role of these apps in recording accurate data that could then be reported effectively to their pediatrician:

I have used this for the last six months with no troubles. It's quick, intuitive, and easy. I have three kids and this baby was a newborn ICU baby, and this made life simpler with multiple caretakers/doctors who want all the details (and not great communication between all the people involved in baby's life.) and exhaustion from all the demands. What a lifesaver!

A minority of reviews were initiated by the health care professional, where they specified either using these apps for information seeking or monitoring patients' progress. Health care professionals have also provided recommendations for certain apps to their patients.

Customer Care Support

Users particularly cared about customer care responsiveness in addressing their questions, promptness in fixing issues, as well as fulfilling their requests for additional features. Consumers expressed distress over situations where they paid for an app but did not receive adequate developer support:

Doesn't work at all, I purchased this app and was so excited after reading the reviews. However, this app is crap! Took my money but does not work at all, nothing. Stays on one page and that's it. I can't get in touch with app support either. Don't waste $$$ or time. I have tried reinstalling it several times and still crap.

Discussion

Principal Findings

To our knowledge, this is the first study to evaluate consumer preferences for mobile apps targeted toward maternal and infant health using publicly available user review data. The elements of maternal and infant health apps that users were most interested in were related to the functionality of apps, that is, app features. Users were satisfied with apps that offered advanced features such as data monitoring and tracking or data synchronizing abilities across different devices. Similarly, users were highly critical of apps that did not meet user expectations in terms of their functionality and were prompted to discontinue the use of apps with limited functionalities. A large number of comments were concerned with the loss of data or inability to edit an incorrect entry. The overall emphasis on app functionality is consistent with previous literature that highlights the role of user satisfaction with app functions as one of the major caveats in consumer app use [14,23]. In addition, users also offered suggestions or recommendations in terms of app features that would improve their overall experience with the app. This is again consistent with previous reports in other domains [14,24,25], indicating that consumers are not fully content with the available features and their needs are often inadequately addressed.

Consumers prefer apps that are easy to use, with an esthetically pleasing interface, and occupy less memory space. The reasons cited for discontinuation of app use centered around technical issues such as crashing or freezing of apps while in use, slow download speed, and too many pop-ups or forced ratings [9,26,27]. An app that crashes often is deleted by users before any engagement with key functions has occurred [23]. Thus, app developers may give careful consideration to the technical aspects of these apps before releasing them for use. One of the major aspects of consideration for app use is the quality of information or content provided by these apps. When it comes to content, users typically value tailored information pertaining to their condition and actionable solutions for its effective management. Pregnant women and new mothers often seek information from apps on topics such as establishing breastfeeding, solving breastfeeding problems, infant health issues, and topics that are uncomfortable to discuss with health care providers [28]. Overall, women reported a positive experience with pregnancy apps, but a few reported issues with what they perceived as validity, accuracy, and timeliness of the information that was being presented by certain apps.

App cost was recognized as an important consideration for app adoption and use. Most users preferred free apps; however, users were willing to pay for apps if they offered sophisticated features and comprehensive information and were of superior quality. These results are consistent with those of previous studies on maternal and infant health apps [12]. Although some users discussed a negative experience with app-based communities, most women expressed their enthusiasm in terms of the social support they received from fellow mothers undergoing similar experiences. Women value peer experiences and knowledge on important topics such as breastfeeding or infant care and, in turn, offer similar support to others in need. Women often use these platforms to discuss topics that are sensitive in nature, such as sexual activity during and after pregnancy or feelings of hopelessness [29]. The anonymity offered in such communities provides an opportunity to raise issues that they otherwise would not discuss with family, friends, or health care professionals during in-person visits [29]. Aside from being a part of an ongoing community of app users, women

have also cherished the ability to engage their partners and other family members in maternal and infant health care and support. Information seeking during pregnancy and postpartum was relatively higher among first-time mothers who use these apps to track stages of pregnancy on a weekly basis or monitor infant development. Women also report using these apps for multiple pregnancies, and some engage with maternal and infant health apps almost daily.

Increasingly, women are using maternal and infant health apps to assist them in health literacy, monitoring, self-management, as well as consumer decision-making. Reviewers elucidated the role of maternal and infant health apps in improving patient-provider communication by aiding in data tracking, which helps improve recall and increase preparedness for doctor visits. Certain apps offer additional features where users can print reports or directly email them to their providers. This increased use of mobile apps during pregnancy and postpartum periods also highlights the importance of providing evidence-based information, especially because of the vulnerable nature of these phases. Very few reviewers discussed the availability of evidence-based content or expressed a desire to identify the scientific sources of information that are being presented to them. This may be credited to the fact that users are more concerned with the overall appeal of an app in terms of functionality, features, content, or usability, as opposed to verifying the credibility of this information. Therefore, health care professionals, app developers, and policy makers may consider strategies to review and promote apps to consumers based on information accuracy and trustworthiness. Future research may focus on evaluating the quality of maternal and infant health app content and information.

Furthermore, there is increased satisfaction among users whose needs and viewpoints were adequately addressed by the app developers. This highlights the fact that consumers value increased involvement in the app development and delivery process, which may increase their engagement and long-term use.

These results may provide meaningful information for app developers and other stakeholders regarding consumer needs and expectations pertaining to maternal and infant health apps, which are often complex and multidimensional. App developers may also consider strategies to collaborate with health care professionals for app content development and technical experts for app interface and usability to provide evidence-based information with superior functionality to users at low cost.

Strengths and Limitations

A major strength of this study is that it analyzes consumer attitudes and perspectives toward maternal and infant health apps using a large and diverse sample of publicly available user review data from Apple App and Google Play stores. This study uses a systematic approach of sampling a comprehensive set of user reviews for analysis. Importantly, to our knowledge, this is the first study that uses this particular methodology and sampling strategy to study users' perspectives on maternal and infant health apps.

This study had some limitations. First, it is likely that user reviews on the app store may not be representative of the larger population of app users. User opinion on a given app may be affected by gender, economic status, education level, or other demographics. However, these data elements were not publicly available and were not considered to draw conclusions. The study, therefore, uses information similar to what would normally be available to consumers in a *real-world* context before downloading an app. In addition, the study used a random sample of user reviews using a strategic approach to capture a more diverse sample of app users, which probably could not be achieved using other exploratory methods such as interviews or focus groups. Second, our results may not be generalizable to apps belonging to other health domains; for example, app use for monitoring infant needs may not be applicable to other populations such as for diabetes or weight loss. Third, the motivation to post may be to give a negative review to express dissatisfaction with the app [30]. However, in our random sample of apps, positive reviews were more common, which is consistent with previous reports [14,31]. Finally, there is a possibility of fake reviews to increase the app demand [32]; however, the large number of reviews per app should compensate for this challenge.

Conclusions

This review extends the literature by emphasizing the features of maternal and infant health apps that are particularly important to users. These results may be beneficial for app developers to consider during app development. Overall, consumers value low-cost apps that have high-quality content, superior features, and smooth technical aspects and are easy to use. Users consider app developer responsiveness an integral part of app use, as it empowers them in the process of app development and delivery. These consumer perspectives are essential for mobile health sustenance, as no best practice guidelines currently exist for the app environment. Users are increasingly using apps for health care management and informed decision-making. Thus, health care professionals, app developers, and policy makers may consider strategies to review and promote evidence-based and trustworthy apps to consumers. Future studies may focus on assessing user experiences by using other qualitative methods to garner detailed perspectives on long-term app use. This study should also be replicated in other health domains to gain a greater sense of consumer perspectives in the field of mobile health. Future studies may also focus on developing a framework for consumers to evaluate app quality for effective app comparisons and use decisions.

Acknowledgments

The authors would like to thank the Department of Health Policy and Management at Indiana University Richard M Fairbanks School of Public Health and Regenstrief Institute, Inc, for their ongoing support during the research process. KSW was supported by a training grant (award T15LM012502) from the National Library of Medicine of the National Institutes of Health.

Authors' Contributions

The primary author, RB, proposed and completed this study as part of her doctoral dissertation. The coauthors KSW, JRV, BED, TC, and CAH (chair) contributed significantly to the acquisition, analysis, and interpretation of data; rigor of the methodology and analysis; and writing of the manuscript.

Conflicts of Interest

JRV is a founder and MND equity holder in Uppstroms, a health technology company. Other authors have declared no conflict of interest.

Multimedia Appendix 1
List of apps included in the qualitative analysis.
[DOCX File , 26 KB - jmir_v23i9e27403_app1.docx]

References

1. Tamrat T, Kachnowski S. Special delivery: an analysis of mHealth in maternal and newborn health programs and their outcomes around the world. Matern Child Health J 2012 Jul;16(5):1092-1101. [doi: 10.1007/s10995-011-0836-3] [Medline: 21688111]
2. Rotheram-Borus MJ, Tomlinson M, Swendeman D, Lee A, Jones E. Standardized functions for smartphone applications: examples from maternal and child health. Int J Telemed Appl 2012;2012:973237 [FREE Full text] [doi: 10.1155/2012/973237] [Medline: 23304136]
3. Declercq ER, Sakala C, Corry MP, Applebaum S, Herrlich A. Listening to Mothers III: Pregnancy and birth. Childbirth Connection. 2013 May. URL: https://www.nationalpartnership.org/our-work/resources/health-care/maternity/listening-to-mothers-iii-pregnancy-and-birth-2013.pdf [accessed 2018-02-15]
4. O'Higgins A, Murphy OC, Egan A, Mullaney L, Sheehan S, Turner MJ. The use of digital media by women using the maternity services in a developed country. Ir Med J 2014;107(10):313-315. [Medline: 25556255]
5. Lee Y, Moon M. Utilization and content evaluation of mobile applications for pregnancy, birth, and child care. Healthc Inform Res 2016 Apr;22(2):73-80 [FREE Full text] [doi: 10.4258/hir.2016.22.2.73] [Medline: 27200216]
6. Hearn L, Miller M, Lester L. Reaching perinatal women online: the Healthy You, Healthy Baby website and app. J Obes 2014;2014:573928 [FREE Full text] [doi: 10.1155/2014/573928] [Medline: 24872891]
7. Hebden L, Cook A, van der Ploeg HP, Allman-Farinelli M. Development of smartphone applications for nutrition and physical activity behavior change. JMIR Res Protoc 2012 Aug 22;1(2):e9 [FREE Full text] [doi: 10.2196/resprot.2205] [Medline: 23611892]
8. Lupton D, Thomas GM. Playing Pregnancy: The ludification and gamification of expectant motherhood in smartphone apps. M/C Journal 2015 Oct 01;18(5):1 [FREE Full text] [doi: 10.5204/mcj.1012]
9. Derbyshire E, Dancey D. Smartphone medical applications for women's health: what is the evidence-base and feedback? Int J Telemed Appl 2013;2013:782074 [FREE Full text] [doi: 10.1155/2013/782074] [Medline: 24454354]
10. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
11. Buijink AW, Visser BJ, Marshall L. Medical apps for smartphones: lack of evidence undermines quality and safety. Evid Based Med 2013 Jun;18(3):90-92. [doi: 10.1136/eb-2012-100885] [Medline: 22923708]
12. Biviji R, Vest JR, Dixon BE, Cullen T, Harle CA. Factors related to user ratings and user downloads of mobile apps for maternal and infant health: cross-sectional study. JMIR Mhealth Uhealth 2020 Jan 24;8(1):e15663 [FREE Full text] [doi: 10.2196/15663] [Medline: 32012107]
13. Biviji R, Vest JR, Dixon BE, Cullen T, Harle CA. Content analysis of behavior change techniques in maternal and infant health apps. Transl Behav Med 2021 Mar 16;11(2):504-515. [doi: 10.1093/tbm/ibaa039] [Medline: 32491165]
14. Nicholas J, Fogarty AS, Boydell K, Christensen H. The reviews are in: a qualitative content analysis of consumer perspectives on apps for bipolar disorder. J Med Internet Res 2017 Apr 07;19(4):e105 [FREE Full text] [doi: 10.2196/jmir.7273] [Medline: 28389420]
15. iTunes. Apple Inc. URL: https://itunes.apple.com/us/genre/ios/id36?mt=8 [accessed 2018-02-01]
16. Google Play. Google Inc. URL: https://play.google.com/store/apps [accessed 2018-01-31]
17. Node. OpenJS Foundation. URL: https://nodejs.org/en/ [accessed 2018-02-01]
18. Shah S, Adhikari S, Subedi S, Chalise U. A major project report on web virtual assistance. Bachelor's Project. Tribhuvan University Institute of Engineering, Lalitpur, Nepal. 2017 Aug. URL: https://www.researchgate.net/profile/Samip-Subedi/publication/319182991_Web_Virtual_Assistant_using_Semantic_Web_Technology_and_Web_Scraping/links/599928a50f7e9b3edb166f49/Web-Virtual-Assistant-using-Semantic-Web-Technology-and-Web-Scraping.pdf [accessed 2021-05-21]
19. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]

20. Cho JY, Lee E. Reducing confusion about grounded theory and qualitative content analysis: similarities and differences. Qual Rep 2014 Oct 15;19(32):1-20 [FREE Full text] [doi: 10.46743/2160-3715/2014.1028]
21. Kondracki NL, Wellman NS, Amundson DR. Content analysis: review of methods and their applications in nutrition education. J Nutr Educ Behav 2002;34(4):224-230. [doi: 10.1016/s1499-4046(06)60097-3] [Medline: 12217266]
22. Dedoose. URL: https://www.dedoose.com [accessed 2018-03-15]
23. Milward J, Khadjesari Z, Fincham-Campbell S, Deluca P, Watson R, Drummond C. User preferences for content, features, and style for an app to reduce harmful drinking in young adults: analysis of user feedback in app stores and focus group interviews. JMIR Mhealth Uhealth 2016 May 24;4(2):e47 [FREE Full text] [doi: 10.2196/mhealth.5242] [Medline: 27220371]
24. Iacob C, Harrison R. Retrieving and analyzing mobile apps feature requests from online reviews. In: Proceedings of the 10th Working Conference on Mining Software Repositories (MSR). 2013 Presented at: 10th Working Conference on Mining Software Repositories (MSR); May 18-19, 2013; San Francisco, CA, USA. [doi: 10.1109/MSR.2013.6624001]
25. Scheibe M, Reichelt J, Bellmann M, Kirch W. Acceptance factors of mobile apps for diabetes by patients aged 50 or older: a qualitative study. Med 2 0 2015 Mar 02;4(1):e1 [FREE Full text] [doi: 10.2196/med20.3912] [Medline: 25733033]
26. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
27. Frie K, Hartmann-Boyce J, Jebb S, Albury C, Nourse R, Aveyard P. Insights from Google Play Store user reviews for the development of weight loss apps: mixed-method analysis. JMIR Mhealth Uhealth 2017 Dec 22;5(12):e203 [FREE Full text] [doi: 10.2196/mhealth.8791] [Medline: 29273575]
28. Guerra-Reyes L, Christie VM, Prabhakar A, Harris AL, Siek KA. Postpartum health information seeking using mobile phones: experiences of low-income mothers. Matern Child Health J 2016 Nov;20(Suppl 1):13-21 [FREE Full text] [doi: 10.1007/s10995-016-2185-8] [Medline: 27639571]
29. Lupton D. 'It Just Gives Me a Bit of Peace of Mind': Australian women's use of digital media for pregnancy and early motherhood. Societies 2017 Sep 15;7(3):25. [doi: 10.3390/soc7030025]
30. Sefferman A. Mobile Ratings: The good, the bad, and the ugly. Apptentive. 2016 Jun 23. URL: https://www.apptentive.com/blog/2016/06/23/mobile-ratings-good-bad-ugly/ [accessed 2021-06-01]
31. Pagano D, Maalej W. User feedback in the appstore: an empirical study. In: Proceedings of the 21st IEEE International Requirements Engineering Conference (RE). 2013 Presented at: 21st IEEE International Requirements Engineering Conference (RE); July 15-19, 2013; Rio do Janeiro, Brazil. [doi: 10.1109/RE.2013.6636712]
32. Hill S. Can you really trust app store ratings? We asked the experts. Digital Trends. 2018. URL: https://www.digitaltrends.com/android/can-you-really-trust-app-store-ratings/ [accessed 2018-05-15]

Edited by R Kukafka; submitted 23.01.21; peer-reviewed by G Hu, H Luo; comments to author 08.03.21; revised version received 05.06.21; accepted 05.07.21; published 01.09.21.

Please cite as:
Biviji R, Williams KS, Vest JR, Dixon BE, Cullen T, Harle CA
Consumer Perspectives on Maternal and Infant Health Apps: Qualitative Content Analysis
J Med Internet Res 2021;23(9):e27403
URL: https://www.jmir.org/2021/9/e27403
doi:10.2196/27403
PMID:34468323

Original Paper

Implementing mHealth-Enabled Integrated Care for Complex Chronic Patients With Osteoarthritis Undergoing Primary Hip or Knee Arthroplasty: Prospective, Two-Arm, Parallel Trial

Jordi Colomina[1], MD; Reis Drudis[2], MD; Montserrat Torra[2], MD; Francesc Pallisó[1], PhD; Mireia Massip[3], BScN; Eloisa Vargiu[4], PhD; Nuria Nadal[5], MD; Araceli Fuentes[6], MD; Marta Ortega Bravo[7,8,9], MD; Felip Miralles[4], PhD; Ferran Barbé[3,10], MD; Gerard Torres[3,10], MD; Jordi de Batlle[3,10], PhD; The CONNECARE-Lleida Group[11]

[1]Servei de Cirurgia Ortopèdica i Traumatologia, Hospital Universitari de Santa Maria de Lleida, Universitat de Lleida, Lleida, Spain

[2]Servei Anestesiologia Reanimació i Clínica del Dolor, Hospital Universitari de Santa Maria de Lleida, Universitat de Lleida, Lleida, Spain

[3]Group of Translational Research in Respiratory Medicine, Institut de Recerca Biomedica de Lleida (IRBLleida), Lleida, Spain

[4]eHealth Unit, Eurecat Centre Tecnòlogic de Catalunya, Barcelona, Spain

[5]Gerència Territorial de Barcelona, Institut Català de la Salut, Barcelona, Spain

[6]Atenció Primària Àmbit Lleida, Institut Català de la Salut, Lleida, Spain

[7]Research Support Unit Lleida, Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Lleida, Spain

[8]Centre d'Atenció Primària Cappont, Gerència Territorial de Lleida, Institut Català de la Salut, Barcelona, Spain

[9]Universitat de Lleida, Lleida, Spain

[10]Center for Biomedical Network Research in Respiratory Diseases (CIBERES), Madrid, Spain

[11]see Authors' Contributions

Corresponding Author:
Jordi de Batlle, PhD
Group of Translational Research in Respiratory Medicine
Institut de Recerca Biomedica de Lleida (IRBLleida)
Rovira Roure 80
Lleida, 25198
Spain
Phone: 34 645624734
Email: jordidebatlle@gmail.com

Abstract

Background: Osteoarthritis is a disabling condition that is often associated with other comorbidities. Total hip or knee arthroplasty is an effective surgical treatment for osteoarthritis when indicated, but comorbidities can impair their results by increasing complications and social and economic costs. Integrated care (IC) models supported by eHealth can increase efficiency through defragmentation of care and promote patient-centeredness.

Objective: This study aims to assess the effectiveness and cost-effectiveness of implementing a mobile health (mHealth)–enabled IC model for complex chronic patients undergoing primary total hip or knee arthroplasty.

Methods: As part of the Horizon 2020 Personalized Connected Care for Complex Chronic Patients (CONNECARE) project, a prospective, pragmatic, two-arm, parallel implementation trial was conducted in the rural region of Lleida, Catalonia, Spain. For 3 months, complex chronic patients undergoing total hip or knee arthroplasty and their caregivers received the combined benefits of the CONNECARE organizational IC model and the eHealth platform supporting it, consisting of a patient self-management app, a set of integrated sensors, and a web-based platform connecting professionals from different settings, or usual care (UC). We assessed changes in health status (12-item short-form survey [SF-12]), unplanned visits and admissions during a 6-month follow-up, and the incremental cost-effectiveness ratio.

Results: A total of 29 patients were recruited for the mHealth-enabled IC arm, and 30 patients were recruited for the UC arm. Both groups were statistically comparable for baseline characteristics, such as age; sex; type of arthroplasty; and Charlson index, American Society of Anesthesiologists classification, Barthel index, Hospital Anxiety and Depression scale, Western Ontario and McMaster Universities Osteoarthritis Index, and Pfeiffer mental status questionnaire scores. Patients in both groups had

significant increases in the SF-12 physical domain and total SF-12 score, but differences in differences between the groups were not statistically significant. IC patients had 50% fewer unplanned visits (P=.006). Only 1 hospital admission was recorded during the follow-up (UC arm). The IC program generated savings in different cost scenarios, and the incremental cost-effectiveness ratio demonstrated cost-effectiveness.

Conclusions: Chronic patients undergoing hip or knee arthroplasty can benefit from the implementation of patient-centered mHealth-enabled IC models aimed at empowering patients and facilitating transitions from specialized hospital care to primary care. Such models can reduce unplanned contacts with the health system and reduce overall health costs, proving to be cost-effective. Overall, our findings support the notion of system-wide cross-organizational care pathways supported by mHealth as a successful way to implement IC for patients undergoing elective surgery.

(J Med Internet Res 2021;23(9):e28320) doi:10.2196/28320

KEYWORDS

mHealth; osteoarthritis; arthroplasty; health plan implementation; chronic disease; mobile phone

Introduction

The progressive aging of populations has led to an increased burden of chronic diseases [1]. Osteoarthritis (OA) is one of the most disabling chronic diseases in developed countries. Worldwide estimates show that 10% of men and 18% of women aged >60 years have symptomatic OA [2]. Knees and hip are the most affected locations. Among the people with OA, 80% have movement limitations and 25% are disabled to perform major activities of daily living [2]. OA is associated with increased comorbidities and mortality. More than half of people with OA have another chronic medical condition, and approximately one-third of people have five or more chronic conditions [3].

Total hip arthroplasty (THA) and total knee arthroplasty (TKA) are effective surgical treatments for end-stage OA, improving joint function and health-related quality of life (QoL) [4]. Since 2000, the number of hip and knee replacements has increased rapidly in most countries in the Organization for Economic Co-operation and Development [5]. On average, hip replacement rates increased by 30% between 2007 and 2017, and knee replacement rates increased by 40% [5]. Comorbidities are independent predictors of increased postoperative complications and nonhomebound discharge in patients undergoing shoulder, hip, or knee arthroplasty [6] and are associated with physical function and pain after primary TKA [6].

So far, the strategies aiming to improve the outcomes of elective surgeries have mainly focused on enhanced recovery protocols, prehabilitation, and postoperative rehabilitation protocols, which have been proven effective for lower limb arthroplasty [7,8]. Nevertheless, there are aspects of the entire care process that can still be improved. Traditional care models suffer from care fragmentation, with the different care levels failing to communicate effectively. After transitioning from hospital care to primary care, patients can have changes in prescribed drugs or repetition of tests because of a lack of communication across health settings [9]. Moreover, traditional models still focus on diseases rather than patients, which makes patients and their caregivers passive actors in the process [9]. Therefore, there is a need for a profound redesign of how care is provided to chronic older patients to ensure quality and sustainability [10]. Integrated care (IC) models aim to increase health efficiency through defragmentation of care, including the promotion of collaboration across settings, promotion of patient-centeredness, and prioritization of preventive strategies [11]. Interestingly, the use of eHealth tools could be the key to enabling such models [12], as demonstrated by projects such as Personalized Connected Care for Complex Chronic Patients (CONNECARE) [13] or BeyondSilos [14], which have proven the feasibility of eHealth-enabled IC in the Catalan setting. The adoption of efficient IC models can result in better outcomes for THA and TKA. In this regard, a 2017 systematic review on the benefits of telerehabilitation after orthopedic surgery showed promising results for patients undergoing THA or TKA [15]. More recently, Jonker et al [16] confirmed the feasibility of perioperative eHealth interventions for the management of older surgical patients. However, there are very few initiatives that fully embrace the use of eHealth-enabled IC models in older patients undergoing THA or TKA.

The CONNECARE project is a European Union Horizon 2020 Research and Innovation project aiming to co-design, develop, deploy, and evaluate a novel smart and adaptive organizational IC model for complex chronic patients (CCPs) [17]. From 2016 to 2019, the project co-designed an organizational model for IC and an eHealth platform supporting it by means of an iterative patient-centered process involving patients and stakeholders across different health settings. The resulting IC model promoted collaboration among professionals in different care settings (hospital specialists, family physicians, and social workers), prioritizing home-based prevention strategies over institutional reactive care and fostering patient empowerment. A Smart Adaptive Case Management (SACM) system is offered as a web-based platform to professionals from different settings, and a patient-empowering mobile health (mHealth) self-management system allowing for three-level monitoring features and advanced communication is offered to patients.

As part of the CONNECARE project, a novel mHealth-enabled IC model was implemented in Lleida, Spain, targeting older CCPs undergoing elective THA or TKA. The existing care model for THA and TKA in Lleida is an enhanced recovery after surgery (ERAS) pathway based on different interventions to reduce perioperative stress; maintain and support homeostasis and physiological function; and accelerate the achievement of discharge criteria, including minimizing complications and readmission [18,19]. Although it produces good results with low transfusion, low complications, and decreased length of

stay, it is limited by the scarce communication between professionals in different care settings, mostly hospital and primary care professionals, with different electronic medical records (EMRs) systems in hospitals and primary care centers (Argos SAP and ECAP [20], respectively). Therefore, we hypothesized that the new mHealth-enabled IC model could result in better outcomes for our patients.

This paper describes the results in terms of effectiveness and cost-effectiveness of the implementation of an mHealth-enabled IC model for the prevention of hospital readmissions in CCPs undergoing THA and TKA.

Methods

Study Design

This was a prospective, pragmatic, two-arm, parallel implementation trial comparing usual care (UC) with a 3-month mHealth-enabled IC intervention. The study was conducted from July 2018 to August 2019 at the University Hospital of Santa Maria (Lleida, Spain) and its network of primary care centers. This corresponds to a large rural area accounting for more than 236,000 citizens with a life expectancy of 80 and 86 years for men and women, respectively [21].

Target Population

The eligibility criteria were home-dwelling patients elected for primary THA or TKA at the University Hospital of Santa Maria; aged >65 years; being defined as CCP (Charlson index score ≥3, taking four or more pills per day, and having had contact with the health system at least two times in the last 12 months); being classified according to the American Society of Anesthesiologists (ASA) classification as ASA II (mild systemic disease) or ASA III (severe systemic disease); and successfully passing a basic technological test, aimed to ensure the availability of internet connection at home as well as patients' or caregivers' competence with the use of a smartphone, tablet, or computer. The basic technological test can be found in Multimedia Appendix 1 [20,22].

Recruitment

The recruitment was done in several waves to match the pace of the CONNECARE project technological developments. In each wave, consecutive potential participants were contacted by a case manager during preoperative assessment at the anesthesiology outpatient clinic. The case manager explained the study protocol and obtained informed consent. These patients formed the intervention arm. After the recruitment of each patient included in the intervention arm, an active search for a control with similar characteristics from the surgery waitlist of the Orthopedics Department of University Hospital of Santa Maria began. This enhanced the similarity of patients in the intervention and control arms, although it implied a certain lag in the recruitment of controls (from some days to few weeks). All patients and their caregivers, regardless of study arm, received a face-to-face explanation about the study and provided informed consent.

Intervention

Patients in the intervention arm were attended using an IC model, including (1) preliminary assessment of the patient's health status using several questionnaires, tests, and indices specific to their main chronic diseases and social needs; (2) a self-management app, with status and performance reports, a virtual coach with customizable automated feedback, and full communication with the care team; (3) a Fitbit Flex 2 digital activity tracker [23], fully integrated into the self-management app; (4) a patient's profile in the SACM web-based platform, accessible to all members of the care team (family physicians, hospital surgical team, and social workers) and used to coordinate professionals in the different settings and enable a communication channel among them and with the patient, when needed; and (5) assignment of a case manager in charge of supervising the whole process and being the main patient contact point. Additional details on the CONNECARE IC model and the supporting eHealth platform can be found in Multimedia Appendix 1. Patients in the control arm were attended using care as usual, being managed from primary care after hospital discharge. After discharge from the initial 90 days of UC or IC management, all patients had 3 months of additional passive follow-up.

Data Collection

Variables characterizing the patients were collected at recruitment using the SACM on tablet and/or desktop computers. Collected variables included main baseline characteristics, such as age, sex, main chronic diseases, Charlson index of comorbidities [24], ASA scores [25], Barthel index for activities of daily living [26], Hospital Anxiety and Depression scale [27], assessment of the dwelling characteristics, main medications, Pfeiffer mental status questionnaire [28], and tobacco and alcohol consumption.

The cost estimation of the IC program and used health care resources is described in Multimedia Appendix 1. Briefly, IC costs included the costs of newly required medical personnel (hospital-based nurse case managers) and the costs of licensing and running the IC platform. In contrast, the one-off costs associated with the development of the IC model and supporting technology or the tuning of the system were not considered. Similarly, the required restructuring of health care professionals' time to include new tasks related to the IC model was assumed to be covered by the health system at no additional cost, as we assumed that such restructuring would not imply any additional time. Costs associated with the use of health care resources, such as medical visits and hospital admissions, were based on official data from the Catalan Health Department [22]. All costs were originally in Euros (€) and were converted into US $ (conversion used: €1=US $1.21).

The main outcomes were (1) intervention effectiveness, as measured by the changes in the 12-item short-form survey (SF-12) health questionnaire's physical and mental domains (baseline vs discharge) [29]; (2) use of health care resources after 6 months; and (3) cost-effectiveness, based on the improvement in QoL relative to costs, assessed by means of the incremental cost-effectiveness ratio (ICER). The source of health information was the EMR, and the collected information

included hospital admissions, emergency room visits, visits to primary care, and visits to hospital specialists.

Statistical Analyses

Participants' baseline characteristics were described by n (%), mean (SD), or median (P25-P75), as appropriate. Comparisons between IC and control patients' baseline characteristics were performed using the chi-square test, *t* test, or Kruskal-Wallis test, as appropriate. A paired *t* test comparing baseline with discharge was used to assess changes in the SF-12 domains. Linear regression models were used to assess the differences in the changes experienced by IC and control patients. Negative binomial regression models were used to assess the differences in the number of visits and admissions. Models were adjusted for age, sex, and Charlson score. ICER was calculated in relation to the SF-12 total score. Sensitivity analyses assuming two different scenarios, 150% and 200% estimated cost of the IC program, were performed to explore the cost-effectiveness performance of the IC model under unplanned increases in the implementation costs. Data analyses were conducted using Stata version 12.1 (StataCorp). The threshold for significance was set at a *P* value of .05. All *t* tests were two-tailed.

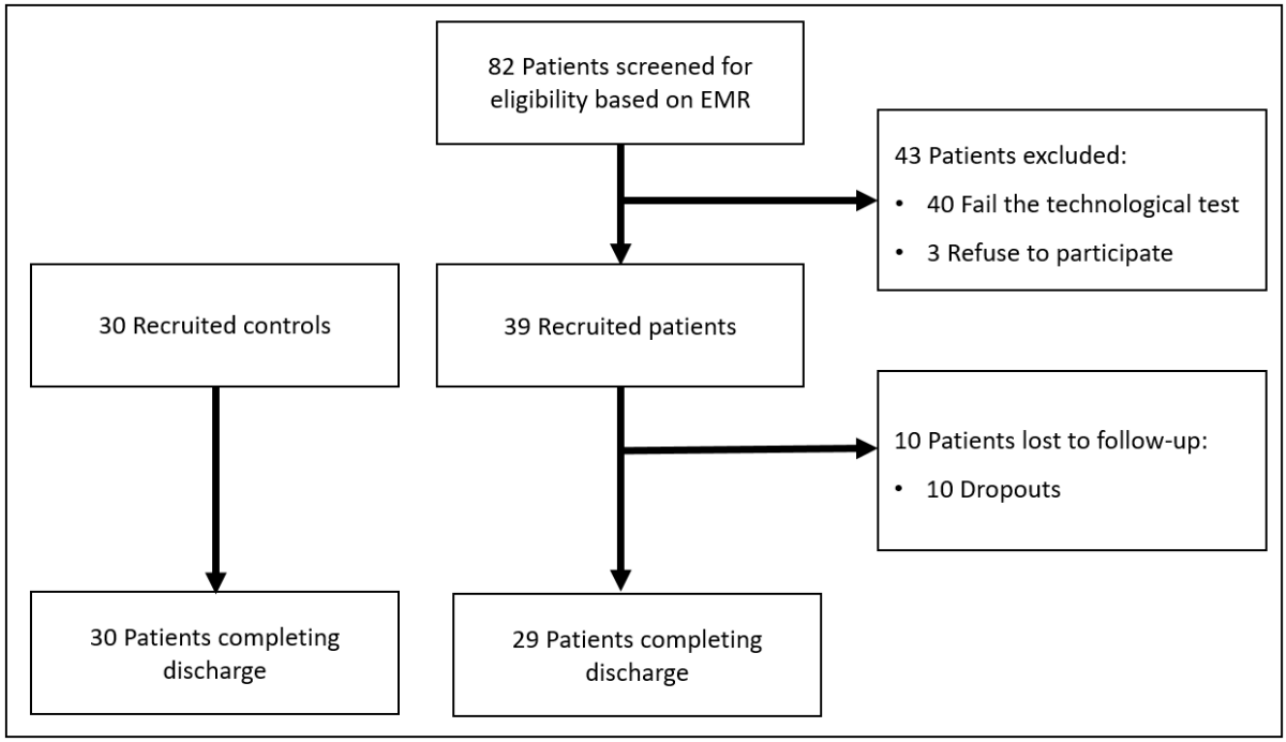
Ethical Considerations

This study was approved by the Ethics Committee of Hospital Arnau de Vilanova (CEIC-1685), and all patients provided written informed consent. All collected data were handled and stored in accordance with the current national and international legislation.

Results

Up to 82 patients were screened for eligibility. Of them, 49% (40/82) failed the technological test because they did not have an internet connection and 4% (3/82) refused to participate. Therefore, 39 patients were recruited for the mHealth-enabled IC arm and 30 for the UC arm. Final analyses were based on 29 IC and 30 control patients who completed the follow-up (Figure 1).

Figure 1. Study flowchart. EMR: electronic medical record.



The baseline characteristics of the patients are presented in Table 1. Although the proportion of knee surgeries was slightly higher in the IC arm, no significant differences were found in the baseline characteristics of the patients in the two arms.

Table 1. Baseline characteristics of the patients in the usual care and integrated care arms (N=59).

Characteristic	Usual care (n=30)	Integrated care (n=29)	*P* value[a]
Sex (male), n (%)	8 (27)	12 (41)	.23
Age (years), mean (SD)	74 (8)	72 (9)	.46
Charlson score, mean (SD)	4.3 (1.7)	4.2 (1.5)	.81
American Society of Anesthesiologists class, n (%)			.52
II	10 (33)	12 (41)	
III	20 (67)	17 (59)	
Western Ontario and McMaster Universities Osteoarthritis Index score, mean (SD)			
Pain	9.5 (3.4)	9.6 (3.8)	.93
Function	38.7 (12.8)	39.8 (14.2)	.77
Stiffness	2.3 (2.1)	3.0 (2.8)	.29
Barthel score, median (IQR)	95 (90-100)	100 (95-100)	.16
HAD[b] scale anxiety score, mean (SD)	6.2 (4.9)	5.0 (3.9)	.34
HAD scale depression score, mean (SD)	5.0 (2.3)	5.1 (2.9)	.88
Pfeiffer intact intellectual functioning, n (%)	30 (100)	27 (93)	.14
Surgery location: knee, n (%)	22 (73)	26 (90)	.11

[a]Chi-square test, *t* test, or Kruskal-Wallis equality-of-populations rank test, as appropriate.

[b]HAD: Hospital Anxiety and Depression.

Table 2 shows the changes in QoL (SF-12 domains) from baseline to discharge. Regardless of the study arm, all patients showed a substantial increase in the SF-12 physical domain. The analysis of the differences in differences in QoL between IC and control patients showed no differences. Stratification according to the surgical location (hip or knee) reported similar findings.

Table 2. Changes in health status in the usual care and integrated care arms.

12-item short-form survey score	Baseline, mean (SD)	Discharge, mean (SD)	Change, mean (SD)	*P* value
Physical				
UC[a]	27.9 (6.4)	42.0 (7.7)	14.1 (9.0)	<.001[b]
IC[c]	29.9 (10.0)	45.3 (9.8)	15.4 (11.7)	<.001[b]
Difference	2.0 (2.2)	3.3 (2.2)	1.4 (3.2)	.79[d]
Mental				
UC	48.1 (14.2)	50.2 (13.5)	2.0 (11.9)	.35[b]
IC	52.1 (14.2)	52.8 (12.9)	0.8 (15.2)	.79[b]
Difference	3.9 (3.6)	2.6 (3.6)	−1.3 (5.1)	.94[d]
Total				
UC	76.1 (15.6)	92.2 (18.1)	16.1 (14.8)	<.001[b]
IC	81.9 (18.8)	98.1 (15.6)	16.2 (14.3)	<.001[b]
Difference	5.9 (4.5)	6.0 (4.5)	0.1 (6.3)	.94[d]

[a]UC: usual care.

[b]Paired *t* test comparing baseline and discharge measures.

[c]IC: integrated care.

[d]Linear regression predicting the difference in the changes experienced by each arm, adjusted by age, sex, and Charlson index.

Table 3 shows that IC patients had 50% fewer unplanned visits than patients in the UC arm, either when considering visits

directly related to the surgical procedure or any visits. One patient in the UC arm required hospital admission, whereas no admissions were recorded among IC arm patients.

Table 3. Total use of health services during the follow-up period (N=59).

Health service	Usual care (n=30), mean (SD)	Integrated care (n=29), mean (SD)	*P* value[a]	Adjusted *P* value[b]
All unplanned visits	1.4 (1.5)	0.7 (1.0)	.03	.006
Unplanned visits directly related to the surgery procedure	0.8 (1.2)	0.4 (0.7)	.10	.02
All hospital admissions	0.03 (0.2)	0 (0)	N/A[c]	N/A
Hospital admissions directly related to the surgery procedure	0.03 (0.2)	0 (0)	N/A	N/A

[a]Negative binomial regression model.

[b]Negative binomial regression model adjusted for age, sex, and Charlson comorbidity index.

[c]N/A: not applicable.

The analyses of within-trial costs and cost-effectiveness for all unplanned visits and hospital admissions are summarized in Table 4 and those for unplanned visits and hospital admissions related to the arthroplasty procedure are summarized in Table 5. The IC program generated savings from €109.88 (US $132.96) to €126.99 (US $153.66) per patient, depending on the nature of unplanned visits and hospitalizations. The IC program was cost-effective according to ICER, performing similar in terms of QoL gain while reducing overall expenses because of the reduction of unplanned visits and hospital admissions.

Table 4. Changes in health-related quality of life, within-trial costs (average cost per patient), and cost-effectiveness considering all unplanned visits and hospital admissions.

Variables	Usual care (n=30)	Integrated care (n=29)	Difference	ICER[a]
Changes in the 12-item short-form survey score, mean (SD)	16.1 (14.8)	16.2 (14.3)	0.1 (6.3)	N/A[b]
Unplanned visits costs[c] (US $)	106.07	51.74	−54.33	N/A
Hospital admissions costs[c] (US $)	185.25	0	−185.25	N/A
Total medical costs per patient (US $)	291.32	51.74	−239.58	N/A
Personalized Connected Care for Complex Chronic Patients program cost (US $)	0	85.92	85.92	N/A
Total costs per patient (US $)	291.32	137.66	−153.66	−1920.73

[a]ICER: incremental cost-effectiveness ratio; incremental cost associated with one additional point gain in the 12-item short-form survey.

[b]N/A: not applicable.

[c]Costs based on the Catalan Institute of Health official pricing.

Table 5. Changes in health-related quality of life, within-trial costs (average cost per patient), and cost-effectiveness considering unplanned visits and hospital admissions related to the surgical intervention.

Variables	Usual care (n=30)	Integrated care (n=29)	Difference	ICER[a]
Changes in the 12-item short-form survey score, mean (SD)	16.1 (14.8)	16.2 (14.3)	0.1 (6.3)	N/A[b]
Unplanned visits costs[c] (US $)	64.68	31.05	−33.63	N/A
Hospital admissions costs[c] (US $)	185.25	0	−185.25	N/A
Total medical costs per patient (US $)	249.93	31.05	−218.88	N/A
Personalized Connected Care for Complex Chronic Patients program cost (US $)	0	85.92	85.92	N/A
Total costs per patient (US $)	249.93	116.97	−132.96	−1661.94

[a]ICER: incremental cost-effectiveness ratio; incremental cost associated with one additional point gain in the 12-item short-form survey.

[b]N/A: not applicable.

[c]Costs based on the Catalan Institute of Health official pricing.

Sensitivity analyses assuming two different cost scenarios, 150% and 200% estimated cost of the IC program, thus exploring cost-effectiveness under unplanned increases in the implementation costs, showed savings and cost-effectiveness, as shown in Tables S1 and S2 in Multimedia Appendix 1.

Discussion

Principal Findings

The prospective assessment of the implementation of an mHealth-enabled IC program for TKA and THA management showed a reduction in the number of unplanned contacts with the health system after the surgery; generated substantial savings for the health system, while not having any negative impact on QoL or clinical outcomes; and demonstrated cost-effectiveness.

Strengths and Limitations

The implemented IC model had several strengths that must be highlighted. First, there was an effort to involve all the stakeholders from different organizations that would be actors in a large-scale deployment of the mHealth-enabled IC program since the very early stages. This is key, as the lack of cooperation among professionals, teams, and organizations is a recurrent barrier for effective IC implementation [11]. Second, informal caregivers played a role in the IC process, as close relatives of the patients can be their link with the health system. Moreover, informal caregivers can play a key role in facilitating the use of self-management apps in older patients. Third, the self-management app for patients (allowing active monitoring, communication with the care team, and a virtual coach with customizable automated feedback) enhanced doctor-patient relationships [30] and facilitated early detection of any problem in the surgical recovery process [31]. Fourth, the assessment and promotion of patients' physical activity is an effective way of preventing any mobility impairment, which is found in one-third of people aged >65 years [32]. Finally, the region selected for the deployment of the IC program, a large rural area, was especially appropriate as their citizens can benefit the most from community-based IC initiatives that can avoid unnecessary travel to the hospital.

Regarding this study, several strengths and limitations should be noted. Among the strengths, we note the use of a prospective study design with a comparator arm; the use of objectively measured endpoints, such as visits and admissions, in contrast to patient-reported outcomes; and cost and cost-effectiveness assessments. Concerning the limitations, the technological platform supporting the implemented IC model showed substantial improvements throughout the implementation period. This implied that patients recruited near the end of the implementation study had a richer IC experience than those recruited at the very beginning. Similarly, this had an impact on health care professionals, who had to cope with a platform under development and not fully integrated with existing EMRs. Nevertheless, participating patients and professionals showed great acceptability of the IC model and setting [33]. Moreover, directly participating in a dynamic development and implementation process fostered professionals' engagement and allowed them to propose changes and new features to be developed, which resulted in not a single professional dropping out of the study. Next, the limited number of patients involved in the study had an impact on statistical power. Nevertheless, the current number of participants sufficed to show a significant reduction in robust endpoints, such as unplanned visits and hospital admissions, and showed the cost-effectiveness of the IC program. Finally, only direct costs were considered, although the inclusion of indirect or societal costs would most likely favor the cost-effectiveness of the IC model.

Comparison With Existing Literature

This study aimed to assess the impact of the implementation of an IC model in three domains: (1) patients' QoL, (2) patients' use of health services, and (3) health economics. Regarding the QoL domain, the IC model performed as good as the UC arm in the differences-in-differences analysis. This result is in line with the mixed results found in a 2017 review on the impact of IC interventions on QoL [34]. However, it is worth noting that the great increase in QoL obtained after a successful THA or TKA is likely to mask any minor increase in QoL caused by being managed in an IC model. Next, regarding the use of health services, patients in the IC model required 50% fewer unplanned visits after the surgery. This corresponds to the upper margin of benefits reported in reviews about IC interventions between 2000 and 2015, which reported significant reductions in hospital activity ranging from 15% to 50% [35]. These excellent results could be in line with the notion of system-wide cross-organizational care pathways as a successful way to implement IC, in contrast to smaller and narrow interventions [9]. Finally, regarding health economics, IC generated savings from €109.88 (US $132.96) to €126.99 (US $153.66) per patient and was deemed cost-effective. This is in line with reviews stating the potential cost-effectiveness of IC in the management of chronic diseases [36]. However, these savings are lower than the range of US $584-$1434 obtained when applying the same IC model to CCPs with chronic obstructive pulmonary disease or heart failure [13], who are more prone to experiencing unplanned visits and hospitalizations.

When specifically focusing on THA and TKA, previous studies have suggested the usefulness of the different potential components of an eHealth IC model, including telerehabilitation [15], care pathways [18,37], education [38], patient-centeredness [39], mHealth continuous monitoring [40,41], and cross-setting integration [42]. However, to our knowledge, this study is the first to include all these components in a single mHealth-enabled IC model. Overall, our results are similar to or better than those reported in these previous studies, suggesting a moderate additive effect of combining the reported interventions into a single mHealth-enabled IC model.

Implications for Research and/or Practice

The World Health Organization has already stated the need for patient-centered IC models to satisfy the health needs of older populations with chronic diseases while keeping costs sustainable [29,43]. With this premise in mind, the CONNECARE project tackled the task of iteratively co-designing a mHealth-enabled IC model with the participation of all key stakeholders: patients; hospital-based surgeons, anesthesiologists, nurses, physiotherapists, and case managers; primary care physicians and nurses; social caregivers; and

managers, technical staff, developers, and researchers. This multidisciplinary team envisioned a system-wide cross-organizational patient-centered care pathway, in line with the principles of the 2015 World Report on Ageing and Health [1]. The experience acquired during the co-design and testing process allows us to highlight some key features that future IC models for the management of older citizens undergoing THA or TKA should consider (1) a common cross-setting web-based platform is key for a successful coordination of care across settings and patient monitoring; (2) the habilitation of functional communication channels for the patients can be a key source of savings, as most savings are achieved through the avoidance of unplanned visits; and (3) involving informal caregivers, such as younger family members, can facilitate the adoption of mHealth tools, such as sensors and self-management apps, and make the overall user experience very satisfactory [44]. Moreover, it is key to ensure access to an internet connection at home, as this was the main criterion halting the participation of potential users.

Conclusions

The implementation of a patient-centered mHealth-enabled IC model for the management of patients undergoing THA or TKA successfully empowered patients, effectively connected the different care settings involved, reduced unplanned contacts with the health system, reduced health costs, and was cost-effective. This supports the use of mHealth tools for the implementation of system-wide cross-organizational IC models.

Acknowledgments

This work was supported by the European Union's Horizon 2020 Research and Innovation Program (under grant agreement GA-689802). JDB acknowledges receiving financial support from the Catalan Health Department (Pla Estratègic de Recerca i Innovació en Salut 2016: SLT002/16/00364) and Instituto de Salud Carlos III (Miguel Servet 2019: CP19/00108), which is cofunded by the European Social Fund, *Investing in your future*. Funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. The authors would like to express their sincere gratitude to all the participating patients and their caregivers as well as to all the involved health and social care professionals. All of them were key in the co-design of the integrated care model and its implementation in the region of Lleida.

Authors' Contributions

JDB, EV, NN, FP, FM, FB, and GT participated in the conceptualization of the project. JC, RD, MT, MM, EV, AF, MOB, and GT conducted the data collection. MM, EV, and JDB participated in the data curation. JDB conducted all statistical analyses. JC and JDB wrote the original draft of the manuscript. All authors reviewed the final manuscript. JDB, EV, FM, FB, and GT secured funding for the project. The CONNECARE-Lleida Group consists of: Maria Aguilà Balastegui, Sandra Alexandre Loxano, Laila Al-Jouja Llorente, Tomás Alonso Sancho, Enrique Aparicio Bañeres, Ana Arce Vila, Jose Maria Baron Burriel, Ramon Bascompte Claret, Albert Bigorda Sague, Emilia Blanco Ponce, Maria Boldú Franque, Àngels Bosch Roig, Carmen Bravo Santiago, Alba Capdevila Sarramona, Aida Castelló Corretge, Montse Coma Gassó, Fina Cregenzan Ortiz, Dolors Del Pozo Garcia, Mireia Falguera Vilamajó, Pere Farre Pagés, Yolanda Fauria Garcia, Anabel Fusalba Canales, Jara Gayan Ordas, Sergi Godia Lopez, Irene Gomez Companys, Jessica Gonzàlez Gutierrez, Anna Gort Oromí, Carme Jorge Tufet, Mercé Lavega Llorens, Laia Llort Samsó, Maria Rosa Lopez Cervelló, Belen Malla Clua, Josep Maria Marsol Mas, Teresita Martí Ribes, Diana Martin Capella, José Maria Martínez Barriuso, Esther Mateus Solé, Ramon Mazana Novellon, Petra Merino De los Santos, Miquel Mesas Julio, Sonia Minguet Vidal, Nuria Moles Porta, Luis Miguel Montaña Esteban, Dolors Morera Roset, Meritxell Moyà Oro, Irene Muñoz Del Campo, Francisco Nicolás Sánchez, Inés Ortiz Catalán, Mireia Ortiz Valls, Sonia Ortiz Congost, Jose Maria Palacin Peruga, Eugeni Paredes Costa, Pablo Pastor Pueyo, Ana Pérez Sainz, Antonio Plana Blanco, Anna Planas Hiraldo, Pepita Pont Aldoma, Marife Quelle Alonso, Rebeca Ramirez Molinero, Maria Àngels Revés Juanbaro, Anna Ribé Miró, Eva Ribó Caubet, Rebeca Rodriguez Corbaton, Marina Rué Florensa, Oscar Sacristán García, Irene Sanmartí Forns, Maria Cruz Sanz Martinez, Neus Sendra Bordes, Maria Cruz Urgelés Castillón, Laia Utrillo Montagut, and Montse Vidal Ballesté.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary information.
[DOCX File , 1565 KB - jmir_v23i9e28320_app1.docx]

References

1. World Health Organization. World Report on Ageing and Health. Geneva: World Health Organization; 2015:1-246.
2. Safiri S, Kolahi AA, Smith E, Hill C, Bettampadi D, Mansournia MA, et al. Global, regional and national burden of osteoarthritis 1990-2017: a systematic analysis of the Global Burden of Disease Study 2017. Ann Rheum Dis 2020 Jun;79(6):819-828. [doi: 10.1136/annrheumdis-2019-216515] [Medline: 32398285]

3. Osteoarthritis: a serious disease, submitted to the U.S. Food and Drug Administration. Osteoarthritis Research Society International. 2016. URL: https://oarsi.org/sites/default/files/library/2018/pdf/oarsi_white_paper_oa_serious_disease121416_1.pdf [accessed 2021-02-13]
4. Ethgen O, Bruyère O, Richy F, Dardennes C, Reginster J. Health-related quality of life in total hip and total knee arthroplasty. A qualitative and systematic review of the literature. J Bone Joint Surg Am 2004 May;86(5):963-974. [doi: 10.2106/00004623-200405000-00012] [Medline: 15118039]
5. Organisation for Economic Cooperation and Development (OECD). Health at a Glance 2019: OECD Indicators. Paris: OECD Publishing; 2019:1-243.
6. Hilton ME, Gioe T, Noorbaloochi S, Singh JA. Increasing comorbidity is associated with worsening physical function and pain after primary total knee arthroplasty. BMC Musculoskelet Disord 2016 Oct 07;17(1):421 [FREE Full text] [doi: 10.1186/s12891-016-1261-y] [Medline: 27717340]
7. Saunders R, Seaman K, Ashford C, Sullivan T, McDowall J, Whitehead L, et al. An ehealth program for patients undergoing a total hip arthroplasty: protocol for a randomized controlled trial. JMIR Res Protoc 2018 Jun 12;7(6):e137 [FREE Full text] [doi: 10.2196/resprot.9654] [Medline: 29895515]
8. Nunns M, Shaw L, Briscoe S, Thompson CJ, Hemsley A, McGrath JS, et al. Multicomponent hospital-led interventions to reduce hospital stay for older adults following elective surgery: a systematic review. Health Serv Deli Res 2019;7(40):07400. [doi: 10.3310/hsdr07400] [Medline: 31869021]
9. Harvey G, Dollard J, Marshall A, Mittinty MM. Achieving integrated care for older people: shuffling the deckchairs or making the system watertight for the future? Int J Health Policy Manag 2018 Apr 01;7(4):290-293 [FREE Full text] [doi: 10.15171/ijhpm.2017.144] [Medline: 29626395]
10. World Health Organization. Integrated Care for Older People: Guidelines on Community-Level Interventions to Manage Declines in Intrinsic Capacity. Geneva: World Health Organization; 2017:1-46.
11. Dates M, Lennox-chhugani N, Sant H, Pereira A, Tedeschi M. Health System Performance Assessment – Integrated Care Assessment. Brussels, Belgium: European Commission; 2018:1-108.
12. Chouvarda IG, Goulis DG, Lambrinoudaki I, Maglaveras N. Connected health and integrated care: toward new models for chronic disease management. Maturitas 2015 Sep;82(1):22-27. [doi: 10.1016/j.maturitas.2015.03.015] [Medline: 25891502]
13. de Batlle J, Massip M, Vargiu E, Nadal N, Fuentes A, Bravo MO, CONNECARE-Lleida Group. Implementing mobile health-enabled integrated care for complex chronic patients: intervention effectiveness and cost-effectiveness study. JMIR Mhealth Uhealth 2021 Jan 14;9(1):e22135 [FREE Full text] [doi: 10.2196/22135] [Medline: 33443486]
14. Piera-Jiménez J, Daugbjerg S, Stafylas P, Meyer I, Müller S, Lewis L, et al. BeyondSilos, a telehealth-enhanced integrated care model in the domiciliary setting for older patients: observational prospective cohort study for effectiveness and cost-effectiveness assessments. JMIR Med Inform 2020 Oct 06;8(10):e20938 [FREE Full text] [doi: 10.2196/20938] [Medline: 33021490]
15. Pastora-Bernal JM, Martín-Valero R, Barón-López FJ, Estebanez-Pérez MJ. Evidence of benefit of telerehabitation after orthopedic surgery: a systematic review. J Med Internet Res 2017 Dec 28;19(4):e142 [FREE Full text] [doi: 10.2196/jmir.6836] [Medline: 28455277]
16. Jonker LT, Haveman ME, de Bock GH, van Leeuwen BL, Lahr MM. Feasibility of perioperative eHealth interventions for older surgical patients: a systematic review. J Am Med Dir Assoc 2020 Dec;21(12):1844-1851 [FREE Full text] [doi: 10.1016/j.jamda.2020.05.035] [Medline: 32694000]
17. The CONNECARE project. URL: http://www.connecare.eu [accessed 2021-02-13]
18. Wainwright TW. Consensus statement for perioperative care in total hip replacement and total knee replacement surgery: Enhanced Recovery After Surgery (ERAS®) Society recommendations. Acta Orthop 2020 Jun 14;91(3):363 [FREE Full text] [doi: 10.1080/17453674.2020.1724674] [Medline: 32056486]
19. Jones EL, Wainwright TW, Foster JD, Smith JR, Middleton RG, Francis NK. A systematic review of patient reported outcomes and patient experience in enhanced recovery after orthopaedic surgery. Ann R Coll Surg Engl 2014 Mar;96(2):89-94 [FREE Full text] [doi: 10.1308/003588414X13824511649571] [Medline: 24780662]
20. eCAP. Department of Health. URL: https://salutweb.gencat.cat/ca/ambits_actuacio/linies_dactuacio/tecnologies_informacio_i_comunicacio/ecap/ [accessed 2021-02-13]
21. CatSalut Memòria 2018 Regió Sanitària Lleida. Generalitat de Catalunya. URL: https://catsalut.gencat.cat/ca/coneix-catsalut/catsalut-territori/lleida/publicacions/#bloc1 [accessed 2021-02-13]
22. Resolution SLT/353/2013, of February 13, on the review of public prices for the services provided by the Catalan Health Institute. CVE-DOGC-A-13051031-2013. URL: https://portaldogc.gencat.cat/utilsEADOP/PDF/6326/1287494.pdf [accessed 2021-02-13]
23. Fitbit. URL: https://www.fitbit.com/pr/flex2 [accessed 2021-02-13]
24. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [doi: 10.1016/0021-9681(87)90171-8] [Medline: 3558716]
25. Mayhew D, Mendonca V, Murthy BV. A review of ASA physical status - historical perspectives and modern developments. Anaesthesia 2019 Mar;74(3):373-379 [FREE Full text] [doi: 10.1111/anae.14569] [Medline: 30648259]

26. Collin C, Wade DT, Davies S, Horne V. The Barthel ADL Index: a reliability study. Int Disabil Stud 1988;10(2):61-63. [doi: 10.3109/09638288809164103] [Medline: 3403500]
27. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
28. Pfeiffer E. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. J Am Geriatr Soc 1975 Oct;23(10):433-441. [doi: 10.1111/j.1532-5415.1975.tb00927.x] [Medline: 1159263]
29. Ware J, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
30. Józsa T. The relevance of mobile applications helping in doctor-patient relationships. Br J Gen Pract 2020 Mar;70(692):109 [FREE Full text] [doi: 10.3399/bjgp20X708377] [Medline: 32107220]
31. Aznar-Tortonda V, Palazón-Bru A, la Rosa DM, Espínola-Morel V, Pérez-Pérez BF, León-Ruiz AB, et al. Detection of frailty in older patients using a mobile app: cross-sectional observational study in primary care. Br J Gen Pract 2020 Jan;70(690):29-35 [FREE Full text] [doi: 10.3399/bjgp19X706577] [Medline: 31685541]
32. Mitra S, Sambamoorthi U. Disability prevalence among adults: estimates for 54 countries and progress toward a global estimate. Disabil Rehabil 2014;36(11):940-947. [doi: 10.3109/09638288.2013.825333] [Medline: 23962193]
33. de Batlle J, Massip M, Vargiu E, Nadal N, Fuentes A, Bravo MO, CONNECARE-Lleida Group. Implementing mobile health-enabled integrated care for complex chronic patients: patients and professionals' acceptability study. JMIR Mhealth Uhealth 2020 Nov 20;8(11):e22136 [FREE Full text] [doi: 10.2196/22136] [Medline: 33216004]
34. Flanagan S, Damery S, Combes G. The effectiveness of integrated care interventions in improving patient quality of life (QoL) for patients with chronic conditions. An overview of the systematic review evidence. Health Qual Life Outcomes 2017 Sep 29;15(1):188 [FREE Full text] [doi: 10.1186/s12955-017-0765-y] [Medline: 28962570]
35. Damery S, Flanagan S, Combes G. Does integrated care reduce hospital activity for patients with chronic diseases? An umbrella review of systematic reviews. BMJ Open 2016 Dec 21;6(11):e011952 [FREE Full text] [doi: 10.1136/bmjopen-2016-011952] [Medline: 27872113]
36. Cronin J, Murphy A, Savage E. Can chronic disease be managed through integrated care cost-effectively? Evidence from a systematic review. Ir J Med Sci 2017 Nov;186(4):827-834. [doi: 10.1007/s11845-017-1600-5] [Medline: 28477328]
37. Mertes SC, Raut S, Khanduja V. Integrated care pathways in lower-limb arthroplasty: are they effective in reducing length of hospital stay? Int Orthop 2013 Jun;37(6):1157-1163 [FREE Full text] [doi: 10.1007/s00264-013-1829-1] [Medline: 23456085]
38. McDonald S, Page MJ, Beringer K, Wasiak J, Sprowson A. Preoperative education for hip or knee replacement. Cochrane Database Syst Rev 2014 May 13(5):CD003526. [doi: 10.1002/14651858.CD003526.pub3] [Medline: 24820247]
39. Olsson L, Karlsson J, Berg U, Kärrholm J, Hansson E. Person-centred care compared with standardized care for patients undergoing total hip arthroplasty - a quasi-experimental study. J Orthop Surg Res 2014 Oct 09;9:95 [FREE Full text] [doi: 10.1186/s13018-014-0095-2] [Medline: 25359278]
40. Ramkumar PN, Muschler GF, Spindler KP, Harris JD, McCulloch PC, Mont MA. Open mHealth architecture: a primer for tomorrow's orthopedic surgeon and introduction to its use in lower extremity arthroplasty. J Arthroplasty 2017 Apr;32(4):1058-1062. [doi: 10.1016/j.arth.2016.11.019] [Medline: 27956125]
41. Ramkumar PN, Haeberle HS, Ramanathan D, Cantrell WA, Navarro SM, Mont MA, et al. Remote patient monitoring using mobile health for total knee arthroplasty: validation of a wearable and machine learning-based surveillance platform. J Arthroplasty 2019 Oct;34(10):2253-2259. [doi: 10.1016/j.arth.2019.05.021] [Medline: 31128890]
42. Bethge M, Bartel S, Streibelt M, Lassahn C, Thren K. [Improved outcome quality following total knee and hip arthroplasty in an integrated care setting: results of a controlled study]. Rehabilitation (Stuttg) 2011 Apr;50(2):86-93. [doi: 10.1055/s-0030-1265144] [Medline: 21503861]
43. Desmedt M, Vertriest S, Hellings J, Bergs J, Dessers E, Vankrunkelsven P, et al. Economic impact of integrated care models for patients with chronic diseases: a systematic review. Value Health 2016;19(6):892-902 [FREE Full text] [doi: 10.1016/j.jval.2016.05.001] [Medline: 27712719]
44. Vargiu E, Dorado A, Massip M, Gómez-Martínez M, Torres G, de Batlle J, et al. User experience in CONNECARE: results and lessons learnt in Lleida. XPA & Health Communication. 2019. URL: http://www.xpabcn.com/revista/index.php/XPAHC/article/view/22 [accessed 2021-02-13]

Abbreviations

ASA: American Society of Anesthesiologists
CCP: complex chronic patient
CONNECARE: Personalized Connected Care for Complex Chronic Patients
EMR: electronic medical record
ERAS: enhanced recovery after surgery
IC: integrated care
ICER: incremental cost-effectiveness ratio

mHealth: mobile health
OA: osteoarthritis
QoL: quality of life
SACM: smart adaptive case management
SF-12: 12-item short-form survey
THA: total hip arthroplasty
TKA: total knee arthroplasty
UC: usual care

Edited by R Kukafka; submitted 01.03.21; peer-reviewed by G Carot-Sans, J Roca; comments to author 27.03.21; revised version received 17.05.21; accepted 14.06.21; published 02.09.21.

Please cite as:
Colomina J, Drudis R, Torra M, Pallisó F, Massip M, Vargiu E, Nadal N, Fuentes A, Ortega Bravo M, Miralles F, Barbé F, Torres G, de Batlle J, The CONNECARE-Lleida Group
Implementing mHealth-Enabled Integrated Care for Complex Chronic Patients With Osteoarthritis Undergoing Primary Hip or Knee Arthroplasty: Prospective, Two-Arm, Parallel Trial
J Med Internet Res 2021;23(9):e28320
URL: https://www.jmir.org/2021/9/e28320
doi:10.2196/28320
PMID:34473068

Original Paper

Evaluation of Changes in Depression, Anxiety, and Social Anxiety Using Smartphone Sensor Features: Longitudinal Cohort Study

Jonah Meyerhoff[1], PhD; Tony Liu[2], BA; Konrad P Kording[3,4], PhD; Lyle H Ungar[2], PhD; Susan M Kaiser[1], MPH; Chris J Karr[5], MA; David C Mohr[1], PhD

[1]Center for Behavioral Intervention Technologies, Department of Preventive Medicine, Northwestern University, Chicago, IL, United States

[2]Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, United States

[3]Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, United States

[4]Department of Neuroscience, University of Pennsylvania, Philadelphia, PA, United States

[5]Audacious Software, Chicago, IL, United States

Corresponding Author:
David C Mohr, PhD
Center for Behavioral Intervention Technologies
Department of Preventive Medicine
Northwestern University
750 N Lake Shore Dr
10th Floor
Chicago, IL, 60611
United States
Phone: 1 312 503 1403
Email: d-mohr@northwestern.edu

Abstract

Background: The assessment of behaviors related to mental health typically relies on self-report data. Networked sensors embedded in smartphones can measure some behaviors objectively and continuously, with no ongoing effort.

Objective: This study aims to evaluate whether changes in phone sensor–derived behavioral features were associated with subsequent changes in mental health symptoms.

Methods: This longitudinal cohort study examined continuously collected phone sensor data and symptom severity data, collected every 3 weeks, over 16 weeks. The participants were recruited through national research registries. Primary outcomes included depression (8-item Patient Health Questionnaire), generalized anxiety (Generalized Anxiety Disorder 7-item scale), and social anxiety (Social Phobia Inventory) severity. Participants were adults who owned Android smartphones. Participants clustered into 4 groups: multiple comorbidities, depression and generalized anxiety, depression and social anxiety, and minimal symptoms.

Results: A total of 282 participants were aged 19-69 years (mean 38.9, SD 11.9 years), and the majority were female (223/282, 79.1%) and White participants (226/282, 80.1%). Among the multiple comorbidities group, depression changes were preceded by changes in GPS features (*Time*: $r=-0.23$, $P=.02$; *Locations*: $r=-0.36$, $P<.001$), exercise duration ($r=0.39$; $P=.03$) and use of active apps ($r=-0.31$; $P<.001$). Among the depression and anxiety groups, changes in depression were preceded by changes in GPS features for *Locations* ($r=-0.20$; $P=.03$) and *Transitions* ($r=-0.21$; $P=.03$). Depression changes were not related to subsequent sensor-derived features. The minimal symptoms group showed no significant relationships. There were no associations between sensor-based features and anxiety and minimal associations between sensor-based features and social anxiety.

Conclusions: Changes in sensor-derived behavioral features are associated with subsequent depression changes, but not vice versa, suggesting a directional relationship in which changes in sensed behaviors are associated with subsequent changes in symptoms.

(J Med Internet Res 2021;23(9):e22844) doi:10.2196/22844

KEYWORDS

mHealth; personal sensing; digital phenotyping; passive sensing; ecological momentary assessment; depression; anxiety; digital biomarkers; digital phenotyping; mental health assessment; mobile device; mobile phone; internet technology; psychiatric disorders; mobile phone

Introduction

Background

Behaviors such as levels of activity and social engagement are associated with common mental health conditions such as depression and anxiety [1-4]. Retrospective evaluations of these behaviors through self-report measures or interviews rely heavily on retrospective recall, which is subject to systematic biases [5,6]. Even more accurate methods, such as ecological momentary assessment (EMA), which acquire self-reported experiences in the course of peoples' lives [7,8], have not proven practical over extended periods outside of research settings [7]. Accelerometry data from wearable devices that measure activity levels have also been associated with depression [9], but many people stop wearing the devices within the first weeks [10].

Smartphones are becoming ubiquitous. As of 2019, 81% of Americans owned a smartphone [11], as did 76% of people in countries with advanced economies, and 45% in emerging nations [12]. Smartphones are fully integrated into our lives, supporting a growing number of activities. Smartphones contain embedded networked sensors that provide continuous, objective data without user effort, which can be used to produce behavioral markers. A growing body of research suggests that these sensor data can be associated with common mental health problems [13]. Location features derived from GPS actual measurement of patterns of locations visited, time in locations, and travel in phone communications such as text messaging and phone and app use have been associated with depression, anxiety, and social anxiety [14-17].

The potential for personal mobile sensing to improve our understanding of the relationship between behavior and mental health, as well as to advance clinical care, has been widely recognized [18]. However, although there is promise, research to date has had a number of weaknesses. Many of these studies have been conducted in small, relatively homogenous groups, such as students [15,16,19-21]. Although there has been some specificity, with location features tending to identify depression [14,15,19,22,23] and communication features tending to predict social anxiety [17], there have also been a few studies that have found the opposite [16,24]. This may be because of the heterogeneity in symptoms and comorbidities [25], which are common and have not been considered in existing research [13].

To date, studies have focused on using sensed behavioral features to estimate a person's state, either the presence or absence of a condition or symptom severity. With some exceptions, they have generally not evaluated the capacity for sensed behavioral features to predict whether symptoms will increase or decrease in the future. Among the few studies that have examined the capacity of sensed features to predict symptom change, one small study of 18 patients with bipolar disorder found that greater inconsistencies in rates of typing on a smartphone keyboard were related to future greater depressive symptom severity [26]. Relative to studies that use sensed behavior to estimate a person's state, the temporal relationship between sensed behaviors and symptom change has received relatively little attention.

Objectives

In this study, we examine the temporal relationship between changes in sensor features and subsequent changes in mental health symptoms in a large sample of participants. The aim of this exploratory study is to evaluate whether changes in classes of smartphone sensor features were associated with changes in symptom severity for depression, anxiety, and social anxiety, across all participants as well as within groups clustered based on symptoms.

Methods

Participants

Participants were recruited from July 15 to July 26, 2019, through ResearchMatch, a National Institute of Health-funded volunteer network, and the Center for Behavioral Intervention Technologies research registry. Participants were included if they were US citizens and residents, age ≥18 years, could read English, and had an Android smartphone. Participants were excluded if they endorsed, via self-report, having been diagnosed with a severe mental illness, defined as bipolar disorder, schizophrenia, or other psychotic disorder. Participants were compensated for completing measures at set assessment points as well as for completing EMA check-ins. Compensation for completing assessments increased as the study period progressed, such that participants were compensated relatively less for early assessment points and relatively more for assessments toward the end of the study. No single assessment was compensated at more than US $32.50 per assessment time point. Recruitment was advertised as a study on depression and deliberately oversampled depressed participants such that at least 50% of the sample experienced at least moderate depression symptom severity according to the 8-item Patient Health Questionnaire (PHQ-8).

Procedures

Participants downloaded the Passive Data Kit [27] mobile app, which unobtrusively collects phone sensor data and administers surveys. Web-based questionnaires were administered every 3 weeks. Participants were enrolled in the study for 16 weeks. All procedures were approved by the Northwestern University Institutional Review Board, and informed consent was obtained from all participants before participation.

Measures

Participants completed web-based symptom severity assessments at baseline and every 3 weeks until the end of the study period (ie, weeks 4, 7, 10, 13, and 16). Symptom measures included depression severity (PHQ-8) [28], generalized anxiety disorder (Generalized Anxiety Disorder 7-item scale [GAD-7]) [29], and social anxiety disorder (Social Phobia Inventory [SPIN]) [30]. The PHQ-8 was administered as an EMA survey after the baseline assessment point and, subsequently, had one additional assessment point (week 1) relative to other symptom measures.

Phone sensor data included GPS coordinates sampled once every 5 minutes, communication information (ie, phone logs and duration, text message logs, and length), and open apps. Assessment weeks occurred every 3 weeks, during which

participants were asked each evening to label the semantic location (type of location) that they had visited for more than 10 minutes [31]. A series of maps identifying each location were presented, and participants selected the category of each place (eg, home, work, errand, entertainment, place of worship, etc).

Data Analyses

Phone Sensor Feature Transformation

Overview

We considered four categories of phone features for our analysis: GPS-derived movement and location information, semantic locations, app use, and phone-based communication (calls and texts).

To increase interpretability and reduce the number of sensor features, we aggregated features first based on their phone sensor source, as different sensor sets provide unique information. Within each sensor set, we used unit weightings that maximized the interpretability for each feature aggregation. Where possible, we used existing theory to guide our unit aggregations. All sensor features were standardized (mean-centered with unit variance) across the full sample and averaged to produce sensor groupings within the four sensor categories. A full list of features, feature calculations, and their groupings can be found in Multimedia Appendix 1.

GPS-Derived Location and Movement

Following the methodology of Saeb et al [20], we computed high-level features from GPS data that measure participant movement, including location variance (variability in GPS location), total unique location clusters, location entropy (variability in time spent at location clusters), normalized entropy, total distance traveled, average velocity, and circadian movement (extent to which sequence of locations followed a 24-hour pattern). The features were aggregated into *Locations* (location cluster and location variance; represents the number and variability in locations visited), *Time* (total entropy, normalized entropy, and circadian movement; represents the variability in time spent across locations), and *Transitions* (distance traveled and velocity; represents travel between locations).

Semantic Location

Labels for semantic location categories included home duration, work duration, shopping duration, social activities duration (eg, friends' homes and entertainment), religious activities duration (eg, place of worship), and exercise location duration (eg, gyms). During the nonassessment weeks, semantic labels were assigned to locations visited using GPS coordinates assigned during the assessment weeks. This allowed us to estimate the daily duration of time participants spent in each semantic location category.

Communication

The number of incoming and outgoing calls and texts, call duration, and text message length were summed to obtain daily aggregates. The feature groups were *Calls* and *Text Messages*.

App Use

Apps running in the foreground of the phone were sampled every 5 minutes. We aggregated to produce daily app use duration measurements. We grouped apps of interest into 3 categories based on previous theory that certain apps facilitate active use, whereas others elicit more passive use [32,33]. This theoretical underpinning resulted in 3 categories of app use that were manually constructed using unit weighting. Final categorizations included: *Active Apps* (eg, messaging, email, and maps), which required active engagement to complete the primary essential task of each app, *Information Consumption Apps* (eg, YouTube and web browsers) where the primary purpose was more passive consumption of information or entertainment and *Social Apps* (eg, Facebook, Instagram, and Snapchat), which were considered social media apps, and were generally considered as a unique category of apps [34].

Population Clustering

Heterogeneity in underlying symptom patterns may impede the ability to observe clinically meaningful relationships between sensor features and symptom severity [13,35]. We used a data-driven approach, performing k-means clustering on the baseline PHQ-8, GAD-7, and SPIN items [36]. We chose k=4 using the elbow heuristic to choose the number of clusters (Multimedia Appendix 2). Qualitative analysis of these clusters showed that the 4 groups roughly corresponded to (1) a *Minimal Symptom* cluster (n=88), comprising participants characterized by low mean scores on all outcome measures; (2) a *Depression andSocial Anxiety* cluster (n=71) that included participants with predominantly moderate severity scores on the PHQ-8 and the SPIN, but low scores on the GAD-7 measures; (3) a *Depression and Anxiety* cluster (n=69), characterized by generally moderate-severe symptoms on the PHQ-8, moderate symptoms on the GAD-7, but mild ratings on the SPIN; and a (4) *Multiple Comorbidities* cluster (n=54) characterized by elevated ratings across all three symptom measures, with a substantial proportion scoring in the severe range.

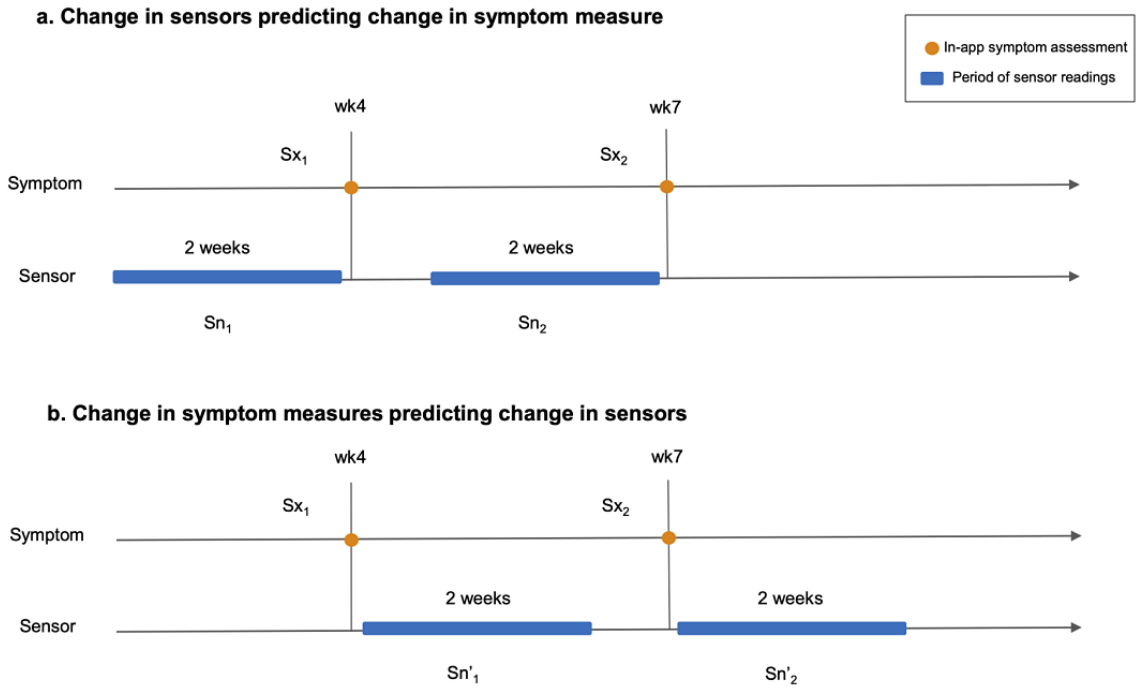
Statistical Methods: Correlation of Changes in Sensors to Changes in Symptom Severity

Figure 1 shows the strategy we used to lag, by 2 weeks, repeated measure correlations [37] of the changes in phone sensor features with changes in symptom severity. The 2-week window for sensor features, consistent with previous research [14,20], was used to allow for sufficient quality of sensor readings to match the retrospective time spans of self-report questionnaires and to maximize data available for analysis while preventing overlapping data sources (ie, symptom outcomes and concurrent sensed behavioral data) from being used at different time points. For the PHQ-8, we had six check-ins across the entire study, yielding five pairs of changes for each participant, whereas for GAD-7 and SPIN, we had five check-ins, yielding four pairs of changes for each participant. For analyses, in which changes in sensor features were used to estimate subsequent symptom severity, Sn_2-Sn_1 was correlated with Sx_2-Sx_1. For analyses in which changes in symptom severity were used to estimate changes in subsequent sensor features, Sx_2-Sx_1 was correlated with $Sn'_2-Sn'_1$. To correct for multiple comparisons, we

computed adjusted *P* values using the Benjamini-Hochberg procedure to control the false discovery rate [38].

If one assessment check-in was missing from a given pair of check-in dates, we used a within-person mean-fill method for the missing assessment. Any pair of assessment check-ins that had missing phone sensor data was discarded from analyses. Power calculations revealed that a sample size of 255 would be required to detect an effect size (correlation, $|\rho|$) of 0.2 at an α of .05 and power (β) of .90.

Figure 1. The sensor window preceding (a) and proceeding (b) the assessment check-ins. Correlations are run as corrected (Sx_2-Sx_1, Sn_2-Sn_1) and corrected (Sx_2-Sx_1, $Sn'_2-Sn'_1$).

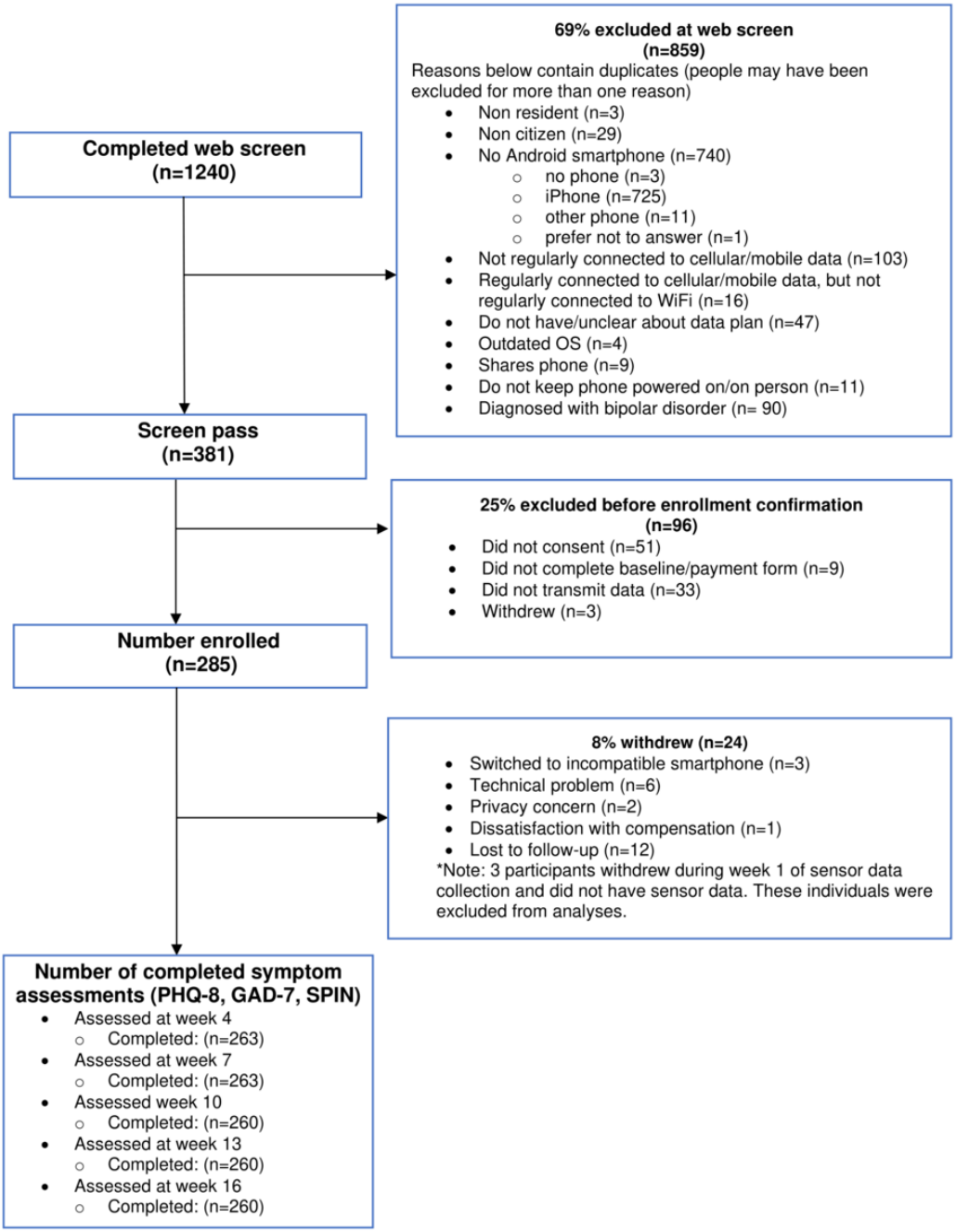


Results

Participants

The flow of participants in this study is shown in Figure 2. Participant demographic and baseline symptom severity characteristics across the entire sample and participant clusters are detailed in Multimedia Appendix 3.

Figure 2. Participant flow diagram. GAD-7: Generalized Anxiety Disorder 7-item scale; PHQ-8: 8-item Patient Health Questionnaire; SPIN: Social Phobia Inventory.



Symptom Change Over Time

Ordinary least square regression revealed no significant change in symptom severity as a function of time (PHQ-8: *P*=.80; GAD-7: *P*=.83; SPIN: *P*=.57). However, there was substantial within-participant variability depending on the symptom measure, with mean SDs of 2.66, 3.50, and 5.90, for the PHQ-8, GAD-7, and SPIN, respectively.

Association Between Sensor-Derived Behavioral Feature Changes and Subsequent Symptom Severity Change

Overview

Table 1 displays the repeated measure correlations primary outcomes by symptom cluster.

Table 1. Repeated measure correlations between sensor and symptom changes and symptom and sensor changes[a].

Characteristics	Repeated measure correlations									
	Change in sensor features association with change in symptom measure					Change in symptom measure association with change in sensor features				
	Value, n	dof[b]	r_{rm}[c]	*P* value (uncorrected)	*P* value (corrected)[d]	Value, n	dof[b]	r_{rm}	*P* value (uncorrected)	*P* value (corrected)[d]
Symptom measure: PHQ-8[e]										
Full sample[f]										
GPS features										
Locations	223	802	−0.17	<.001	<.001	225	801	<0.001	.98	.98
Time	223	802	−0.12	<.001	.003	225	801	−0.006	.86	.97
Transitions	223	802	−0.12	<.001	.003	225	801	0.020	.56	.94
Semantic location										
Home duration	−225	806	0.054	.13	.23	224	801	0.017	.64	.95
Work duration	192	700	0.026	.50	.53	190	691	−0.012	.74	.97
Shopping duration	212	767	0.009	.80	.80	210	210	−0.005	.88	.97
Social activities duration	219	790	−0.062	.08	.17	207	207	−0.021	.57	.94
Religious activities duration	54	195	−0.084	.24	.36	44	44	−0.14	.08	.31
Exercise location duration	85	314	0.18	.001	.005	81	81	−0.13	.02	.30
Communication										
SMS text messages	223	796	−0.034	.34	.46	221	790	−0.062	.08	.31
Calls	225	802	0.03	.40	.50	221	786	−0.023	.52	.94
App use										
Active apps	226	809	−0.041	.24	.36	225	807	−0.004	.90	.97
Information consumption apps	225	805	0.026	.47	.53	226	809	−0.072	.04	.30
Social apps	208	748	0.073	.05	.14	207	746	−0.021	.57	.94
Subgroups (features with corrected *P*≥.1 omitted)										
Multiple comorbidities										
Locations	41	143	−0.36	<.001	<.001	41	143	0.021	.80	.93
Time	41	143	−0.23	.005	.02	41	143	−0.061	.46	.70
Transitions	41	143	−0.18	.03	.07	41	143	0.051	.55	.74
Exercise location duration	11	41	0.39	.01	.03	10	33	−0.13	.45	.70
Active apps	42	146	−0.31	<.001	<.001	42	146	0.10	.22	.70
Depression and anxiety										
Locations	56	204	−0.20	.005	.03	56	202	0.022	.75	.94
Transitions	56	204	−0.21	.002	.03	56	202	−0.17	.02	.22
Depression and social anxiety										
Locations	61	218	−0.17	.01	.08	62	218	0.007	.92	.93

Characteristics	Repeated measure correlations									
	Change in sensor features association with change in symptom measure					Change in symptom measure association with change in sensor features				
	Value, n	dof[b]	r_{rm}[c]	*P* value (uncorrected)	*P* value (corrected)[d]	Value, n	dof[b]	r_{rm}	*P* value (uncorrected)	*P* value (corrected)[d]
Time	61	218	−0.16	.02	.08	62	218	0.025	.71	.93
Social Activities Duration	60	214	−0.17	.01	.08	57	208	0.006	.93	.93
Symptom measure: SPIN[g]										
Depression and social anxiety (features with corrected *P*≥.1 omitted)										
Calls	66	195	0.25	<.001	.005	66	193	-0.045	.53	.73

[a]There were no significant associations between sensor features and subsequent 8-item Patient Health Questionnaire (PHQ-8) symptoms or PHQ-8 symptoms and subsequent sensor features within the *minimal symptoms* group. There were also no significant associations between sensor features and subsequent Generalized Anxiety Disorder 7-item scale (GAD-7) symptoms or GAD-7 symptoms and subsequent sensor features for any subgroup.

[b]dof = n (k − 1), where n is the total number of participants and k is the average number of observations per participant.

[c]Repeated measures correlation coefficient.

[d]Benjamini-Hochberg corrected *P* values. Identical *P* values are due to the recursive definition of the Benjamini-Hochberg correction; it is possible for corrected *P* values to be identical, especially for nonsignificant correlations.

[e]PHQ-8: Patient Health Questionnaire-8 item.

[f]Full results are provided for Patient Health Questionnaire-8 item results in the full sample only to display all sensed behavioral features. Thereafter only relationships with corrected *P*<.1 are displayed.

[g]SPIN: Social Phobia Inventory.

Location Features

Patterns in individuals' movements were associated with subsequent changes in psychological symptoms, specifically depressive symptoms. Changes in GPS-derived *Location* were negatively associated with changes in the PHQ-8 in two of three symptom clusters, *Multiple Comorbidities* (r=−0.36; P<.001), *Depression and Anxiety* (r=-.20; P=.03), and the full sample (r=−0.17; P<.001) and trended toward significance in the *Depression and Social Anxiety* cluster (r=−0.16; P=.08), but the results were not statistically significant. Changes in GPS-derived *Time* were negatively associated with changes in PHQ-8 within the *Multiple Comorbidities* cluster (r=−0.23; P=.02) and the full sample (r=−0.12; P=.003) and trended toward significance in the *Depression and Social Anxiety* cluster (r=−0.16; P=.08), but the results were not statistically significant. Changes in GPS-derived *Transitions* were negatively correlated with changes in PHQ-8 for the *Depression and Anxiety* cluster (r=−0.21; P=.03) and the full sample (r=−0.12; P=.003) and trended toward significance in the *Multiple Comorbidities* cluster (r=−0.18; P=.07), but the results were not statistically significant. There were no significant relationships for the *Minimal Symptom* cluster for the PHQ-8 and no significant relationships between GPS features and subsequent changes in the GAD-7 or SPIN.

Certain types of semantic locations were also associated with PHQ-8 changes. In the depression and social anxiety cluster, there was a trend toward significance for *Social Activity Duration*, which was negatively correlated with changes in the PHQ-8 (r=−0.17; P=.08), but the results were not statistically significant. Within the *Multiple Comorbidities* symptom cluster and the full sample, *Exercise Location Duration* was positively associated with subsequent changes in PHQ-8 scores (*Multiple Comorbidities* cluster: r=0.39, P=.03; full sample: r=0.18, P=.005). This unexpected association between changes in exercise locations and changes in PHQ-8 was likely due to a preponderance of individuals who spent no time in exercise-based locations (ie, zero change in *Exercise Location Duration* from zero time spent in exercise locations), thus overweighting the data of some individuals who saw increased time spent in exercise-based locations with increased PHQ-8. There were no significant relationships within the minimal symptom cluster for the PHQ-8, and no significant relationships were found between semantic location features and subsequent changes in the GAD-7 or SPIN.

Telephone Calls

Within the *Depression and Social Anxiety* baseline cluster, increases in calls were associated with increases in SPIN scores (r=0.25; P<.001).

App Use

Changes in *Active App Use* were negatively correlated (r=−0.31; P<.001) with changes in PHQ-8 scores within the *Multiple Comorbidities* cluster.

Association Between Changes in Symptom Severity and Subsequent Change in Sensor-Derived Behavioral Features

There were no significant correlations between changes in any symptom severity measures and subsequent changes in sensor-derived behavioral features.

Missing Data

Across all treatment weeks, missingness (ie, the number of missing observations/total number of possible observations for

all 282 participants) was higher for the PHQ-8 (277/1692, 16.37%) assessments than for the GAD-7 (104/1410, 7.38%) and SPIN (104/1410, 7.38%). PHQ-8 assessments were administered through our smartphone app, whereas the GAD-7 and SPIN were administered through REDCap (Research Electronic Data Capture) [39]. In addition, across symptom outcome measures and relative to baseline symptom levels (Multimedia Appendix 3), those with missing assessments tended to have slightly higher baseline symptom severity (PHQ-8 range: 10.94-12.82; GAD-7 range: 9.95-11.5; and SPIN range: 24.52-27.48).

Discussion

Principal Findings

Changes in numerous phone sensor–derived behavioral features were associated with subsequent changes in mental health symptoms among people with elevated symptoms of depression. However, changes in symptoms were not associated with subsequent changes in behavioral features. GPS location features were fairly consistently and negatively, albeit modestly, related to subsequent changes in depression severity across symptom groups. This is consistent with a number of previous relatively small studies showing correlations between GPS features and depression [14,19,20,22,40,41]. This larger study confirms these earlier findings, and importantly, indicates a directional relationship in which GPS features are associated with subsequent increases or decreases in depressive symptoms, but not with anxiety or social anxiety.

The types of locations (work, shopping, etc) people visited were less consistently related to changes in depression. This does not necessarily mean that specific locations are unimportant at the individual level: one person may like shopping, whereas another may detest it. However, this suggests that patterns of movement through geographic space, irrespective of the destinations or locations to which one travels, are indicators of symptom change among people with depression. We speculate that this may reflect a loss of motivation expressed through geographic movement. Perhaps more speculatively, it may also be that changes in neurocognitive mechanisms, such as executive control, affect, and motivation, impact both depression and basic mechanisms involved in movement through geographic space [42,43].

The different constellations of symptoms that participants experience impact the salience of some sensed behaviors in predicting outcomes. For example, only within participants in the *Depression and Social Anxiety* symptom cluster was the amount of time spent in locations related to social activities associated with (at the trend level) subsequent change in depressive symptoms, suggesting that although locations are generally useful for depression prediction, social activities may be particularly useful when social anxiety symptoms are present. *Active App Use* (texting, email, and mapping) was associated with depression change among those with multiple and more severe comorbidities. Although GPS features were generally useful, features integrating time and location were not useful among those with comorbid generalized anxiety, and features measuring transitioning between locations were not useful for those with comorbid social anxiety. Thus, there was support for the notion that symptom constellations are important considerations for some features.

Increases in telephone calls were associated with increases in social anxiety symptoms among clusters characterized by depression and social anxiety. This finding notwithstanding, the capacity for sensor-derived features to be associated with changes in social anxiety symptoms was not as consistent as that for depressive symptoms. Furthermore, we did not find any associations between sensor-derived features and generalized anxiety symptom changes.

These findings indicate that sensor-derived behavioral features, which are objective and can be acquired with reduced participant effort, can be a useful tool for investigating the role of some behaviors in changing depressive symptom severity. There has been much speculation about the clinical potential of mobile sensing [18,44]. The effect sizes are modest, albeit consistent with many other studies that have examined the use of sensed behavioral features to estimate the presence or severity of symptoms [16,19]. This study does not support the use of phone sensor data alone in monitoring symptom changes in mental health populations; however, phone sensor data may be useful in conjunction with other networked sensing tools such as wearables. Phone sensor data may be useful for digital mental health interventions [45]. Just-in-time adaptive interventions [46,47] use individualized data to predict risk and deliver context-aware intervention material that is adaptive. These approaches are increasingly applied in digital health interventions, such as identifying when to send messages to increase physical activity such as step counts [48]. The promise of delivering motivational messages at opportune moments that reinforce behavioral activation strategies, such as visiting someplace new, spending more time outside of the home or work, or engaging socially, has the potential to improve engagement with these tools and reduce depression.

Limitations

This study had several limitations. First, the exploratory nature of this study requires interpreting results with caution and necessitates that future work must explicitly test the a priori hypotheses arising from these results. Next, although we establish significant temporal relationships between sensed behavior changes and subsequent changes in symptom severity, our study is not experimental and does not establish causal relationships. Furthermore, our sensor feature aggregations were limited to single sensor sources and were constructed to maximize interpretability; however, future studies that use data-driven aggregations are necessary to help inform feature aggregations across sensor modalities. Although aggregating across sensor modalities presents a challenge for interpretability, future work that examines cross-sensor aggregations could yield improved estimation of sensed behaviors and, subsequently, more robust associations with changes in symptoms. In addition, although these findings provide some support for the hypothesis that sensed behavior change is associated with subsequent changes in depression and not vice versa, this study examined the associations between changes in sensor features and subsequent changes in symptom severity measures lagged by

2 weeks, and therefore should not be generalized to periods beyond the 2-week window. Another limitation is that our sample only included those who used Android devices and agreed to participate in this research. App use and communication data are not readily available for iOS devices. Regarding data missingness, across all the surveys, individuals who had missing data had higher baseline symptom severity than the overall sample, though not dramatically so; thus, data were missing not at random. These missingness rates are in line with established criteria that are often used as the standard for good trial data [49]. Finally, although we controlled for multiple analyses, we nonetheless caution against overinterpretation of more isolated findings that need to be replicated in future studies.

Conclusions

The ubiquity of smartphones with networked sensors has opened up new opportunities to identify behavioral markers related to mental health that can be acquired continuously and effortlessly. Changes in movement through geographic space were consistently associated with subsequent changes in depressive symptoms; however, there was no evidence that changes in depression were associated with subsequent changes in sensed behaviors. This supports a directional relationship in which changes in movement patterns precede symptom change, but symptom change does not precede changes in movement behaviors.

Acknowledgments

This work was supported by a grant from the National Institute of Mental Health (5R01MH111610) to DCM and KPK. DCM had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. JM is supported by a grant from the National Institute of Mental Health (T32MH115882). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

DCM has accepted consulting fees from Apple Inc, Pear Therapeutics, Otsuka Pharmaceuticals, and the One Mind Foundation. He also accepted royalties from Oxford Press and has an ownership interest in Adaptive Health, Inc. SMK accepted the consulting fees from Adaptive Health, Inc. CJK is the founder and sole developer at Audacious Software, LLC, which provides academic and research software development to a variety of academic, nonprofit, and commercial entities that are similar to the contributions to this work. None of the other authors have any competing interests to declare.

Multimedia Appendix 1
Sensor features and groupings.
[PDF File (Adobe PDF File), 139 KB - jmir_v23i9e22844_app1.pdf]

Multimedia Appendix 2
Symptom cluster elbow plot.
[PDF File (Adobe PDF File), 132 KB - jmir_v23i9e22844_app2.pdf]

Multimedia Appendix 3
Demographics and baseline characteristics.
[PDF File (Adobe PDF File), 135 KB - jmir_v23i9e22844_app3.pdf]

References

1. Cacioppo JT, Hawkley LC, Thisted RA. Perceived social isolation makes me sad: 5-year cross-lagged analyses of loneliness and depressive symptomatology in the Chicago Health, Aging, and Social Relations Study. Psychol Aging 2010 Jun;25(2):453-463 [FREE Full text] [doi: 10.1037/a0017216] [Medline: 20545429]
2. Hiles SA, Lamers F, Milaneschi Y, Penninx BW. Sit, step, sweat: longitudinal associations between physical activity patterns, anxiety and depression. Psychol Med 2017 Jun;47(8):1466-1477. [doi: 10.1017/S0033291716003548] [Medline: 28137333]
3. Pereira SM, Geoffroy M, Power C. Depressive symptoms and physical activity during 3 decades in adult life: bidirectional associations in a prospective cohort study. JAMA Psychiatry 2014 Dec 01;71(12):1373-1380. [doi: 10.1001/jamapsychiatry.2014.1240] [Medline: 25321867]
4. Da Silva MA, Singh-Manoux A, Brunner EJ, Kaffashian S, Shipley MJ, Kivimäki M, et al. Bidirectional association between physical activity and symptoms of anxiety and depression: the Whitehall II study. Eur J Epidemiol 2012 Jul;27(7):537-546 [FREE Full text] [doi: 10.1007/s10654-012-9692-8] [Medline: 22623145]
5. Tversky A, Kahneman D. Availability: A heuristic for judging frequency and probability. Cognit Psychol 1973 Sep;5(2):207-232. [doi: 10.1016/0010-0285(73)90033-9]

6. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
7. Csikszentmihalyi M. In: de Vries MW, editor. The Experience of Psychopathology: Investigating Mental Disorders in their Natural Settings. Cambridge: Cambridge University Press; 1992.
8. Killingsworth MA, Gilbert DT. A wandering mind is an unhappy mind. Science 2010 Nov 12;330(6006):932 [FREE Full text] [doi: 10.1126/science.1192439] [Medline: 21071660]
9. Cruz BD, Alfonso-Rosa RM, McGregor D, Chastin SF, Palarea-Albaladejo J, Cruz JD. Sedentary behaviour is associated with depression symptoms: Compositional data analysis from a representative sample of 3233 US adults and older adults assessed with accelerometers. J Affect Disord 2020 Mar 15;265:59-62. [doi: 10.1016/j.jad.2020.01.023] [Medline: 31959584]
10. Lee H, Lee Y. A look at wearable abandonment. In: Proceedings of the 18th IEEE International Conference on Mobile Data Management (MDM). 2017 Presented at: 18th IEEE International Conference on Mobile Data Management (MDM); May 29-June 1, 2017; Daejeon, Korea (South) p. 392-393. [doi: 10.1109/mdm.2017.70]
11. Demographics of mobile device ownership and adoption in the United States. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-08-10]
12. Silver L. Smartphone ownership is growing rapidly around the world, but not always equally. Pew Research Center. 2019 Feb. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2021-08-10]
13. Mohr DC, Zhang M, Schueller SM. Personal Sensing: Understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 May 08;13:23-47. [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
14. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
15. Ware S, Yue C, Morillo R, Lu J, Shang C, Bi J, et al. Predicting depressive symptoms using smartphone data. Smart Health 2020 Mar;15:100093. [doi: 10.1016/j.smhl.2019.100093]
16. Chow PI, Fua K, Huang Y, Bonelli W, Xiong H, Barnes LE, et al. Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students. J Med Internet Res 2017 Mar 03;19(3):e62 [FREE Full text] [doi: 10.2196/jmir.6820] [Medline: 28258049]
17. Gao Y, Li A, Zhu T, Liu X, Liu X. How smartphone usage correlates with social anxiety and loneliness. PeerJ 2016;4:e2197 [FREE Full text] [doi: 10.7717/peerj.2197] [Medline: 27478700]
18. Insel TR. Digital Phenotyping: Technology for a new science of behavior. J Am Med Assoc 2017 Oct 03;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
19. Farhan A, Yue C, Morillo R, Ware S, Lu J, Bi J, et al. Behavior vs. introspection: refining prediction of clinical depression via smartphone sensing data. In: Proceedings of the IEEE Wireless Health (WH). 2016 Presented at: IEEE Wireless Health (WH); Oct. 25-27, 2016; Bethesda, MD, USA. [doi: 10.1109/wh.2016.7764553]
20. Saeb S, Lattie EG, Schueller SM, Kording KP, Mohr DC. The relationship between mobile phone location sensor data and depressive symptom severity. PeerJ 2016;4:e2537 [FREE Full text] [doi: 10.7717/peerj.2537] [Medline: 28344895]
21. Seppälä J, De Vita I, Jämsä T, Miettunen J, Isohanni M, Rubinstein K, M-RESIST Group, et al. Mobile phone and wearable sensor-based mHealth approaches for psychiatric disorders and symptoms: systematic review. JMIR Ment Health 2019 Feb 20;6(2):e9819 [FREE Full text] [doi: 10.2196/mental.9819] [Medline: 30785404]
22. Palmius N, Tsanas A, Saunders KE, Bilderbeck AC, Geddes JR, Goodwin GM, et al. Detecting bipolar depression from geographic location data. IEEE Trans Biomed Eng 2017 Dec;64(8):1761-1771 [FREE Full text] [doi: 10.1109/TBME.2016.2611862] [Medline: 28113247]
23. Canzian L, Musolesi M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp '15: The 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 7 - 11, 2015; Osaka Japan p. 1293-1304. [doi: 10.1145/2750858.2805845]
24. Razavi R, Gharipour A, Gharipour M. Depression screening using mobile phone usage metadata: a machine learning approach. J Am Med Inform Assoc 2020 Apr 01;27(4):522-530. [doi: 10.1093/jamia/ocz221] [Medline: 31977041]
25. Vittengl JR, Clark LA, Smits JA, Thase ME, Jarrett RB. Do comorbid social and other anxiety disorders predict outcomes during and after cognitive therapy for depression? J Affect Disord 2019 Jan 01;242:150-158 [FREE Full text] [doi: 10.1016/j.jad.2018.08.053] [Medline: 30176494]
26. Stange JP, Zulueta J, Langenecker SA, Ryan KA, Piscitello A, Duffecy J, et al. Let your fingers do the talking: Passive typing instability predicts future mood outcomes. Bipolar Disord 2018 May 08;20(3):285-288 [FREE Full text] [doi: 10.1111/bdi.12637] [Medline: 29516666]
27. Passive data kit. Audacious Software. 2018. URL: https://passivedatakit.org/ [accessed 2021-08-10]
28. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]

29. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
30. Connor KM, Davidson JR, Churchill LE, Sherwood A, Foa E, Weisler RH. Psychometric properties of the Social Phobia Inventory (SPIN). New self-rating scale. Br J Psychiatry 2000 Apr 02;176(4):379-386. [doi: 10.1192/bjp.176.4.379] [Medline: 10827888]
31. Saeb S, Lattie EG, Kording KP, Mohr DC. Mobile phone detection of semantic location and its relationship to depression and anxiety. JMIR Mhealth Uhealth 2017 Aug 10;5(8):e112 [FREE Full text] [doi: 10.2196/mhealth.7297] [Medline: 28798010]
32. Montague E, Xu J. Understanding active and passive users: the effects of an active user using normal, hard and unreliable technologies on user assessment of trust in technology and co-user. Appl Ergon 2012 Jul;43(4):702-712 [FREE Full text] [doi: 10.1016/j.apergo.2011.11.002] [Medline: 22192788]
33. Escobar-Viera CG, Shensa A, Bowman ND, Sidani JE, Knight J, James AE, et al. Passive and active social media use and depressive symptoms among United States adults. Cyberpsychol Behav Soc Netw 2018 Jul;21(7):437-443. [doi: 10.1089/cyber.2017.0668] [Medline: 29995530]
34. Harari GM, Müller SR, Stachl C, Wang R, Wang W, Bühner M, et al. Sensing sociability: Individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. J Pers Soc Psychol 2020 Jul;119(1):204-228. [doi: 10.1037/pspp0000245] [Medline: 31107054]
35. Chekroud AM, Gueorguieva R, Krumholz HM, Trivedi MH, Krystal JH, McCarthy G. Reevaluating the efficacy and predictability of antidepressant treatments: a symptom clustering approach. JAMA Psychiatry 2017 Apr 01;74(4):370-378 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.0025] [Medline: 28241180]
36. Taylor S, Jaques N, Nosakhare E, Sano A, Picard R. Personalized multitask learning for predicting tomorrow's mood, stress, and health. IEEE Trans Affect Comput 2020;11(2):200-213. [doi: 10.1109/TAFFC.2017.2784832] [Medline: 32489521]
37. Bakdash JZ, Marusich LR. Repeated measures correlation. Front Psychol 2017 Apr 7;8:00456. [doi: 10.3389/fpsyg.2017.00456]
38. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc Series B (Methodol) 2018 Dec 05;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
39. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
40. Pratap A, Atkins DC, Renn BN, Tanana MJ, Mooney SD, Anguera JA, et al. The accuracy of passive phone sensors in predicting daily mood. Depress Anxiety 2019 Jan 21;36(1):72-81. [doi: 10.1002/da.22822] [Medline: 30129691]
41. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 2015 Sep;38(3):218-226 [FREE Full text] [doi: 10.1037/prj0000130] [Medline: 25844912]
42. Pykett J. Geography and neuroscience: Critical engagements with geography's “neural turn”. Trans Inst Br Geogr 2017 Oct 06;43(2):154-169. [doi: 10.1111/tran.12213]
43. Clark L, Chamberlain SR, Sahakian BJ. Neurocognitive mechanisms in depression: implications for treatment. Annu Rev Neurosci 2009;32:57-74. [doi: 10.1146/annurev.neuro.31.060407.125618] [Medline: 19400725]
44. Torous J, Staples P, Onnela J. Realizing the potential of mobile mental health: new methods for new data in psychiatry. Curr Psychiatry Rep 2015 Aug;17(8):602. [doi: 10.1007/s11920-015-0602-0] [Medline: 26073363]
45. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. NPJ Digit Med 2019;2:88 [FREE Full text] [doi: 10.1038/s41746-019-0166-1] [Medline: 31508498]
46. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. Health Psychol 2015 Dec;34 Suppl:1220-1228. [doi: 10.1037/hea0000305] [Medline: 26651463]
47. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. Ann Behav Med 2016 Sep 23. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
48. Liao P, Greenewald K, Klasnja P, Murphy S. Personalized HeartSteps: A reinforcement learning algorithm for optimizing physical activity. ArXiv: Computer Science - Machine Learning. 2019. URL: http://arxiv.org/abs/1909.03539 [accessed 2020-03-03]
49. Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. Phys Ther 2003 Aug;83(8):713-721. [Medline: 12882612]

Abbreviations

EMA: ecological momentary assessment
GAD-7: Generalized Anxiety Disorder 7-item scale
PHQ-8: 8-item Patient Health Questionnaire

REDCap: Research Electronic Data Capture
SPIN: Social Phobia Inventory

Edited by R Kukafka; submitted 24.07.20; peer-reviewed by K Huckvale, J Zulueta, B Teachman; comments to author 03.09.20; revised version received 29.10.20; accepted 19.07.21; published 03.09.21.

Please cite as:
Meyerhoff J, Liu T, Kording KP, Ungar LH, Kaiser SM, Karr CJ, Mohr DC
Evaluation of Changes in Depression, Anxiety, and Social Anxiety Using Smartphone Sensor Features: Longitudinal Cohort Study
J Med Internet Res 2021;23(9):e22844
URL: https://www.jmir.org/2021/9/e22844
doi:10.2196/22844
PMID:34477562

Original Paper

Automated Detection of Acute Myocardial Infarction Using Asynchronous Electrocardiogram Signals—Preview of Implementing Artificial Intelligence With Multichannel Electrocardiographs Obtained From Smartwatches: Retrospective Study

Changho Han[1*], MD, MSc; Youngjae Song[2*], PhD; Hong-Seok Lim[3*], MD, PhD; Yunwon Tae[2], MSc; Jong-Hwan Jang[1], PhD; Byeong Tak Lee[2], MSc; Yeha Lee[2], PhD; Woong Bae[2], MSc; Dukyong Yoon[1,4], MD, PhD

[1]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Yongin, Republic of Korea

[2]VUNO Inc, Seoul, Republic of Korea

[3]Department of Cardiology, Ajou University School of Medicine, Suwon, Republic of Korea

[4]Center for Digital Health, Yongin Severance Hospital, Yonsei University Health System, Yongin, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Dukyong Yoon, MD, PhD
Department of Biomedical Systems Informatics
Yonsei University College of Medicine
363, Dongbaekjukjeon-daero, Giheung-gu
Yongin, 16995
Republic of Korea
Phone: 82 3151898450
Email: dukyong.yoon@yonsei.ac.kr

Abstract

Background: When using a smartwatch to obtain electrocardiogram (ECG) signals from multiple leads, the device has to be placed on different parts of the body sequentially. The ECG signals measured from different leads are asynchronous. Artificial intelligence (AI) models for asynchronous ECG signals have barely been explored.

Objective: We aimed to develop an AI model for detecting acute myocardial infarction using asynchronous ECGs and compare its performance with that of the automatic ECG interpretations provided by a commercial ECG analysis software. We sought to evaluate the feasibility of implementing multiple lead–based AI-enabled ECG algorithms on smartwatches. Moreover, we aimed to determine the optimal number of leads for sufficient diagnostic power.

Methods: We extracted ECGs recorded within 24 hours from each visit to the emergency room of Ajou University Medical Center between June 1994 and January 2018 from patients aged 20 years or older. The ECGs were labeled on the basis of whether a diagnostic code corresponding to acute myocardial infarction was entered. We derived asynchronous ECG lead sets from standard 12-lead ECG reports and simulated a situation similar to the sequential recording of ECG leads via smartwatches. We constructed an AI model based on residual networks and self-attention mechanisms by randomly masking each lead channel during the training phase and then testing the model using various targeting lead sets with the remaining lead channels masked.

Results: The performance of lead sets with 3 or more leads compared favorably with that of the automatic ECG interpretations provided by a commercial ECG analysis software, with 8.1%-13.9% gain in sensitivity when the specificity was matched. Our results indicate that multiple lead-based AI-enabled ECG algorithms can be implemented on smartwatches. Model performance generally increased as the number of leads increased (12-lead sets: area under the receiver operating characteristic curve [AUROC] 0.880; 4-lead sets: AUROC 0.858, SD 0.008; 3-lead sets: AUROC 0.845, SD 0.011; 2-lead sets: AUROC 0.813, SD 0.018; single-lead sets: AUROC 0.768, SD 0.001). Considering the short amount of time needed to measure additional leads, measuring at least 3 leads—ideally more than 4 leads—is necessary for minimizing the risk of failing to detect acute myocardial infarction occurring in a certain spatial location or direction.

Conclusions: By developing an AI model for detecting acute myocardial infarction with asynchronous ECG lead sets, we demonstrated the feasibility of multiple lead-based AI-enabled ECG algorithms on smartwatches for automated diagnosis of cardiac disorders. We also demonstrated the necessity of measuring at least 3 leads for accurate detection. Our results can be used as reference for the development of other AI models using sequentially measured asynchronous ECG leads via smartwatches for detecting various cardiac disorders.

(J Med Internet Res 2021;23(9):e31129) doi:10.2196/31129

KEYWORDS

wearables; smartwatches; asynchronous electrocardiogram; artificial intelligence; deep learning; automatic diagnosis; myocardial infarction; timely diagnosis; machine learning; digital health; cardiac health; cardiology

Introduction

Wearable devices, simply referred to as "wearables," are smart electronics or computers that are integrated into clothing and other accessories that can be worn on or attached to the body [1]. The consumer adoption of wearable technology for health care services is skyrocketing owing to increasing interest in personalized health management, disease prevention, and fitness [2,3]. One such technology is continuous/day-to-day measurement of single-lead electrocardiograms (ECGs) via smartwatches or other portable/handheld devices [4-6]. These devices can provide a novel opportunity for facilitating timely diagnostics by extending the availability of ECG measurement to the general population outside the hospital.

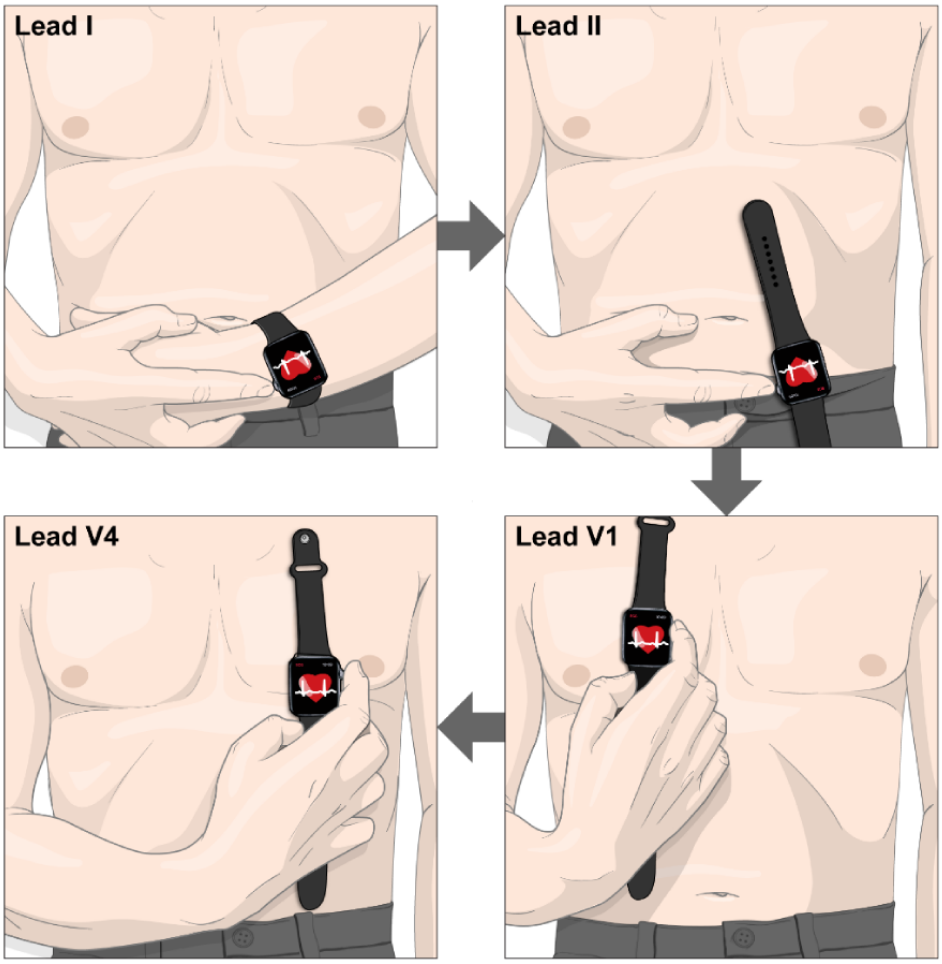
Smartwatches and other portable/handheld ECG devices measure single-lead ECG when the 2 electrode detectors are attached to 2 different parts of the body [5]. However, useful information from other leads can potentially be neglected when only a single lead is evaluated [7]. Analyzing electrical activity of the heart from different spatial locations by measuring multiple leads is necessary for accurate and robust detection of cardiac disorders, such as myocardial infarction, pulmonary embolism, and acute left or right heart failure [8,9]. Accordingly, the standard 12-lead ECG is the most commonly used assessment among physicians for evaluation of the heart.

Previous studies have explored the possibility and described the methodology of measuring multiple ECG leads using smartwatches [9,10]. Multiple ECG leads can be obtained from smartwatches by sequentially placing the smartwatch on different parts of the body (Figure 1). The ECG signals from different leads are asynchronous when measured in this way. There are also reports evaluating the concordance of multiple-lead ECG obtained by smartwatches compared with the standard 12-lead ECG in detecting conditions related to ischemic heart disease when read by physicians [11-13].

To the best of our knowledge, previous studies on automated diagnosis or classification of ECGs using artificial intelligence (AI) have utilized either single-lead ECGs or synchronous multiple-lead ECG signals as input [14-19]. Application of asynchronous ECG signals for AI model development is largely unexplored. Such an application needs to be assessed to ensure that multiple lead-based AI-enabled ECG models can be implemented on smartwatches. Moreover, the adequate number of sequentially recorded leads from smartwatches that would ensure sufficient diagnostic power of the AI-enabled ECG model needs to be verified.

In this study, we aimed to develop an AI model for detecting acute myocardial infarction using asynchronous ECG lead sets and then compare the performance of our model with that of an automatic ECG interpretation provided by a commercial ECG analysis software. Such a model could prove the feasibility of AI-enabled ECG algorithms on smartwatches. As a prerequisite to develop such a model, we derived asynchronous ECG signals from standard 12-lead ECG reports to simulate a situation similar to the sequential recording of ECG leads via smartwatches. Moreover, we aimed to find the optimal number of leads for sufficient diagnostic power by randomly masking each lead channel during the training phase and validating/testing our model with various targeting lead sets (and masking the remaining lead channels).

Figure 1. Example of measuring multi-lead electrocardiogram (ECG) from a smartwatch. Multiple-lead ECG can be obtained from smartwatches by sequentially placing the smartwatch on different parts of the body. The figure depicts an example of measuring leads I, II, V1, and V4 sequentially. Lead I can be recorded with the smartwatch on the left wrist and the right index finger on the crown. Then, after removing the smartwatch from the left wrist, lead II can be recorded with the smartwatch on the left lower quadrant of the abdomen and the right index finger on the crown. Next, leads V1 and V4 can be recorded with the smartwatch on the fourth intercostal space at the right sternal border and fifth intercostal space at the midclavicular line, respectively, with the right index finger on the crown in both cases.



Methods

Ethics Approval

The Institutional Review Board of Ajou University Hospital approved this study (protocol AJIRB-MED-MDB-20-597) and waived the requirement for informed consent because only anonymized data were used retrospectively.

Data Sample and Labeling

We utilized standard 12-lead ECG reports collected from General Electric (GE) ECG machines at Ajou University Medical Center (AUMC), a tertiary teaching hospital in South Korea. These ECG reports of AUMC originally exist as PDFs and are stored in a database. Thus far, the ECG database contains a total of 1,039,550 ECGs from 447,445 patients, collected between June 1994 and January 2018. A previous study extracted raw waveforms, demographic information, and ECG measurement parameters/automatic ECG interpretations made by the GE Marquette 12SL ECG Analysis Program from these reports [20]. In these reports, each lead is 2.5 seconds in duration and sampled at 500 Hz. We also collected clinical data, such as emergency room visit time or the diagnosis of the patients, from the AUMC Electronic Medical Records database.

For our study, we identified and extracted ECGs recorded within 24 hours from each visit to the emergency room between June 1994 and January 2018 from patients aged 20 years or older. For each visit to the emergency room, all diagnoses made during the stay in hospital were collected. If either International Classification of Diseases, Tenth Revision (ICD-10) code I21 (acute myocardial infarction) or I22 (subsequent ST elevation and non-ST elevation) was entered, the ECGs for those visits were labeled as having acute myocardial infarction. For visits that had neither of the 2 ICD-10 codes entered, the ECGs for those visits were labeled as not having acute myocardial infarction.

We split the data into training/validation (80%) and independent hold-out test (20%) sets, and then further split the training/validation set into training (85%) and validation (15%) sets. To reduce ambiguity, we excluded patients whose time of registration for the ICD-10 codes for acute myocardial infarction (I21 or I22) was either “null” (meaning that the registration time was not entered and thus is unknown) or not within 24 hours of ECG measurement.

After model development, we compared the performance of our model with that of the automatic ECG interpretation provided by the GE ECG analysis program. To derive the performance of the automatic ECG interpretation for detecting acute myocardial infarction, we categorized the interpretations in 2 different ways. First, the automatic ECG interpretation was categorized as myocardial infarction if the interpretation included at least one of the following three phrases: “ACUTE MI,” “ST elevation,” and “infarct.” The second categorizing criterion consisted of the 3 phrases in the first labeling criterion along with the following three phrases: “T wave abnormality,” “ST abnormality,” and “ST depression.” We thus derived 2 distinct performance indices from these 2 categories.

Deriving Asynchronous Lead Sets From ECG Reports

Multimedia Appendix 1 shows an example of a standard 12-lead ECG report used at AUMC. These ECG reports are asynchronous as a whole while being synchronous when grouped into 4 subsets of 3 leads each. The x-axis of the ECG report represents time flow; the waveforms on the left side are recorded earlier than those on the right side. The total recorded time of this ECG report is 10 seconds. In Multimedia Appendix 1, leads I, II, and III are shown to have been recorded 2.5 seconds earlier than leads aVR, aVL, and aVF, which were recorded 2.5 seconds earlier than leads V1, V2, and V3, which in turn had been recorded 2.5 seconds earlier than leads V4, V5, and V6.

As previously mentioned, asynchronous ECG lead sets can be derived from ECG reports to simulate a situation similar to the sequential recording of ECG leads via smartwatches. For example, a 4-lead subset consisting of leads I, aVR, V1, and V4 from the ECG report is completely asynchronous. According to the Einthoven law and Goldberger equation, for the 6 limb leads (leads I, II, III, aVR, aVL, and aVF), the remaining 4 leads can be calculated even if only 2 leads are available [21,22]. Here, we trained/validated our AI model by randomly masking each lead channel and then tested our model with various target lead sets (while masking the remaining lead channels) to determine the optimal number of leads for sufficient diagnostic power. The tested lead sets are specified in Multimedia Appendix 2. For the multiple-lead sets, we included lead I in all cases, given that lead I is the most basic lead channel that can be measured from a smartwatch: lead I can be measured by placing the right index finger on the crown without removing the smartwatch from the left wrist. The lead channels in each 4-, 3-, and 2-lead set are completely asynchronous. Thus, the lead channels included in the 4-lead sets were leads I and II (calculated from leads aVR, aVL, and aVF) for the limb leads and all the possible combinations of 2 precordial leads that could be derived from the ECG report while maintaining complete asynchrony. The lead channels included in the 3-lead sets were leads I and lead II (calculated from leads aVR, aVL, and aVF) for the limb leads and 1 precordial lead. The lead channels included in the 2-lead sets were lead I and either lead II (calculated from leads aVR, aVL, and aVF) or 1 precordial lead. We also tested 2 single-lead cases (lead I or II).

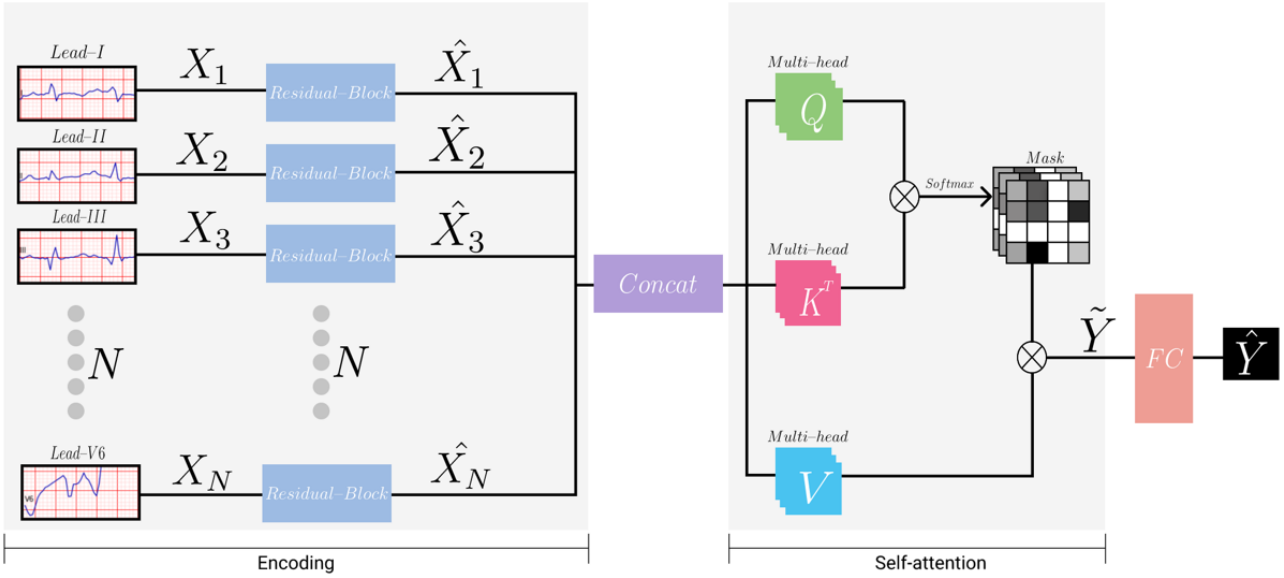
Primary and Secondary Aims of the Study

Our primary aim was to develop an AI model for detecting acute myocardial infarction from asynchronous ECG signals, which outperforms the automatic ECG interpretation provided by the GE ECG analysis program. Our secondary aim was to determine the optimal number of leads required for sufficient diagnostic power. Model performances were assessed using the following statistics: area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPRC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

Neural Network Architecture and Training

Figure 2 illustrates the architecture of the neural network model used in our study. The model was divided into two phases: encoding and self-attention.

Figure 2. Illustration of the neural network's architecture. The encoding phase encodes each lead channel with a weight-shared structure. The self-attention phase captures the relation between each lead channel.



Encoding Phase

The model took the input of 2.5 seconds from each 12-lead ECG channel, which was downsampled from 500 Hz to 250 Hz. Each lead was processed in a separate but weight-shared encoder. Details of the architecture of the encoder are summarized in Multimedia Appendix 3. The encoder consisted of 16 residual blocks with 2 or three 1-dimensional convolutional neural network (CNN) layers in each block [23]. Every CNN layer was followed by a batch normalization layer and a ReLU activation function. All the CNN layers had a kernel size of 7 and the "same" padding. In the first residual block, the output of the first ReLU activation function was connected to the block's output via a pooling layer. In the following 15 residual blocks, the input and output of each block were connected via skip connection. CNN layers with a stride of 2 were applied every 2 or 4 residual blocks. The depth (number of features) of the CNN layers increased by a factor of 2 per 4 residual blocks. For example, a stride described as "2,1,1" in Multimedia Appendix 3 implies that there are 3 CNN layers in that block and the stride of those CNN layers are 2, 1, and 1, respectively. The "Length" and "Depth" columns in Multimedia Appendix 3 are the length and depth of the output of each block. Each feature of the final output of the encoder was average pooled to obtain length=1.

Self-attention Phase

To capture the associations among each lead channel, we utilized a multi-head self-attention module that consisted of queries, keys, and values. Each query, key, and value represented a single dense layer that took all output from the encoder (ie, ) [24]. We computed the dot products of the query with all keys and applied a softmax function to obtain $N \times N$ attention matrices, where N is the number of lead channels. During the training phase of the model, to ensure generalization and applicability for any lead combinations (eg, various 4-, 3-, 2-, and single-lead sets), we randomly masked each lead channel on the attention matrices. Meanwhile, we masked all the lead channels except for the specific targeting leads during the inference phase. For instance, if the target leads were I and V1, we masked all other leads but leads I and V1 during the inference phase. Any lead combinations can be set as target leads. The specific combinations that we tested are specified in Multimedia Appendix 2. After acquiring the attention matrices, we computed the dot products of the values with attention matrices such that the model could reflect the relation between leads. Afterward, these outputs from all the multi-heads were concatenated and linearly projected so that the final output dimension of the multi-head self-attention module became 512 (the same as the original input of the multi-head self-attention module).

We then flattened the output of lead channels before feeding them into the classifier. The classifier had 2 layers of dense layers, which reduced the dimension from 6144 (512 × 12) to 1, followed by a sigmoid layer that calibrated the probability of acute myocardial infarction (ie, ) range from 0 to 1. We split the data into training/validation (80%) and independent hold-out test (20%) sets, and then further split the training/validation set into training (85%) and validation (15%) sets. For training, we used the Adam optimizer with a batch size of 32 and a learning rate of 0.001. We also applied weight decay and several data augmentation techniques, including random Gaussian noise, time scaling, and signal masking, to prevent overfitting. To tune the hyperparameters, we utilized validation data sets with extensive experiment settings (ie, 12-, 4-, 3-, 2-, and single-lead settings). We implemented the model using the Pytorch library.

Results

Data Set Characteristics

From the AUMC ECG database, we extracted 97,742 patients aged 20 years or older with 183,982 ECGs recorded within 24 hours from each visit to the emergency room (Figure 3). After applying the exclusion criteria, we included 76,829 patients with 138,549 ECGs in the training and validation data set, and 19,109 patients with 34,371 ECGs in the test data set. The data set characteristics are summarized in Table 1. The proportion of ECGs labeled as acute myocardial infarction was 1.78% for the training and validation data set, and 1.61% for the test data set.

Figure 3. Patient flow diagram. The patients were split into training and validation (80%) and test (20%) data sets. ECG: electrocardiogram, ICD-10: International Classification of Diseases, Tenth Revision.

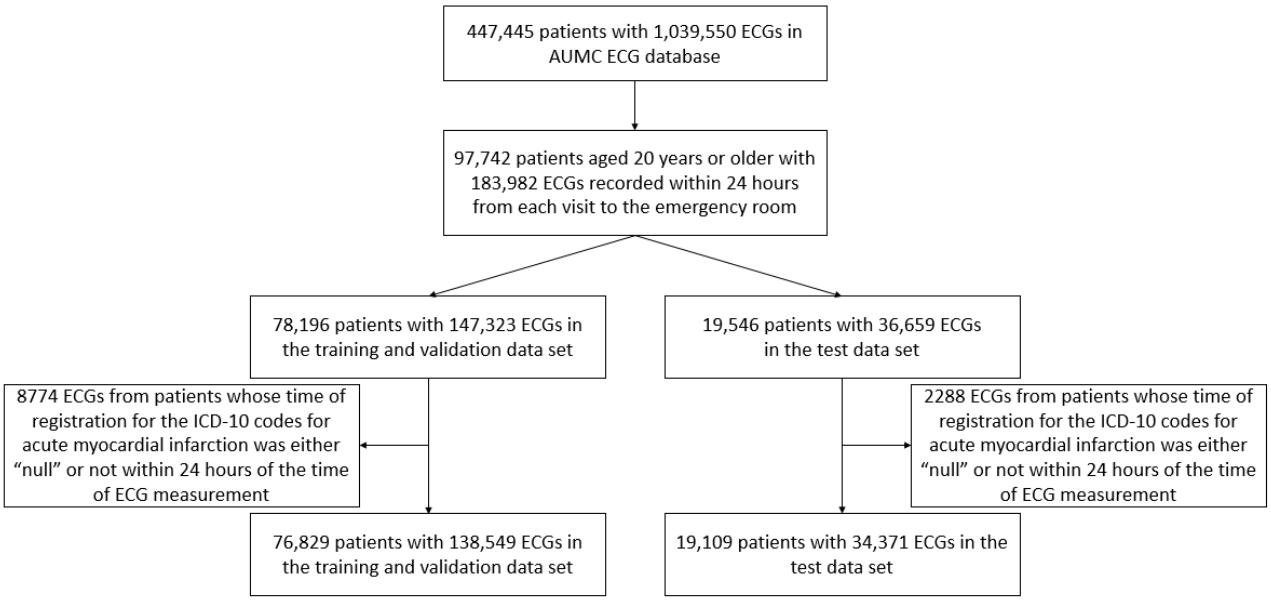


Table 1. Data set characteristics.

Characteristics	Training and validation (n=138,549)	Test (n=34,371)
Patients, n	76,829	19,109
Age (years), mean (SD)	59.00 (16.98)	59.00 (16.95)
Sex		
Male		
Electrocardiographs, n	75,552	18,426
Patients, n	40,662	10,043
Female		
Electrocardiographs, n	64,097	15,945
Patients, n	36,170	9066
Acute myocardial infarction, n	2465	554

Model Performance

Figures 4 and 5 show the receiver operating characteristic (ROC) and the precision-recall (PR) curves for the various target lead sets. The dots indicate the performance of the automatic ECG interpretations provided by the GE ECG analysis program. The sensitivity, specificity, PPV, and NPV of the first labeling criterion of the automatic ECG interpretation were 0.579, 0.866, 0.066, and 0.992, respectively. The corresponding values of the second labeling criterion of the automatic ECG interpretation were 0.765, 0.647, 0.034, and 0.996, respectively. Lead sets with 3 or more leads had a better performance than the automatic interpretations: their corresponding ROC and PR curves consistently lay above the corresponding dots of the automatic ECG interpretations. Similarly, the single-lead sets had worse performance than the automatic ECG interpretations: the corresponding ROC and PR curves lead sets lay below the corresponding dots of the automatic ECG interpretations. For the 2-lead sets, some of the ROC and PR curves lay above and some below the corresponding dots of the automatic ECG interpretations, which implied that not all the 2-lead sets had a better performance than the automatic interpretations.

Figure 4. ROC curves for the various target lead sets. The plot on the upper left shows the average ROC curves according to the number of leads. The solid lines depict the average ROC curves, and the shaded areas depict 1 SD of the ROC curves. The rest of the plots show the ROC curves for the 12-, 4-, 3-, 2-, and single-lead sets, respectively. In all plots, the performance of the automatic ECG interpretations is depicted as dots. AUROC: area under the receiver operating characteristic curve; ROC: receiver operating characteristic curve.

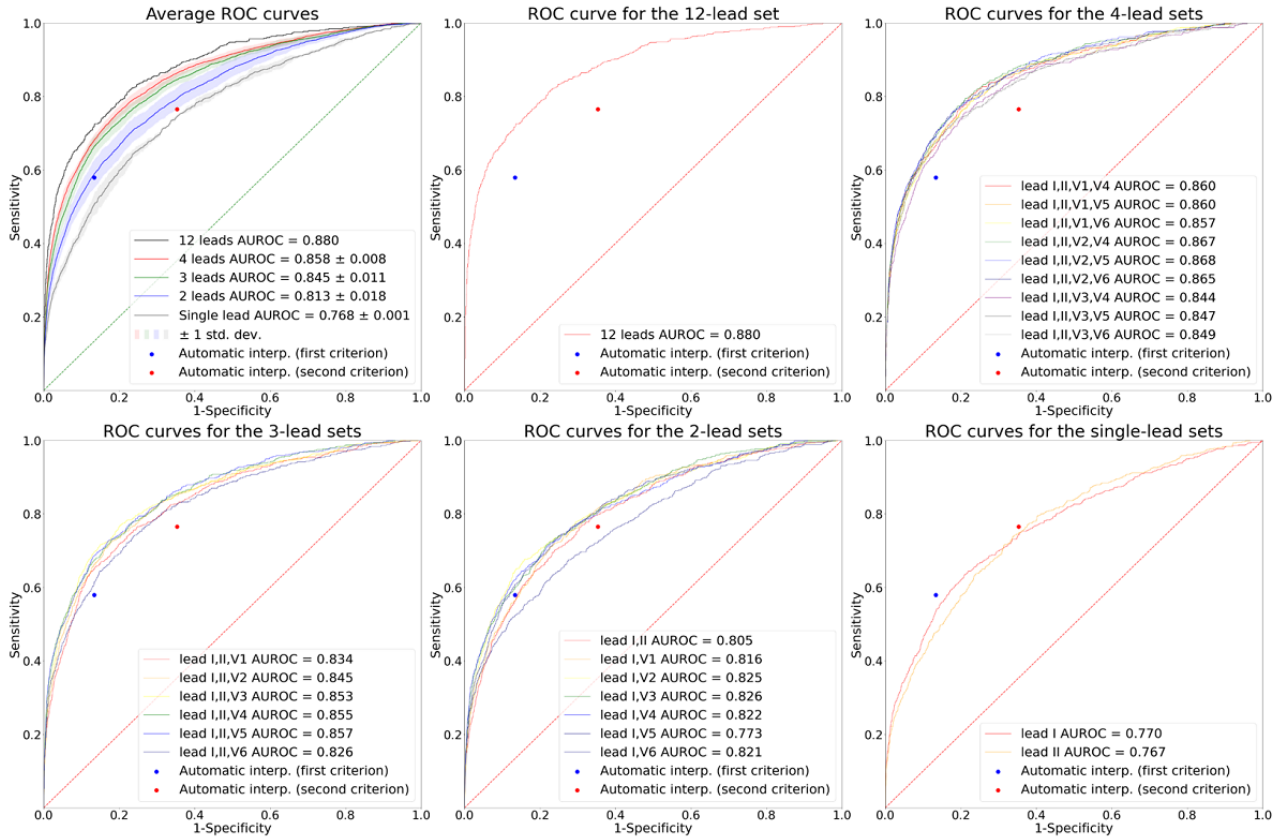
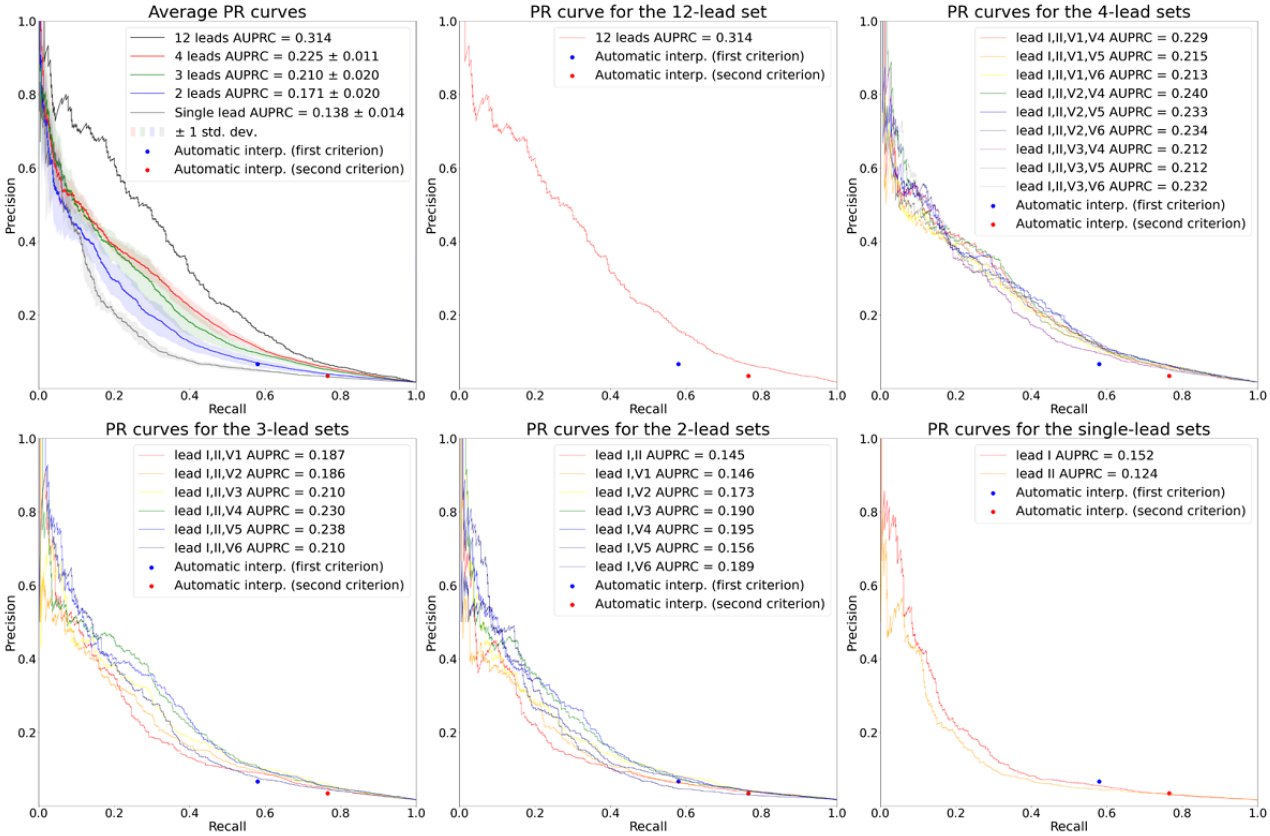


Figure 5. PR curves for the various target lead sets. The plot on the upper left shows the average PR curves according to the number of leads. The solid lines depict the average PR curves, and the shaded areas depict 1 SD of the PR curves. The rest of the plots show the PR curves for the 12-, 4-, 3-, 2-, and single-lead sets, respectively. In all plots, the performance of the automatic ECG interpretations is depicted as dots. AUPRC: area under the precision-recall curve; PR: precision-recall.



The average AUROCs for the 12-, 4-, 3-, 2-, and single-lead sets were 0.880, 0.858 (SD 0.008), 0.845 (SD 0.011), 0.813 (SD 0.018), and 0.768 (SD 0.001), respectively. The average AUPRCs for the 12-, 4-, 3-, 2-, and single-lead sets were 0.314, 0.225 (SD 0.011), 0.210 (SD 0.020), 0.171 (SD 0.020), and 0.138 (SD 0.014), respectively. These values indicate that the average AUROC and AUPRC increased as the number of leads increased. All the comparisons of AUROCs between ROC curves having the median AUROC from lead sets with different numbers of leads ("12-lead set" vs "4-lead set [leads I, II, V1, V5]" vs "3-lead set [leads I, II, V3]" vs "2-lead set [leads I, V6]" vs "single-lead set [lead I]") were statistically significant at a significance level of .05, as revealed through the DeLong test [25]. All the comparisons of AUROCs between ROC curves having the highest AUROC from lead sets with different numbers of leads ("12-lead set" vs "4-lead set [leads I, II, V2, V5]" vs "3-lead set [leads I, II, V5]" vs "2-lead set [leads I, V3]" vs "single-lead set [lead I]") were also statistically significant at a significance level of .05, as revealed through the DeLong test.

When we set the thresholds of the lead sets to match the specificity of the first labeling criteria of the automatic ECG interpretation (specificity=0.866), the 12-, 4-, and 3-lead sets demonstrated an average gain in sensitivity of 13.9%, 10.2% (SD 1.6%), and 8.5% (SD 2.7%), respectively (Table 2), compared to the automatic ECG interpretation, while maintaining a high NPV above 0.99. The results for the second labeling criteria (specificity=0.647) revealed average gains in sensitivity of 11.9%, 9.8% (SD 1.2%), and 8.1% (SD 1.5%) for the lead sets with 12, 4, and 3 leads, respectively (Table 2), while maintaining a high NPV above 0.99. The sensitivities of the 2-lead sets were, on average but not consistently, slightly higher than those of the automatic ECG interpretations when the thresholds of the 2-lead sets were set to match the specificities of the automatic ECG interpretations. Single-lead sets had lower sensitivities than the automatic ECG interpretations when the specificities were matched.

Table 2. Average sensitivity, positive predictive value, and negative predictive value according to the number of leads when the thresholds were set to match the specificity of the first or second labeling criteria of automatic electrocardiogram interpretation.

	Sensitivity		Positive predictive value		Negative predictive value	
	At specificity=0.866 (first labeling criteria)	At specificity=0.647 (second labeling criteria)	At specificity=0.866 (first labeling criteria)	At specificity=0.647 (second labeling criteria)	At specificity=0.866 (first labeling criteria)	At specificity=0.647 (second labeling criteria)
Automatic electrocardiogram interpretation	0.579	0.765	0.066	0.034	0.992	0.996
12-lead set	0.718	0.884	0.081	0.039	0.995	0.997
4-lead sets, mean (SD)	0.681 (0.016)	0.863 (0.012)	0.077 (0.002)	0.039 (0.001)	0.994 (0.000)	0.997 (0.000)
3-lead sets, mean (SD)	0.664 (0.027)	0.846 (0.015)	0.075 (0.003)	0.038 (0.001)	0.994 (0.000)	0.996 (0.000)
2-lead sets, mean (SD)	0.589 (0.038)	0.794 (0.030)	0.067 (0.004)	0.036 (0.001)	0.992 (0.001)	0.995 (0.001)
Single-lead sets, mean (SD)	0.505 (0.029)	0.745 (0.001)	0.058 (0.003)	0.033 (0.000)	0.991 (0.001)	0.994 (0.000)

Discussion

Principal Findings

In this study, we developed an AI model for detecting acute myocardial infarction by randomly masking each lead channel during the training phase and testing the model using various target ECG lead sets with the remaining lead channels masked. First, we found that the performances of lead sets with 3 or more leads compared favorably with that of the automatic ECG interpretations provided by the GE ECG analysis program, with a 8.1%-13.9% gain in sensitivity when the threshold was set to match the specificity of the automatic ECG interpretations, and with the ROC and PR curves lying above the corresponding dots of the automatic ECG interpretations. Only some of the 2-lead sets compared favorably with the automatic ECG interpretations. When only a single lead was evaluated, acute myocardial infarction could be underdiagnosed; thus, useful information from other leads could potentially be neglected. Indeed, single-lead sets performed worse than the automatic ECG interpretations.

Multiple-lead ECG is necessary for the accurate and robust detection of cardiac disorders, particularly acute myocardial infarction. Given that multiple-lead ECGs can be obtained by smartwatches only in an asynchronous manner, our results imply that multiple lead-based AI-enabled ECG algorithms can be implemented on these devices. Such implementation could facilitate timely diagnostics to enhance outcomes and reduce mortality among cardiovascular disease populations outside the hospital.

Second, we found that model performance generally increased as the number of leads increased (12-lead set: AUROC 0.880; 4-lead sets: AUROC 0.858, SD 0.008; 3-lead sets: AUROC 0.845, SD 0.011; 2-lead sets: AUROC 0.813, SD 0.018; single-lead sets: AUC 0.768, SD 0.001). With smartwatches, measuring additional leads would only take less than a minute, and the benefit of doing so would greatly outweigh the risk. In an emergency situation, we suggest measuring at least 3 leads (ie, I, II, and V5) and ideally more than 4 leads (ie, I, II, V2, and V5) to minimize the risk of failing to detect acute myocardial infarction occurring in a certain spatial location or direction.

Previous studies on automated diagnosis or classification of multiple-lead ECGs using AI have used synchronous ECG signals as input. The results from these studies are insufficient for the evaluation of the feasibility of multiple lead-based AI-enabled ECG algorithms on smartwatches since only asynchronous ECG signals can be obtained from smartwatches. To the best of our knowledge, our study is the first to utilize asynchronous ECG signals for AI model development. Future studies could aim at developing AI models with asynchronous ECG signals for detecting cardiac disorders other than acute myocardial infarction, such as cardiac arrhythmias or contractile dysfunctions.

Our study has important medical and economic impacts. First, our model can significantly reduce time to diagnosis, and consequently reduce time to reperfusion, which is the elapsed time between the onset of symptoms and reperfusion and is critical to the clinical outcome of the disease [26]. The ECG is commonly the first diagnostic test in the evaluation of myocardial infarction, and it should be acquired as early as practicable [27]. Traditionally, the bulky ECG equipment and the need for a trained physician for diagnosis have required the transfer of patients to hospitals, even in emergency situations. This practice greatly delays time to diagnosis, which would be most ideal if made directly in the field. With our model implemented on smartwatches, reliable preliminary diagnosis can be made even before contact with emergency services, thereby greatly reducing the time from the onset of symptoms to diagnosis. With the preliminary diagnosis already made, patients can be promptly triaged to the most appropriate form of treatment after accounting for geographical factors and available facilities [26]. The final diagnosis should be made by a trained physician after arriving at the appropriate facility, but with the aid of our model, the time required for the entire process can be greatly reduced. The threshold for a positive result from our model can be altered to balance between over- and undertriage. Second, our model has the potential to greatly reduce mortality and the related economic burden due to acute myocardial infarction. Untimely diagnosis or treatment results in increased myocardial damage and mortality. The extent of myocardial salvage is greatest if patients are reperfused in the first 3 hours after onset of symptoms [28]. For every 30-minute delay in coronary reperfusion, the relative 1-year mortality rate

increases by 7.5% [29]. Our model can reduce mortality due to acute myocardial infarction by facilitating timely diagnosis and reperfusion. Consequently, the economic cost caused by lost productivity from premature mortality due to acute myocardial infarction, which is estimated to be US $40.5 billion annually in the United States, can also be reduced [30]. Third, since our study indicates the feasibility of multiple lead-based AI-enabled ECG algorithms on smartwatches, it can promote the development of AI models with asynchronous ECG signals for detecting cardiac disorders other than acute myocardial infarction, thus accelerating market growth in this field.

Strengths and Limitations

Our study has several strengths. First, our model only takes ECG as input and does not require other additional clinical data. This implies that our model is highly applicable in real-world, real-time settings where no medical practitioners are available. Smartwatches are the only requirement for applying our model. Second, our model is theoretically implementable with all smartwatches, which further strengthens our study in terms of real-world applicability. That is, creating a mobile software app that activates the ECG hardware, instructs the wearer on how to measure the leads, preprocesses the measured leads to satisfy the input conditions of our AI model (eg, resampling the ECG to 250 Hz, snipping 2.5 seconds from each lead), and runs our AI model, would be sufficient for real-world implementation. We believe that with the aid of mobile app developers, such an app would not be technically difficult to develop. We leave this as a subject for further study. Third, we did not exclude ECGs on the basis of waveform abnormalities. This implies that our model is applicable regardless of ECG abnormalities, thereby greatly enhancing the generalizability to real-world settings. Fourth, our model was trained, validated, and tested with a very large data set of 172,920 ECGs recorded from 95,938 patients. A large enough data set can reduce overfitting to the training set, thus increasing generalizability to other data sets [31,32]. Fifth, as mentioned in the Methods section, our model is applicable to any lead combinations (eg, various 4-, 3-, 2-, and single-lead sets). This is because we randomly masked each lead channel on the attention matrices during the training phase. Thus, users would be able to choose any lead combination in accordance with their preferences or situation.

However, our study also has some limitations. First, our labeling method might be problematic. The diagnosis of acute myocardial infarction does not ensure that the patient's initial ECG in the emergency room would show explicit signs of acute myocardial infarction. Thus, some ECGs labeled as acute myocardial infarction in our data set might not explicitly show signs of acute myocardial infarction. Nevertheless, our model showed high performance, with our 12-lead set having an AUROC of 0.880. Second, the 12-lead set is not completely asynchronous. When grouped into 4 subsets with 3 leads in each subset, the ECGs are asynchronous intersubset-wise, while being synchronous intrasubset-wise. Thus, the maximum number of leads that can compose a completely asynchronous lead set in our study was 4. The diagnostic capacity of a model tested with 5 or more completely asynchronous lead sets needs to be evaluated in future studies. Third, our model cannot be deemed as a confirmatory test. The final confirmatory diagnosis should be made by a trained physician after the patient arrives in hospital. However, with the preliminary diagnosis made by our model, patients can be efficiently triaged to get the most appropriate form of treatment after accounting for geographical factors and available facilities, even before contact with emergency services. Finally, our model was not validated with external data sets. In future studies, external validation should be performed to ensure the reliability of our model in new environments.

Conclusions

In conclusion, this study shows the feasibility of multiple lead-based AI-enabled ECG algorithms on smartwatches for the automated diagnosis of cardiac disorders by developing an AI model for detecting acute myocardial infarction with asynchronous ECG signals. We also showed that measuring at least 3 leads, and ideally more than 4 leads, is necessary for accurate detection. Our results show that single-lead sets lack diagnostic performance. From our results, we look forward to the development of other AI models that detect various cardiac disorders using sequentially measured, asynchronous ECG leads from smartwatches. Such models, along with our model, can facilitate timely diagnostics to enhance outcomes and reduce mortality among various cardiac disease populations outside the hospital.

Acknowledgments

This work was supported by the Korea Medical Device Development Fund grant funded by the Korean government (the Ministry of Science and ICT; Ministry of Trade, Industry and Energy; Ministry of Health & Welfare; and Ministry of Food and Drug Safety) (project number 1711138152, KMDF_PR_20200901_0095). This study was also supported by a new faculty research seed money grant of Yonsei University College of Medicine for 2021 (2021-32-0044). We thank Medical Illustration & Design, part of the Medical Research Support Services of Yonsei University College of Medicine, for all artistic support related to this work.

Conflicts of Interest

CH and HSL declare that they have no competing interests. YS, YT, BTL, YL, and WB are employees of VUNO Inc. JHJ is an employee of Medical AI Inc. DY is an employee of BUD.on Inc. VUNO Inc, Medical AI Inc, and BUD.on Inc did not have any role in the study design, analysis, decision to publish, or the preparation of the manuscript. There are no patents, products in development, or marketed products to declare.

Multimedia Appendix 1
Standard 12-lead ECG report example.
[PNG File , 769 KB - jmir_v23i9e31129_app1.png]

Multimedia Appendix 2
Tested lead sets.
[DOCX File , 19 KB - jmir_v23i9e31129_app2.docx]

Multimedia Appendix 3
Architecture of the encoder.
[DOCX File , 20 KB - jmir_v23i9e31129_app3.docx]

References

1. Wright R, Keith L. Wearable Technology: If the Tech Fits, Wear It. J Electron Resour Med Libr 2014 Dec 03;11(4):204-216. [doi: 10.1080/15424065.2014.969051]
2. Lee SM, Lee D. Healthcare wearable devices: an analysis of key factors for continuous use intention. Serv Bus 2020 Oct 15;14(4):503-531. [doi: 10.1007/s11628-020-00428-3]
3. Kalantari M. Consumers' adoption of wearable technologies: literature review, synthesis, and future research agenda. IJTMKT 2017;12(3):274. [doi: 10.1504/ijtmkt.2017.089665]
4. Koltowski L, Balsam P, Glowczynska R, Rokicki JK, Peller M, Maksym J, et al. Kardia Mobile applicability in clinical practice: A comparison of Kardia Mobile and standard 12-lead electrocardiogram records in 100 consecutive patients of a tertiary cardiovascular care center. Cardiol J 2021;28(4):543-548 [FREE Full text] [doi: 10.5603/CJ.a2019.0001] [Medline: 30644079]
5. Benezet-Mazuecos J, García-Talavera CS, Rubio JM. Smart devices for a smart detection of atrial fibrillation. J Thorac Dis 2018 Nov;10(Suppl 33):S3824-S3827 [FREE Full text] [doi: 10.21037/jtd.2018.08.138] [Medline: 30631488]
6. Seshadri DR, Bittel B, Browsky D, Houghtaling P, Drummond CK, Desai MY, et al. Accuracy of Apple Watch for Detection of Atrial Fibrillation. Circulation 2020 Feb 25;141(8):702-703. [doi: 10.1161/CIRCULATIONAHA.119.044126] [Medline: 32091929]
7. Al-Zaiti SS, Callaway CW, Kozik TM, Carey MG, Pelter MM. Clinical Utility of Ventricular Repolarization Dispersion for Real-Time Detection of Non-ST Elevation Myocardial Infarction in Emergency Departments. J Am Heart Assoc 2015 Jul 24;4(7):e002057 [FREE Full text] [doi: 10.1161/JAHA.115.002057] [Medline: 26209692]
8. Meo M, Zarzoso V, Meste O, Latcu DG, Saoudi N. Spatial variability of the 12-lead surface ECG as a tool for noninvasive prediction of catheter ablation outcome in persistent atrial fibrillation. IEEE Trans Biomed Eng 2013 Jan;60(1):20-27. [doi: 10.1109/TBME.2012.2220639] [Medline: 23033326]
9. Samol A, Bischof K, Luani B, Pascut D, Wiemer M, Kaese S. Recording of Bipolar Multichannel ECGs by a Smartwatch: Modern ECG Diagnostic 100 Years after Einthoven. Sensors (Basel) 2019 Jun 30;19(13):2894 [FREE Full text] [doi: 10.3390/s19132894] [Medline: 31261981]
10. Samol A, Bischof K, Luani B, Pascut D, Wiemer M, Kaese S. Single-Lead ECG Recordings Including Einthoven and Wilson Leads by a Smartwatch: A New Era of Patient Directed Early ECG Differential Diagnosis of Cardiac Diseases? Sensors (Basel) 2019 Oct 10;19(20):4377 [FREE Full text] [doi: 10.3390/s19204377] [Medline: 31658713]
11. Spaccarotella CAM, Polimeni A, Migliarino S, Principe E, Curcio A, Mongiardo A, et al. Multichannel Electrocardiograms Obtained by a Smartwatch for the Diagnosis of ST-Segment Changes. JAMA Cardiol 2020 Oct 01;5(10):1176-1180 [FREE Full text] [doi: 10.1001/jamacardio.2020.3994] [Medline: 32865545]
12. Avila CO. Novel Use of Apple Watch 4 to Obtain 3-Lead Electrocardiogram and Detect Cardiac Ischemia. Perm J 2019;23:19-25 [FREE Full text] [doi: 10.7812/TPP/19-025] [Medline: 31314734]
13. Muhlestein JB, Le V, Albert D, Moreno FL, Anderson JL, Yanowitz F, et al. Smartphone ECG for evaluation of STEMI: results of the ST LEUIS Pilot Study. J Electrocardiol 2015;48(2):249-259. [doi: 10.1016/j.jelectrocard.2014.11.005] [Medline: 25601407]
14. Ribeiro AH, Ribeiro MH, Paixão GMM, Oliveira DM, Gomes PR, Canazart JA, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nat Commun 2020 Apr 09;11(1):1760 [FREE Full text] [doi: 10.1038/s41467-020-15432-4] [Medline: 32273514]
15. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. Lancet 2019 Sep 07;394(10201):861-867. [doi: 10.1016/S0140-6736(19)31721-0] [Medline: 31378392]
16. Chen Y, Liu C, Tseng V, Hu YF, Chen SA. Large-scale Classification of 12-lead ECG with Deep Learning. 2019 Presented at: IEEE EMBS International Conference on Information Technology Applications in Biomedicine (ITAB); May 19-22, 2019; Chicago, IL. [doi: 10.1109/bhi.2019.8834468]

17. Kwon J, Lee SY, Jeon K, Lee Y, Kim K, Park J, et al. Deep Learning-Based Algorithm for Detecting Aortic Stenosis Using Electrocardiography. J Am Heart Assoc 2020 Apr 07;9(7):e014717 [FREE Full text] [doi: 10.1161/JAHA.119.014717] [Medline: 32200712]
18. Adedinsewo D, Carter RE, Attia Z, Johnson P, Kashou AH, Dugan JL, et al. Artificial Intelligence-Enabled ECG Algorithm to Identify Patients With Left Ventricular Systolic Dysfunction Presenting to the Emergency Department With Dyspnea. Circ Arrhythm Electrophysiol 2020 Aug;13(8):e008437. [doi: 10.1161/CIRCEP.120.008437] [Medline: 32986471]
19. Jang J, Kim TY, Yoon D. Effectiveness of Transfer Learning for Deep Learning-Based Electrocardiogram Analysis. Healthc Inform Res 2021 Jan;27(1):19-28 [FREE Full text] [doi: 10.4258/hir.2021.27.1.19] [Medline: 33611873]
20. Chung D, Choi J, Jang J, Kim TY, Byun J, Park H, et al. Construction of an Electrocardiogram Database Including 12 Lead Waveforms. Healthc Inform Res 2018 Jul;24(3):242-246 [FREE Full text] [doi: 10.4258/hir.2018.24.3.242] [Medline: 30109157]
21. Goldberger E. A simple, indifferent, electrocardiographic electrode of zero potential and a technique of obtaining augmented, unipolar, extremity leads. American Heart Journal 1942 Apr;23(4):483-492. [doi: 10.1016/s0002-8703(42)90293-x]
22. Einthoven W. Weiteres über das Elektrokardiogramm. Pflüger Arch 1908 May;122(12):517-584. [doi: 10.1007/bf01677829]
23. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 Presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27-30, 2016; Las Vegas, NV. [doi: 10.1109/cvpr.2016.90]
24. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, et al. Attention is all you need. arXiv Preprint posted online June 12, 2017. [FREE Full text]
25. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988 Sep;44(3):837-845. [Medline: 3203132]
26. Brown AJ, Ha FJ, Michail M, West NEJ. Prehospital Diagnosis and Management of Acute Myocardial Infarction. In: Primary Angioplasty: A Practical Guide. Singapore: Springer; 2018.
27. Link MS, Berkow LC, Kudenchuk PJ, Halperin HR, Hess EP, Moitra VK, et al. Part 7: Adult Advanced Cardiovascular Life Support: 2015 American Heart Association Guidelines Update for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation 2015 Nov 03;132(18 Suppl 2):S444-S464. [doi: 10.1161/CIR.0000000000000261] [Medline: 26472995]
28. Gersh BJ, Stone GW, White HD, Holmes DR. Pharmacological facilitation of primary percutaneous coronary intervention for acute myocardial infarction: is the slope of the curve the shape of the future? JAMA 2005 Feb 23;293(8):979-986. [doi: 10.1001/jama.293.8.979] [Medline: 15728169]
29. De Luca G, Suryapranata H, Ottervanger JP, Antman EM. Time delay to treatment and mortality in primary angioplasty for acute myocardial infarction: every minute of delay counts. Circulation 2004 Mar 16;109(10):1223-1225. [doi: 10.1161/01.CIR.0000121424.76486.20] [Medline: 15007008]
30. Bishu KG, Lekoubou A, Kirkland E, Schumann SO, Schreiner A, Heincelman M, et al. Estimating the Economic Burden of Acute Myocardial Infarction in the US: 12 Year National Data. Am J Med Sci 2020 May;359(5):257-265. [doi: 10.1016/j.amjms.2020.02.004] [Medline: 32265010]
31. Karystinos GN, Pados DA. On overfitting, generalization, and randomly expanded training sets. IEEE Trans Neural Netw 2000;11(5):1050-1057. [doi: 10.1109/72.870038] [Medline: 18249833]
32. Yip KY, Gerstein M. Training set expansion: an approach to improving the reconstruction of biological networks from limited and uneven reliable interactions. Bioinformatics 2009 Jan 15;25(2):243-250 [FREE Full text] [doi: 10.1093/bioinformatics/btn602] [Medline: 19015141]

Abbreviations

AI: artificial intelligence
AUMC: Ajou University Medical Center
AUPRC: area under the precision-recall curve
AUROC: area under the receiver operating characteristic curve
CNN: convolutional neural network
ECG: electrocardiogram
GE: General Electric
NPV: negative predictive value
PPV: positive predictive value
PR: precision-recall
ROC: receiver operating characteristic

Edited by G Eysenbach; submitted 23.06.21; peer-reviewed by HS Kim, Z Su; comments to author 15.07.21; revised version received 27.07.21; accepted 01.08.21; published 10.09.21.

Please cite as:
Han C, Song Y, Lim HS, Tae Y, Jang JH, Lee BT, Lee Y, Bae W, Yoon D
Automated Detection of Acute Myocardial Infarction Using Asynchronous Electrocardiogram Signals—Preview of Implementing Artificial Intelligence With Multichannel Electrocardiographs Obtained From Smartwatches: Retrospective Study
J Med Internet Res 2021;23(9):e31129
URL: https://www.jmir.org/2021/9/e31129
doi:10.2196/31129
PMID:34505839

Original Paper

Identifying Enablers of Participant Engagement in Clinical Trials of Consumer Health Technologies: Qualitative Study of Influenza Home Testing

Spurthy Dharanikota[1], MBA; Cynthia M LeRouge[1], PhD; Victoria Lyon[2], MPH; Polina Durneva[1], BA; Matthew Thompson[2], MPH, MD, PhD

[1]Department of Information Systems and Business Analytics, Florida International University, Miami, FL, United States

[2]Primary Care Innovation Lab, Department of Family Medicine, University of Washington, Seattle, WA, United States

Corresponding Author:
Spurthy Dharanikota, MBA
Department of Information Systems and Business Analytics
Florida International University
11200 SW 8th Street
Miami, FL, 33199
United States
Phone: 1 3057812536
Email: sdhar006@fiu.edu

Abstract

Background: A rise in the recent trend of self-managing health using consumer health technologies highlights the importance of efficient and successful consumer health technology trials. Trials are particularly essential to support large-scale implementations of consumer health technologies, such as smartphone-supported home tests. However, trials are generally fraught with challenges, such as inadequate enrollment, lack of fidelity to interventions, and high dropout rates. Understanding the reasons underlying individuals' participation in trials can inform the design and execution of future trials of smartphone-supported home tests.

Objective: This study aims to identify the enablers of potential participants' trial engagement for clinical trials of smartphone-supported home tests. We use influenza home testing as our instantiation of a consumer health technology subject to trial to investigate the dispositional and situational enablers that influenced trial engagement.

Methods: We conducted semistructured interviews with 31 trial participants using purposive sampling to facilitate demographic diversity. The interviews included a discussion of participants' personal characteristics and external factors that enabled their trial engagement with a smartphone-supported home test for influenza. We performed both deductive and inductive thematic analyses to analyze the interview transcripts and identify enabler themes.

Results: Our thematic analyses revealed a structure of dispositional and situational enablers that enhanced trial engagement. Situationally, clinical affiliation, personal advice, promotional recruitment strategies, financial incentives, and insurance status influenced trial engagement. In addition, digital health literacy, motivation to advance medical research, personal innovativeness, altruism, curiosity, positive attitude, and potential to minimize doctors' visits were identified as the dispositional enablers for trial engagement in our study.

Conclusions: We organized the identified themes for dispositional and situational enablers of trial engagement with a smartphone-supported home test into a research framework that can guide future research as well as the trial design and execution of smartphone-supported home tests. We suggest several trial design and engagement strategies to enhance the financial and scientific viability of these trials that pave the way for advancements in patient care. Furthermore, our study also offers practical strategies to trial organizers to enhance participants' enrollment and engagement in clinical trials of these home tests.

(J Med Internet Res 2021;23(9):e26869) doi:10.2196/26869

KEYWORDS

consumer health care technologies; CHTs; smartphone-supported home tests; Smart-HT; premarket clinical trials; trial engagement; at-home diagnostic testing; mobile phone

Introduction

Background

Consumer health care technologies (CHTs) that could potentially transform the health care industry require proper assessment and evaluation before large-scale use. In 2013, the US Food and Drug Administration (FDA) issued guidelines named *Mobile Medical Applications Guidance for Industry and Food and Drug Administration Staff* to regulate a rapidly growing number of mobile consumer health apps [1]. Furthermore, the FDA defined *mobile medical applications* as medical devices that are mobile apps or as an accessory to regulated medical devices that require FDA approval [1]. Personal health information is increasingly intertwined with clinical applications. In addition, CHTs are increasingly being looked at as a means to support the effective execution of home-based medical diagnostic tests [2-4] and medical procedure preparation kits (eg, colonoscopy preparation) [5]. As a result, an increasing number of CHTs will require assessment and evaluation (eg, clinical and premarket trials). Premarket trials are often an integral part of health care product development, even when regulatory assessments are not formally required. Developers and vendors may seek to pretest CHTs with target users for usability, feasibility, and economic viability [6]. Specifically, they provide ways to evaluate the CHT's impact on health outcomes and assess efficacy and safety before releasing it to the general population [7]. Trials can also provide opportunities for refinements to improve CHTs before their availability for commercial use [8]. Overall, trial engagement involving CHT is an interesting user behavior that deserves more attention in research to facilitate the assessment of health care innovations for the suitability for general use.

We define a potential trial participant as someone willing to try the technology before it becomes available for public and wider-scale use. Trial engagement refers to a potential trial participant successfully completing a trial. Research generally reports inadequate trial engagement as a significant challenge to CHT trial success [9]. For example, potential trial participants may perceive trial participation as a risky activity and may not be willing to enroll [10]. In addition, potential trial participants may determine that incentives or compensation are inadequate compared with their perceived risk of participating in the trial [11]. Even after enrollment, the participants' lack of engagement in trial activities and high dropout rates can impede trial success [12]. In the context of CHT, studies show that inadequate enrollment and engagement can be very costly for CHT manufacturers and researchers involved [13-15]. By and large, trial engagement is crucial for determining the validity, feasibility, and success of CHT clinical trials [16-18].

Unfortunately, there is a lack of research to provide guidance regarding people's perceptions and attitudes toward engaging in CHT trials. Current research that addresses engagement in clinical trials focuses primarily on evaluating attitudes toward traditional in-person trials of pharmaceutical interventions such as pharmaceutical trials [19] and randomized control trials for interventions and care processes related to serious health conditions such as cancer [20-22], cardiovascular disease [23,24], and lung disease [25].

Indeed, there is a shortage of behavioral research related to trials that evaluate engagement in CHT trials involved in the clinical or self-care process [26]. Within the realm of CHT, health apps offer tools, procedures, and communications to support mobile health care practices [27,28]. Moreover, very few studies in the medical informatics literature have explored clinical trials involving health apps [29], and few of the existing studies aim to identify enablers for health apps. Consumer perceptions or beliefs about the trial attributes that influence clinical trial engagement are known as enablers [30]. One notable exception in the medical informatics literature includes a 2-pronged research study, which proposed a recruitment framework for eHealth clinical trials using cost-effective and time-efficient trial recruitment strategies [18,31]. However, the study did not consider the attitudes of potential trial participants toward the trial. Another notable exception is a study that examined the perceptions and experiences of women engaging in a digital technology–based clinical trial in the context of physical activity interventions [32]. The study identified critical factors that enabled participants' continued engagement in physical activity after the trial period but did not investigate the enabling factors for their trial participation. In a different study context, Cohen et al [33] tried to understand patient compliance during 2 digital trials involving 2 pathologies, Parkinson disease and Huntington disease, each lasting 6 months. The study measured patient compliance metrics, namely, daily app-based medication reporting during the 6-month trials; however, it did not attempt to understand the factors that enabled participants' trial engagement.

Identifying engagement enablers can inform future trial design and execution, thereby enhancing the scientific and financial feasibility of CHT trials. Among the increasing technological capabilities associated with CHT innovation is using a smartphone to support home-based diagnostic tests, referred to as smartphone-supported home tests (Smart-HT). This study focuses on trial engagement in the context of CHT used for home-based diagnostics or Smart-HT (ie, diagnostic tests that a user can carry out with the support of a smartphone) [34-37]. Smart-HT usually requires FDA approval and has growing market potential and patient interest. Smart-HT comprises a diagnostic testing kit and a software component (eg, a mobile app providing one or more of the following features: instructions, education about the test or health situation, an indication of results, and results messaging). The user is required to simultaneously perform the necessary physical procedures to complete the test and understand the software component to use Smart-HT to achieve the desired outcome—diagnosis and enablement for the next steps in care.

The growing interest in at-home diagnostics, combined with their potential implications for future pandemic preparedness, merits the focused study of Smart-HT. The at-home diagnostic market is predicted to value >US $6 billion by the end of 2027; the market is anticipated to have a compound annual growth rate of 3.98% during the forecast period of 2020 and 2027 [38]. A recent survey (the 2018 Deloitte Center for Health Solutions) indicated that most people are interested in engaging with new care channels, such as at-home diagnostic testing. Overall, 51% of respondents reported being comfortable using at-home tests

for their current health concerns and identifying potential future health issues. Furthermore, 44% of respondents reported being comfortable using mobile apps connected to at-home diagnostic tests to track and monitor their health trends [39].

Objective

The purpose of this study is to identify the enablers of potential participants' trial engagement in clinical trials of CHT, in particular, Smart-HT trials. We use flu@home, a Smart-HT in trial that comprises a test kit and a mobile app to facilitate home-based testing for the diagnosis of influenza, as our instantiation of CHT to address the following research question: What enablers influence trial engagement for Smart-HT?

The paper is structured as follows. First, we discuss the Smart-HT context of the study (flu@home pilot) used to inform the enablers of Smart-HT trial engagement. We provide details on the interview phase of the pilot, where qualitative data were collected and analyzed to inform the results of this study. Second, we present our results in the form of a Smart-HT trial engagement research framework based on our evidence and provide a discussion of the identified enablers for Smart-HT trial engagement. Finally, we discuss the implications of this study for trial organizers and researchers.

Methods

Smart-HT Context: flu@home

This study leverages the context of influenza (or *flu*) using an instantiation of Smart-HT called flu@home, which has the potential to allow individuals to self-diagnose influenza. Influenza, a contagious respiratory infection caused by influenza viruses, is a serious global health threat with an estimated 1 billion cases each year, of which 18% of people die worldwide [40]. Influenza imposes a substantial economic burden, including health care costs and productivity losses, accounting for US $87 billion in the United States alone [41]. Infectious diseases, including influenza, call attention to the need for speedy diagnosis and patient treatment or isolation [3]. In addition, the drugs available for treating influenza are most effective when used within 48 hours of identifying symptoms [42]. Smart-HT offers an expedient and effective way of diagnosing influenza at home without exposing others to the virus (eg, at a physicians's office).

The Smart-HT in this study, flu@home, comprises two components: a test kit (adapted from the Quidel QuickVue Influenza A+B test) and a mobile app to facilitate testing (Figure 1). The at-home rapid diagnostic testing kit includes materials for an individual to self-test for influenza using a low nasal swab. The testing involves taking a sample from inside one's nose and processing it using a lateral flow assay (Quidel Corporation). The mobile app was available for participants to download on the iOS platform for use on smartphones and tablets. Multimedia Appendix 1 contains detailed information about the technical details associated with the flu@home mobile app.

The flu@home pilot study included two phases: the comparative accuracy phase and the interview phase (Figure 2). The study design was approved by the University of Washington institutional review board STUDY00007627.

Figure 1. Screenshots of flu@home app.

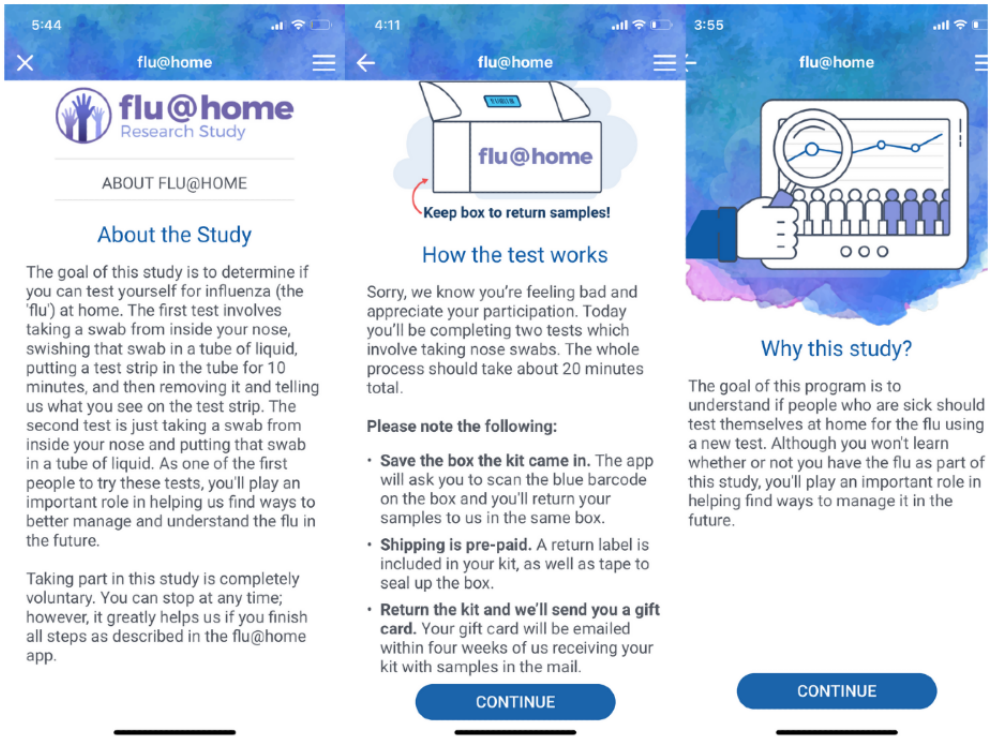
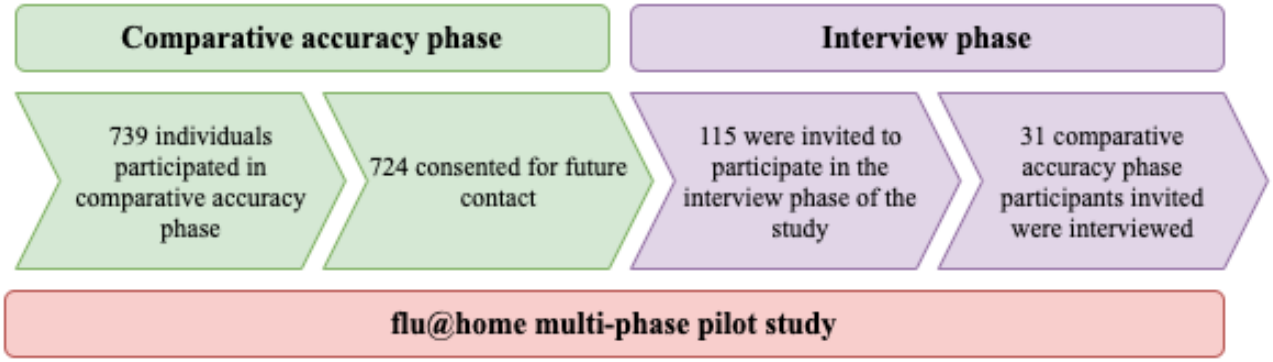


Figure 2. Phases of flu@home pilot study.



During the comparative accuracy phase, participants engaged with flu@home. The inclusion criteria for the comparative accuracy phase of the flu@home pilot study involved eligible participants who were aged ≥18 years, spoke English, had an iPhone or iPad, and had an influenza-like illness, defined as the presence of a cough and at least one or more of the following symptoms: fever, chills or sweats, muscle or body aches, or feeling tired or more tired than usual. Two videos were provided to participants as part of this study: one for recruitment purposes (study overview) and the other for providing step-by-step instructions to engage in the trial (Multimedia Appendices 2 and 3 contain the videos). Trial participants (N=739) completed the following workflow: (1) downloaded the flu@home app to enroll in the trial and order a test kit, (2) completed the consent form and a survey about their symptoms and exposure risk, (3) received a test kit by mail and followed the app instructions to conduct the rapid diagnostic test, and (4) shipped a second nasal swab sample to the research laboratory for reference standard testing. Of the 739 individuals who participated in the comparative accuracy phase of the flu@home trial, 97.9% (724/739) trial participants consented to be contacted for further phases of research.

During the interview phase, we conducted semistructured interviews with a sample of individuals who completed the comparative accuracy phase. We used phenomenology to guide this phase to construct a rich understanding of participants' experiences with flu@home and identify the enablers of Smart-HT trial engagement [43]. Phenomenology, described as the science of phenomena, explores human experience to elicit meanings for individuals through the analysis of their experiences and perceptions [44]. In line with this paradigm, we conducted in-depth interviews that covered experiences and perception questions related to (1) personal characteristics that enabled participants to complete flu@home trials; (2) attitudes toward health and medical research; (3) perceptions toward flu@home, mobile apps, and the overall trial; and (4) factors that enabled their Smart-HT trial engagement.

To fulfill the purposes of this study, we reported results from the interview phase related to understanding the enablers of potential participants' Smart-HT trial engagement. In the following sections, we further describe the data collection and data analysis procedures pertinent to the interview phase.

Data Collection

To meet the eligibility (inclusion) criteria for the interviews, individuals must have participated in the comparative accuracy phase of the pilot and consented in the flu@home app to be contacted for future study efforts. However, participation in the interview phase was not dependent on individuals completing the final steps in the flu@home workflow (ie, it is possible that some participants never mailed their sample to the laboratory; however, they would still be eligible if they opted in for future contact in the app). The flu@home test results were deidentified, so we were unable to determine how many eligible participants may not have performed the final step of mailing back their test results for analysis.

We consulted the literature to determine the total number of participants for the interviews. Data saturation, the point at which additional data collection no longer generates a new understanding [45], is the most common guiding principle for determining the total number of participants in qualitative research [46]. Prior interview study designs reached saturation in as few as 10-40 interviews [47-52]. In considering precedence and study design, we determined a minimum of 20 trial participants to be an adequate and appropriate target number of interviews.

Participants were invited to participate in the interviews via email and offered a US $25 gift card for completing the interview. To ensure diversity in the interviewee sample, we recruited participants in three waves. We followed purposive sampling to maintain diversity in the representation of trial participants in terms of age, race, and geographic location. First, we sorted the trial participants into age groups (18-24 years, 25-34 years, 35-44 years, 45-64 years, and ≥65 years) and randomly selected participants from each age group to send interview invitations. In the two subsequent waves of recruitment, we adjusted the proportion of participants recruited from each age group to ensure sample representation from all age groups. Attempts to fulfill purposeful sampling resulted in 115 invitations to participate and 31 actual interviews. Table 1 details the demographics of the trial participants who were invited and interviewed for the study.

Table 1. Demographics of interview participants in the study.

Characteristics	Invited to participate in study	Completed the interview
Total, n (%)	115 (100)	31 (100)
Age (years), n (%)		
18-24	12 (10.4)	3 (10)
25-34	34 (29.6)	6 (19)
35-44	38 (33)	11 (36)
45-64	21 (18.3)	8 (26)
≥65	10 (8.7)	3 (10)
Ethnicity, n (%)		
White individuals	78 (67.8)	21 (68)
Black or African American	10 (8.7)	6 (19)
Asian	8 (7)	0 (0)
Native Hawaiian or Other Pacific Islander	1 (0.9)	1 (3)
American Indian or Alaska Native	16 (13.9)	1 (3)
N/A[a], other, or prefer not to say	2 (1.7)	2 (6)
Geographic representation, n (%)		
West	43 (37.4)	14 (45)
Midwest	21 (18.3)	5 (16)
Southwest	2 (1.7)	1 (3)
Northeast	32 (27.8)	4 (13)
Southeast	17 (14.8)	7 (23)

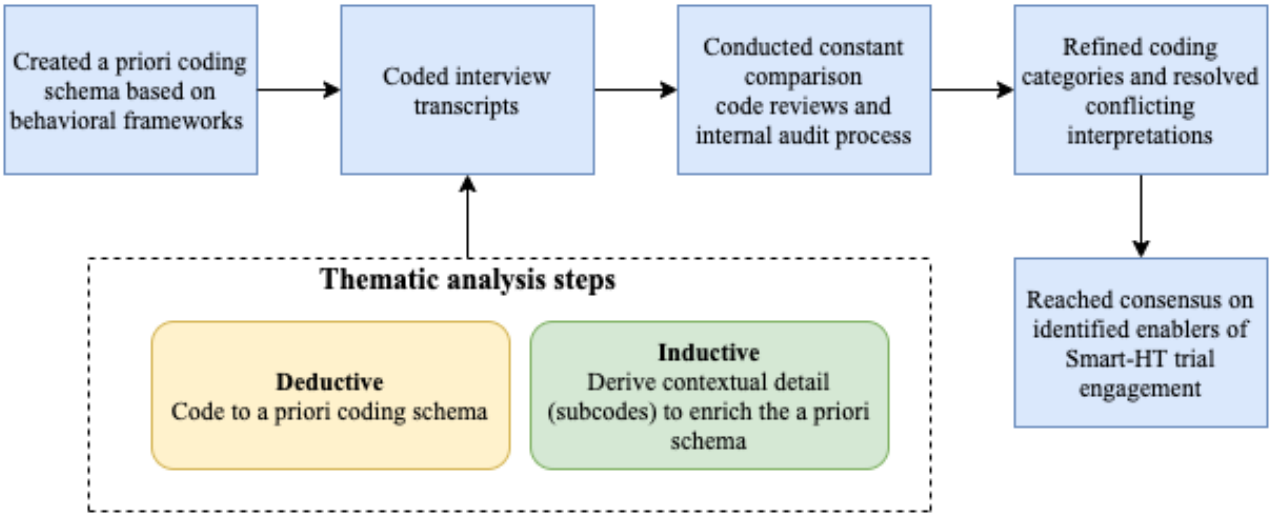
[a]N/A: not applicable.

Semistructured interviews were conducted via the web-based video conferencing tool Zoom from August 12, 2019, to December 19, 2019. Zoom has been an effective means of data collection in various health-related studies [53]. Participants had the option of being on or off camera. Each interview lasted between 40 and 60 minutes. Of the 3 research team members (in a rotating fashion), 2 were present during the interviews, providing direct experience with the actual interviews to the 3 team members subsequently engaged in the analysis process. Of the 2 team members participating in the interview, 1 served as the lead interviewer, whereas the other served the role of *active listener* and notetaker. The *active listener* was invited to ask any follow-up questions needed for clarification during the interview. The 3 team members alternated in serving the primary interviewer and listener roles for each interview. The 2 team members participating in the interview debriefed immediately after each interview to review highlights relevant to the study, establishing early key points to consider for coding purposes. Interviews were recorded and transcribed, and all personal identifiers were removed before analysis. We uploaded deidentified interview transcripts to Dedoose version 7.0.23, a software for qualitative data analysis.

Data Analysis

We conducted a thematic analysis [54] to code the deidentified transcripts. Thematic analysis is defined as a method for identifying, analyzing, organizing, describing, and reporting themes within a data set [55]. To ensure the validity and reliability of the thematic analysis results, we followed the Lincoln and Guba [50] criteria for conducting qualitative research. Specifically, we established confirmability, dependability, and credibility through procedures such as researcher triangulation, code reviews, expert feedback, and resolution meetings [50,56]. We took a hybrid approach to thematic analysis by including both deductive and inductive coding [57]. Figure 3 illustrates this process.

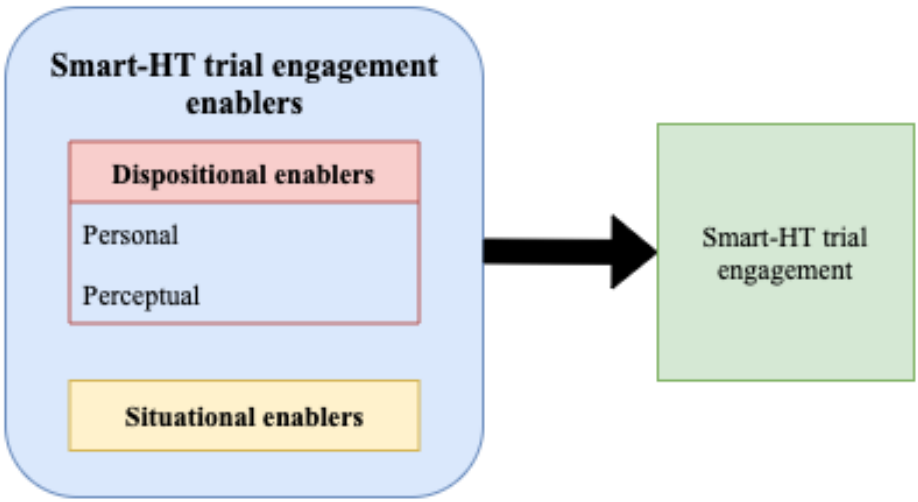
Figure 3. The qualitative thematic analysis process that guided the deductive and inductive coding of the interview transcripts. Smart-HT: smartphone-supported home tests.



We developed the a priori coding schema based on widely used behavioral frameworks that explain an individual's behaviors and actions [58-61] and categorized the factors into dispositional and situational enablers. Specifically, we relied on attribution theory, the digital health engagement model proposed by O'Connor et al [61], and the technology acceptance model [60] to identify generalized enabler constructs that align with our context to include in our a priori coding schema. On the basis of this foundation, our resulting a priori coding schema recognized the possibility of personal, perceptional, and situational factors. Situational enablers referred to external events or environmental factors that contributed to participants' Smart-HT trial engagement [58,59]. Personal enablers related to an individual's abilities, traits, and beliefs, whereas perceptional enablers referred to their perceptions that enabled their Smart-HT trial engagement. To further refine our schema and acknowledge relationships in existing research, we decided to group the personal and perceptional enablers we find in the data as *dispositional*.

Dispositional enablers referred to trial participants' behaviors and perceptions, whereas situational enablers were independent of participants' behavior. We used the skeletal a priori coding schema (Figure 4) as a starting point for the thematic analysis. Using the schema facilitated an initial, agreed-upon conceptual basis for the research team to begin analysis and then determine the operational definitions for each identified variable based on the schema [62,63].

Figure 4. A priori coding schema for thematic analysis. Smart-HT: smartphone-supported home tests.



We started our thematic analysis with a deductive approach [57]. During this analysis step, 2 of our team members independently coded the interview transcripts using the a priori coding schema (Figure 4). We were also open to additional high-level constructs emerging from our study that would extend the a priori schema. To capture key content within each a priori theme (ie, parent code), we extended and enriched the a priori schema codes (ie, parent codes) with inductively identified subcodes (ie, child codes). Explicitly, as we encountered new content during the coding process, we updated our coding schema with additional subcodes to further define the general themes in the a priori schema to the Smart-HT context. These practices further defined and specified associated constructs in

the a priori coding schema and continued until thematic saturation of the Smart-HT trial enablers was reached.

We used the constant comparison method of analysis [63,64] to refine and triangulate the coding. The constant comparison procedure included 2 coders and an internal auditor serving. A research team member with >17 years of qualitative research methods expertise, familiar with the study constructs and context but not engaged in the detailed coding process, served as our internal auditor to ensure expertise, understanding, and independence from the specific assignment of quotes to codes and initial code labels. The team held recurring meetings to compare initial coding determinations to ensure agreement on high-level codes and ratify the addition of subcodes (child codes) supporting the constructs in the initial coding schema. The team also refined the code labeling (names of codes) as needed to ensure understandable terms and themes. The ratification of subcodes included a general review of supporting quotes to support discussion points.

Although various studies finalize coding on this consensus [65], we chose to extend our rigor to support our findings further. Specifically, as a final step, our internal auditor also reviewed the holistic structure, labeling, and syntax of the final reconciled schema and performed a code review of 100% of the coded quotes to ensure alignment with the final coding structure. Any identified issues were brought back to the collective team for final disposition. As noted in prior literature, a feedback process, such as this internal audit process, can serve as a crucial *reader resonation* strategy to further enhance the validity of qualitative findings [66].

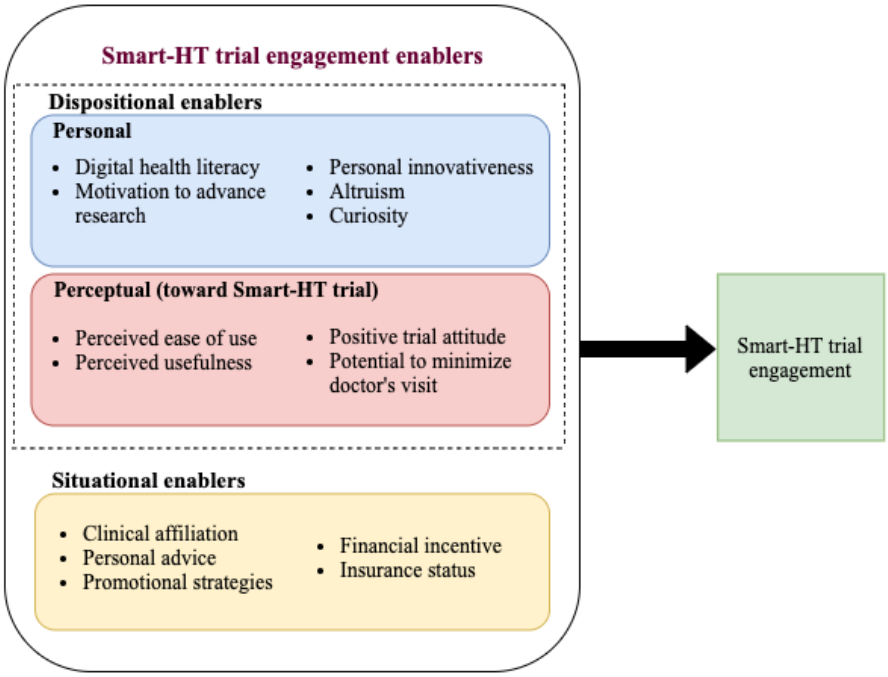
There were no noted major conceptual differences in the constant comparison and final internal review process. Reconciliation and arbitration were primarily focused on combining some subcodes or refining the labels and definitions of added codes. In particular, the team sought to refine the labels, where appropriate, to terms used in generally related research. In completing the process, there were no exceptions to reaching consensus, and, thus, code reviews and resolution meetings established the confirmability and dependability of the findings. We present our results as a research framework for enablers of Smart-HT trial engagement in the following section.

Results

Overview of Results

We synthesized our results into a research framework (Figure 5) that presents pertinent dispositional and situational enablers for Smart-HT trial engagement. This section details our thematic analysis findings under each of the broader enabler themes (dispositional and situational). We include evidence trace tables providing a representative participant quote to demonstrate the study's identified themes, as found in the data in Multimedia Appendix 4. Tables S1, S2, and S3 provided in Multimedia Appendix 4 present the evidence trace table for personal enablers and perceptional and situational enablers, respectively. The research team collectively selected the quotes from the multiple options coded as both a good conceptual fit and easy for the reader to understand outside of providing extensive interview text and context.

Figure 5. Evidence-based smartphone-supported home tests trial engagement research framework. Smart-HT: smartphone-supported home tests.



Dispositional Enablers

Our results supported characterizing dispositional enablers into personal and perceptional.

Personal Enablers

Our results identified the following personal enablers for Smart-HT trial engagement: digital health literacy, motivation to advance medical research, innovativeness, curiosity, and

altruism. Table S1 in Multimedia Appendix 4 presents the evidence trace table for personal enablers identified in the study.

In general, the interviewees in this study were digital health literate. Interviewees commonly expressed their ability to seek, find, understand, and apply health information from a wide range of web-based sources such as Google, health care websites, and WebMD, among others. Moreover, some interviewees even noted their familiarity with using wearables (eg, Apple Watch) and telemedicine for taking care of their health, which evidenced their adequate degree of *digital health literacy.*

Our inductive thematic analysis emphasized *motivation to advance medical research* as a prominent enabling factor in trial engagement for Smart-HT. The data revealed that participants hoped that their feedback after the flu@home trial might benefit vaccine development in the future. An interviewee explained that offering feedback and data that further medical research is vital for the greater good.

Remarkably, nearly all the interview participants acknowledged that they were *inherently innovative*. Their perception of Smart-HT as an innovative and transformative technology seemingly encouraged them to engage in the trial. Our analysis further identified *curiosity* in a relatively new technology (ie, a Smart-HT) as a personal enabler for individuals to engage in the trial. In addition to being curious about the trial and technology, interviewees indicated how their *altruistic notions* encouraged them to participate in the trial.

Perceptual Enablers (Toward Smart-HT Trial)

Our analysis revealed the importance of trial participants' perceptions of Smart-HT in their trial engagement for Smart-HT. The perceptual enablers identified in the study (Table S2 in Multimedia Appendix 4) include ease of use, usefulness, positive attitude, and potential to minimize physician visits.

The interviewees perceived the overall Smart-HT trial to be easy to follow and self-explanatory. For example, many described their use of nasal swabs to self-test for influenza as an *easy process.* Participants also stressed how the test kit packaging with vials to hold to swabs made the whole Smart-HT trial engagement process simpler. Interviewees also commented on the mobile app's interface that guided them with the instructions required to complete the trial as not complicated and that it served the intended purpose.

In addition, *the perceived usefulness* of the Smart-HT and the value of their participation in advancing research seemed to have encouraged interviewees' Smart-HT trial engagement. Interviewees perceived their trial engagement with flu@home as a *useful* exercise contributing to public health management in enhancing the Smart-HT functionalities for large-scale commercial use. Participants also indicated that flu@home was *useful* in reducing the spread of disease in the future.

Overall, the interviewees cited a *positive attitude* toward home-based testing and their trial engagement for the Smart-HT (*positive trial attitude*). Many expressed optimism toward Smart-HT and found it to be a promising and innovative technology with potential uses. Excitement at being offered to be a part of the Smart-HT trial was also a common theme among the interviews. Several interviewees reported their interest in engaging in a trial for a promising Smart-HT that could minimize traditional physician's visits for illness diagnosis. In addition, most interviewees explained how getting a physician's appointment when most needed could be a hassle that they would readily forego for a Smart-HT.

Situational Enablers

Situational enablers referred to external events or environmental features that enabled trial engagement for the Smart-HT. Our analysis revealed clinical affiliation, personal advice, promotion strategies, financial incentives, and insurance status as situational enablers for Smart-HT trial engagement. We present a representative quote for each of the situational enablers identified in this study (Table S3 in Multimedia Appendix 4).

Clinical affiliations of the health care entities associated with the trial prominently influenced interviewees' willingness to engage in the Smart-HT trial. Several interviewees explained that the reputable institutions associated with flu@home's development and the trial execution enhanced their perceived trust in the trial, which resulted in their trial engagement.

Furthermore, interviewees expressed their propensity to trust *personal advice* or recommendations from friends and family members to participate in these Smart-HT trials. A few interviewees shared details about who recommended them to the Smart-HT study. The list of personal recommenders identified included friends, spouses, and family members.

Considering the voluntary nature of trial participation, we identified the salience of effective *promotion and recruitment strategies* in enabling individuals' engagement in the Smart-HT trial. Interviewees detailed how they found out about the Smart-HT study on the web and described how web-based advertising could be a helpful recruitment mode. In addition, the interviewees offered meaningful insights into preferred future web-based recruitment and promotion strategies. From our analysis, social media emerged as a popular recruitment pathway to achieve the desired trial participation. The data also included references to a web-based forum or an email list where people could volunteer for these trials.

The *financial incentive* that was offered in the form of a gift card provided by the research study at completion notably incentivized and compensated the participants to engage in the Smart-HT trial. An interviewee also emphasized how *lack of insurance* coupled with symptoms and incentives encouraged Smart-HT trial engagement.

Discussion

Overview

In this qualitative study, we explored the enablers of trial engagement with a home-based diagnostic supported by a CHT, referred to as Smart-HT. We identified dispositional and situational enablers of Smart-HT trial engagement that can inform future Smart-HT trial design and execution. In the following sections, we discuss these enablers, along with recommendations for trial organizers and researchers.

In reviewing the following discussion of enablers, it is worth considering that many CHTs in various health care contexts involve a software component and are rooted in a traditional health care service offering. Additional forms of CHTs may elicit similar trial engagement responses as those who responded to the Smart-HT context. Therefore, it is plausible that the identified Smart-HT dispositional and situational enablers we discuss may apply to a majority of CHT trials, in general, and merit testing in other CHT contexts.

Dispositional Enablers

Personal Enablers

With clinical trials facing challenges in the form of low retention and high dropout rates of trial participants, our results point to the need for increased attention on personal enablers in future Smart-HT trial recruitment efforts. In alignment with prior studies on digital health technologies [67,68], individuals with digital health literacy (eg, prior experience with digital health tools) generally tended to have a positive experience with Smart-HT. In our study, individuals generally reported to be digitally health literate and did not encounter any major issues during the Smart-HT trial. Experience with accessing health care on the web seemed to have enabled participants to engage with Smart-HT in the trial. Digital health literacy of potential trial participants could inform their Smart-HT trial engagement behaviors. Therefore, as part of the participant recruitment efforts, trial organizers may want to develop screening criteria to recruit individuals with proficient digital health literacy among the target appropriate populations. To reach those with no relevant digital health tool experience, we suggest that the trial organizers offer appropriate training and general promotion concerning Smart-HT use to enhance Smart-HT trial engagement.

Smart-HT trial engagement was also enabled by the trial participants' willingness to play a role in advancing medical research. These results are in line with a survey conducted on public perceptions in the United States toward clinical trials in which 86% of the sample noted that they would participate in clinical trials for advancing scientific research [69]. In the future, researchers should make use of appropriate advertising channels to communicate the potential value of the Smart-HT trial in advancing medical research to prospective trial participants. An increase in the perceived value of the trial might further motivate prospective participants' Smart-HT trial engagement intentions.

Personal attributes, such as inherent innovativeness, curiosity, and altruistic notions, enabled participants' Smart-HT trial engagement. These findings offer practical implications for Smart-HT trial designers and trial organizing entities, as they inform successful trial participant recruitment in future trials. For instance, researchers in future trials could extend a short survey as part of the screening criteria to measure personal innovativeness [70] and curiosity [71] as part of the Smart-HT trial recruitment processes. Potential trial participants who are relatively more innovative and curious about the Smart-HT (in the trial) could then be screened, thereby enhancing the Smart-HT trial engagement and success. We also propound that future trial designers could highlight their choice of individuals who explore and are curious about new technologies in their recruitment campaigns. For instance, clear recruitment messages such as "Do you want to try a new and transformational healthcare technology?" and "Are you interested in contributing to your community" could further encourage potential trial participants' Smart-HT trial engagement.

Perceptual Enablers (Toward Smart-HT Trial)

The successful design of the accompanying mobile app enabled the overall success of the Smart-HT trial. The mobile app instructions were perceived to be self-explanatory and easy to use and enabled the participants' Smart-HT trial engagement**.** Easily understood and navigable mobile app interfaces with clear and step-by-step instructions (which in this case included videos) are critical to ensuring that trial participants complete the trial successfully and do not withdraw from the study because of frustration with the test procedures or technology. This finding further strengthens this study's contribution as it points to the potential of Smart-HT. Namely, the addition of a technology component to a trial or even an existing home diagnostic test to comprise a Smart-HT has the potential to alleviate some of the concerns of errors or process misunderstanding surrounding home diagnostic self-testing in the absence of clinical supervision. The addition of technology to facilitate procedural success can further inform future trials. Future researchers could examine whether trials involving tests and home-based preparation procedures (eg, blood test) combined with the easy-to-use app would have better trial engagement outcomes than trials involving diagnostic tests accompanied by no apps or complicated apps.

Our findings suggest a few considerations regarding the specific design of apps to support trial procedures. First, trial mobile app developers could ensure that the trial's aims and objectives are adequately communicated through the app pages to enhance the Smart-HT trial engagement. Second, the app should contain a frequently asked questions section that addresses the common concerns or risks surrounding the trial. Third, the app should provide an option to call a research coordinator to answer any questions that may arise during the trial. Fourth, interactive videos demonstrating the step-by-step study procedures (eg, nasal swab process) may stimulate and inspire potential trial participants to engage. All these design elements may reduce trial drop-off.

In line with our findings, prior research on the trialability of new technologies has shown that a positive attitude [72,73] toward the technology can ease concerns surrounding a trial of relatively new technology like Smart-HT. Removing this concern eliminates a barrier to motivating individuals to engage with Smart-HT. Furthermore, Smart-HT trial engagement was perceived to contribute to public health management in our study. These findings have significant implications for trial organizers and mobile app developers who could highlight how a specific trial could advance and benefit public health in addition to advancing medical research. Consequently, educating potential trial participants about positive trial consequences may manifest in a positive attitude toward the Smart-HT and the overall trial, encouraging the Smart-HT trial engagement.

Another interesting perceptional enabler theme that emerged from our analysis was the perceived potential of the Smart-HT

to minimize a physician's visit. Smart-HT offers a fundamental change from traditional health care encounters for diagnostic testing. The potential role of Smart-HT in reducing the need to seek face-to-face appointments encouraged the interviewees to engage in the trial. Therefore, Smart-HT trial engagement may be enhanced if potential trial participants are convinced that the Smart-HT is transformational and offers solutions to the existing problems they face in the health care system (eg, access to testing and getting physician's appointments). Researchers in future trials should inform potential trial participants about the possible innovations and improvements that may be solved by the Smart-HT in trial.

Situational Enablers

Awareness of the trial's clinical affiliations enabled Smart-HT trial engagement in our study. The brand reputation associated with these health care and research institutions encouraged potential participants to trust the trial as legitimate and, thus, influenced their trial engagement. These findings are in line with a survey conducted on the perceptions toward clinical trials [69]. For example, 38% of the nationwide sample cited a lack of trust in trials as one of the primary reasons for not considering clinical trial participation. In addition, 91% of the same sample said that the competence and reputation of the trial conducting institutions play an essential role in their decision to participate in a trial. These findings could potentially inform future trial designers and mobile app developers. For example, future recruitment strategies could include detailed descriptions of the trial organizing institutions' previous research breakthroughs and research quality (in terms of patents or publications) that could persuade potential trial participants to perceive the trial as legitimate and to engage in it.

Most of the trial participants heard about the Smart-HT trial in our study from social media. Our analysis affirmed that web-based advertising and recruitment strategies could be effective modes of potential participant recruitment. Web-based medical forums where potential trial participants could volunteer to engage in an advertised trial could offer a potential future mode of participant recruitment for Smart-HT trials. Sending Smart-HT trial invitations using emails to a list of potential trial participants could also enable the Smart-HT trial engagement. Therefore, Smart-HT trial organizers should dedicate significant efforts toward web-based recruitment using available channels such as web-based medical forums, email lists, health care blogs, and social media [74,75].

Personal recommendations to engage in the Smart-HT trial (in the form of referrals from friends and family) and the financial incentive offered for participating in the trial seemed to have motivated some interviewees' engagement. In the future, mobile app developers could add a functionality, allowing one to invite others to sign up for the trial through the app itself. Many interviewees affirmed financial incentives as a critical factor that influenced their Smart-HT trial engagement. Although considering the perceived criticality of financial incentives, there may be value in offering additional incentives for referring other potential participants to the Smart-HT trial. Referral promotions could solve the issues of inadequate participant enrollment and retention rates.

Some participants attributed their trial engagement for the Smart-HT to their lack of health insurance. This study highlights the opportunity to engage participants in Smart-HT clinical trials to improve their access to needed health care. We also identified practical issues for trial designers in terms of trial advertising and recruitment efforts. One of the interviewees explained how the absence of trial participation costs encouraged her to enroll in the Smart-HT trial. Therefore, emphasizing that the Smart-HT trial is free and does not require health insurance in the advertisements and recruitment messages may increase engagement. These techniques could draw the attention of potential trial participants who may assume that they are ineligible for trial as they lack health insurance. Simultaneously, we caution the Smart-HT trial organizers that it would be unethical to coerce potential trial participants to sign up for a clinical study they would otherwise be reluctant to do.

Overall, our evidence-based research framework (Figure 5) highlights the cruciality of the trial participants' personalities, beliefs, and perceptions as drivers of their Smart-HT trial engagement. Our analysis and the resulting framework may help Smart-HT researchers and trial organizers gain valuable insight into the phenomenon of trial engagement. We also encourage future research studies that extend our framework by examining the relative significance of the enablers presented using quantitative research methods.

Limitations

We focused on one example of CHT, Smart-HT, an instantiation of flu@home to identify the dispositional and situational enablers of trial engagement in our study. Although we feel that our results may carry to other CHT trial situations, we did not test other CHT contexts. Therefore, this choice may limit the generalizability of the study's findings. Specifically, trial participants for CHTs that differ from Smart-HT may elicit a different or an extended set of their trial engagement enablers. Although we expect most of the identified dispositional enablers to influence trial engagement for other CHTs and Smart-HTs, there could be additional situational enablers relevant to the functionality and context of the specific technology used. We encourage future research to investigate and enrich the framework in other contexts such as mobile health apps and Smart-HT for other health conditions and clinical trials. Our study focuses on offering an evidence-based Smart-HT trial engagement framework for Smart-HT that requires FDA approval. Future research is required to investigate the enablers for trial engagement for other health and wellness technologies that may not formally require FDA approval or undergo approval through different FDA routes for devices. The need for approval and situation influences resulting in social desirability response bias might have minimally influenced the qualitative study's findings [76]. We encourage future studies to empirically test and extend our evidence-based Smart-HT trial engagement framework using quantitative research methods such as surveys.

Conclusions

Premarket clinical trials play a crucial role in evaluating and assessing medical devices and technologies for effectiveness and safety before their approval and release to the general population. However, inadequate trial engagement remains an

issue that leads to failure in CHT trials. In the field of CHT, there is a dearth of research exploring the factors leading to CHT trial engagement. This exploratory qualitative study reports enablers of trial engagement for Smart-HT, a promising form of CHT, using an instantiation to diagnose influenza, namely, flu@home.

Our study adds to the understanding of people's perceptions toward Smart-HT trials by delineating an organized series of enablers that can impact the decision to participate in a Smart-HT trial. Clarification on these dispositional and situational enablers contributes to behavioral research associated with Smart-HT trials and can inform future Smart-HT trial design and engagement strategies. In addition, our study offers practical implications for trial organizers in the form of trial promotional and recruitment efforts. An important example is developing screening criteria to recruit innovative, curious, and altruistic individuals with proficient digital health literacy among the target population. Our study also confirms the benefit of creating relevant promotion materials that showcase Smart-HT's usefulness, ease of use, benefits, trial organizer's reputation and clinical affiliations, the trial's role in advancing medical research, and finally, Smart-HT's role in solving existing health care issues. Understanding and accounting for a series of internal and external enablers when targeting potential trial participants can address the issue of inadequate Smart-HT trial engagement, thereby leading to effective evaluation of health technologies and timely release of them to the general population.

Acknowledgments

Efforts that helped shape this paper were partly supported by the Seattle Flu Study, funded by Gates Ventures. The funder was not involved in the design of the study, nor did they have any ownership over the management and conduct of the study, the data, or the rights to publish. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Seattle Flu Study investigators. We acknowledge John Tamerius and Quidel Corporation for donating the testing materials for the flu@home pilot. The authors would like to acknowledge Audere for developing the flu@home mobile app and Shawna Cooper for providing app content for this publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The flu@home mobile app.
[DOCX File , 13 KB - jmir_v23i9e26869_app1.docx]

Multimedia Appendix 2
Trial recruitment videos—demo of flu@home.
[MP4 File (MP4 Video), 14643 KB - jmir_v23i9e26869_app2.mp4]

Multimedia Appendix 3
Trial recruitment videos—overview of flu@home research study.
[MP4 File (MP4 Video), 9200 KB - jmir_v23i9e26869_app3.mp4]

Multimedia Appendix 4
Evidence trace tables.
[DOCX File , 17 KB - jmir_v23i9e26869_app4.docx]

References

1. Mobile medical applications: guidance for industry and Food and Drug Administration staff. US Food & Drug Administration. 2013. URL: https://www.fda.gov/media/80958/download [accessed 2014-09-08]
2. Medina-Marino A, de Vos L, Bezuidenhout D, Denkinger CM, Schumacher SG, Shin SS, et al. "I got tested at home, the help came to me": acceptability and feasibility of home-based TB testing of household contacts using portable molecular diagnostics in South Africa. Trop Med Int Health 2021 Mar 04;26(3):343-354 [FREE Full text] [doi: 10.1111/tmi.13533] [Medline: 33289194]
3. Bissonnette L, Bergeron M. Diagnosing infections--current and anticipated technologies for point-of-care diagnostics and home-based testing. Clin Microbiol Infect 2010 Aug;16(8):1044-1053 [FREE Full text] [doi: 10.1111/j.1469-0691.2010.03282.x] [Medline: 20670286]
4. Atchison C, Pristerà P, Cooper E, Papageorgiou V, Redd R, Piggin M, et al. Usability and acceptability of home-based self-testing for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) antibodies for population surveillance. Clin Infect Dis 2021 May 04;72(9):e384-e393 [FREE Full text] [doi: 10.1093/cid/ciaa1178] [Medline: 32785665]

5. Sewitch MJ, Fallone CA, Ghali P, Lee GE. What patients want in a smartphone app that supports colonoscopy preparation: qualitative study to inform a user-centered smartphone app. JMIR Mhealth Uhealth 2019 Jul 02;7(7):e12242 [FREE Full text] [doi: 10.2196/12242] [Medline: 31125310]
6. Juarascio AS, Goldstein SP, Manasse SM, Forman EM, Butryn ML. Perceptions of the feasibility and acceptability of a smartphone application for the treatment of binge eating disorders: Qualitative feedback from a user population and clinicians. Int J Med Inform 2015 Oct;84(10):808-816 [FREE Full text] [doi: 10.1016/j.ijmedinf.2015.06.004] [Medline: 26113461]
7. Cuggia M, Besana P, Glasspool D. Comparing semi-automatic systems for recruitment of patients to clinical trials. Int J Med Inform 2011 Jun;80(6):371-388. [doi: 10.1016/j.ijmedinf.2011.02.003] [Medline: 21459664]
8. Ammenwerth E, Gräber S, Herrmann G, Bürkle T, König J. Evaluation of health information systems—problems and challenges. Int J Med Inform 2003 Sep;71(2-3):125-135. [doi: 10.1016/s1386-5056(03)00131-x]
9. Costenbader KH, Brome D, Blanch D, Gall V, Karlson E, Liang MH. Factors determining participation in prevention trials among systemic lupus erythematosus patients: a qualitative study. Arthritis Rheum 2007 Feb 15;57(1):49-55 [FREE Full text] [doi: 10.1002/art.22480] [Medline: 17266094]
10. Tsungmey T, Kim JP, Dunn LB, Ryan K, Lane-McKinley K, Roberts LW. Negative association of perceived risk and willingness to participate in innovative psychiatric research protocols. J Psychiatr Res 2020 Mar;122:9-16 [FREE Full text] [doi: 10.1016/j.jpsychires.2019.12.010] [Medline: 31891880]
11. Ding EL, Powe NR, Manson JE, Sherber NS, Braunstein JB. Sex differences in perceived risks, distrust, and willingness to participate in clinical trials: a randomized study of cardiovascular prevention trials. Arch Intern Med 2007 May 14;167(9):905-912. [doi: 10.1001/archinte.167.9.905] [Medline: 17502531]
12. Nuttall A. Considerations for improving patient recruitment into clinical trials. RDP Clinical Outsourcing. 2012. URL: http://vertassets.blob.core.windows.net/download/64c39d7e/64c39d7e-c643-457b-aec2-9ff7b65b3ad2/rdprecruitmentwhitepaper.pdf [accessed 2021-08-24]
13. Kearney A, Daykin A, Shaw AR, Lane AJ, Blazeby JM, Clarke M, et al. Identifying research priorities for effective retention strategies in clinical trials. Trials 2017 Aug 31;18(1):406 [FREE Full text] [doi: 10.1186/s13063-017-2132-z] [Medline: 28859674]
14. Salman RA, Beller E, Kagan J, Hemminki E, Phillips RS, Savulescu J, et al. Increasing value and reducing waste in biomedical research regulation and management. Lancet 2014 Jan;383(9912):176-185. [doi: 10.1016/s0140-6736(13)62297-7]
15. Skea ZC, Newlands R, Gillies K. Exploring non-retention in clinical trials: a meta-ethnographic synthesis of studies reporting participant reasons for drop out. BMJ Open 2019 Jun 03;9(6):e021959 [FREE Full text] [doi: 10.1136/bmjopen-2018-021959] [Medline: 31164359]
16. Gupta A, Calfas KJ, Marshall SJ, Robinson TN, Rock CL, Huang JS, et al. Clinical trial management of participant recruitment, enrollment, engagement, and retention in the SMART study using a Marketing and Information Technology (MARKIT) model. Contemp Clin Trials 2015 May;42:185-195 [FREE Full text] [doi: 10.1016/j.cct.2015.04.002] [Medline: 25866383]
17. Alegría M, Carson N, Flores M, Li X, Shi P, Lessios AS, et al. Activation, self-management, engagement, and retention in behavioral health care: a randomized clinical trial of the DECIDE intervention. JAMA Psychiatry 2014 May 01;71(5):557-565 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4519] [Medline: 24647680]
18. Palac HL, Alam N, Kaiser SM, Ciolino JD, Lattie EG, Mohr DC. A practical do-it-yourself recruitment framework for concurrent eHealth clinical trials: simple architecture (part 1). J Med Internet Res 2018 Nov 01;20(11):e11049 [FREE Full text] [doi: 10.2196/11049] [Medline: 30389650]
19. Kemmler G, Hummer M, Widschwendter C, Fleischhacker WW. Dropout rates in placebo-controlled and active-control clinical trials of antipsychotic drugs: a meta-analysis. Arch Gen Psychiatry 2005 Dec 01;62(12):1305-1312. [doi: 10.1001/archpsyc.62.12.1305] [Medline: 16330718]
20. Thompson IM, Tangen CM, Paradelo J, Lucia MS, Miller G, Troyer D, et al. Adjuvant radiotherapy for pathologically advanced prostate cancer: a randomized clinical trial. J Am Med Assoc 2006 Nov 15;296(19):2329-2335. [doi: 10.1001/jama.296.19.2329] [Medline: 17105795]
21. Massarelli E, William W, Johnson F, Kies M, Ferrarotto R, Guo M, et al. Combining immune checkpoint blockade and tumor-specific vaccine for patients with incurable human papillomavirus 16-related cancer: a phase 2 clinical trial. JAMA Oncol 2019 Jan 01;5(1):67-73 [FREE Full text] [doi: 10.1001/jamaoncol.2018.4051] [Medline: 30267032]
22. Agboola SO, Ju W, Elfiky A, Kvedar JC, Jethwani K. The effect of technology-based interventions on pain, depression, and quality of life in patients with cancer: a systematic review of randomized controlled trials. J Med Internet Res 2015 Mar 13;17(3):e65 [FREE Full text] [doi: 10.2196/jmir.4009] [Medline: 25793945]
23. Kasner SE, Del Giudice A, Rosenberg S, Sheen M, Luciano JM, Cucchiara BL, et al. Who will participate in acute stroke trials? Neurology 2009 May 11;72(19):1682-1688. [doi: 10.1212/wnl.0b013e3181a55fbe]
24. Martin SS, Ou F, Newby LK, Sutton V, Adams P, Felker GM, et al. Patient- and trial-specific barriers to participation in cardiovascular randomized clinical trials. J Am Coll Cardiol 2013 Feb 19;61(7):762-769 [FREE Full text] [doi: 10.1016/j.jacc.2012.10.046] [Medline: 23410547]

25. Kinder BW, Sherman A, Young L, Hagaman J, Oprescu N, Byrnes S, et al. Predictors for clinical trial participation in the rare lung disease lymphangioleiomyomatosis. Respir Med 2010 Apr;104(4):578-583 [FREE Full text] [doi: 10.1016/j.rmed.2009.09.023] [Medline: 19962873]
26. Or CK, Tao D. Does the use of consumer health information technology improve outcomes in the patient self-management of diabetes? A meta-analysis and narrative review of randomized controlled trials. Int J Med Inform 2014 May;83(5):320-329. [doi: 10.1016/j.ijmedinf.2014.01.009] [Medline: 24534118]
27. Liu C, Zhu Q, Holroyd KA, Seng EK. Status and trends of mobile-health applications for iOS devices: a developer's perspective. J Sys Soft 2011 Nov;84(11):2022-2033. [doi: 10.1016/j.jss.2011.06.049]
28. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013 Jun 14;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]
29. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
30. Cenfetelli R. Inhibitors and enablers as dual factor concepts in technology usage. J Assoc Inf Syst 2004 Dec;5(11):472-492. [doi: 10.17705/1jais.00059]
31. Lattie EG, Kaiser SM, Alam N, Tomasino KN, Sargent E, Rubanovich CK, et al. A practical do-it-yourself recruitment framework for concurrent eHealth clinical trials: identification of efficient and cost-effective methods for decision making (part 2). J Med Internet Res 2018 Nov 29;20(11):e11050 [FREE Full text] [doi: 10.2196/11050] [Medline: 30497997]
32. Lindgren T, Hooper J, Fukuoka Y. Perceptions and experiences of women participating in a digital technology-based physical activity intervention (the mPED trial): qualitative study. JMIR Public Health Surveill 2019 Dec 20;5(4):e13570 [FREE Full text] [doi: 10.2196/13570] [Medline: 31859677]
33. Cohen S, Waks Z, Elm JJ, Gordon MF, Grachev ID, Navon-Perry L, et al. Characterizing patient compliance over six months in remote digital trials of Parkinson's and Huntington disease. BMC Med Inform Decis Mak 2018 Dec 20;18(1):138 [FREE Full text] [doi: 10.1186/s12911-018-0714-7] [Medline: 30572891]
34. Mauk MG. Calling in the test: Smartphone-based urinary sepsis diagnostics. EBioMedicine 2018 Nov;37:11-12 [FREE Full text] [doi: 10.1016/j.ebiom.2018.10.047] [Medline: 30482724]
35. Snijders T, Bosker R. Smartphone Based Medical Diagnostics. Cambridge: Academic Press Inc; 1999:1-252.
36. Song J, Pandian V, Mauk MG, Bau HH, Cherry S, Tisi LC, et al. Smartphone-based mobile detection platform for molecular diagnostics and spatiotemporal disease mapping. Anal Chem 2018 Apr 03;90(7):4823-4831 [FREE Full text] [doi: 10.1021/acs.analchem.8b00283] [Medline: 29542319]
37. Hernández-Neuta I, Neumann F, Brightmeyer J, Ba Tis T, Madaboosi N, Wei Q, et al. Smartphone-based clinical diagnostics: towards democratization of evidence-based health care. J Intern Med 2019 Jan 12;285(1):19-39 [FREE Full text] [doi: 10.1111/joim.12820] [Medline: 30079527]
38. At home diagnostics market by test type, by form type, by sample type, by usage, by distribution channel and by region: industry analysis, market share, revenue opportunity, competition and forecast 2020 to 2027. FutureWise. URL: https://www.futurewiseresearch.com/healthcare-market-research/At-Home-Diagnostics/181 [accessed 2021-08-24]
39. Betts D, Korenda L. Inside the patient journey: Three key touch points for consumer engagement strategies: Findings from the Deloitte 2018 Health Care Consumer Survey. Deloitte Insights. 2018. URL: https://www2.deloitte.com/us/en/insights/industry/health-care/patient-engagement-health-care-consumer-survey.html [accessed 2021-09-01]
40. World Health Organization. Global influenza strategy 2019-2030. Geneva: World Health Organization; 2019:31.
41. Molinari NM, Ortega-Sanchez IR, Messonnier ML, Thompson WW, Wortley PM, Weintraub E, et al. The annual impact of seasonal influenza in the US: measuring disease burden and costs. Vaccine 2007 Jun 28;25(27):5086-5096. [doi: 10.1016/j.vaccine.2007.03.046] [Medline: 17544181]
42. Reh G. At-home Diagnostic Tools: A New Way to Connect With Consumers. The Wall Street Journal. 2018. URL: https://deloitte.wsj.com/articles/at-home-diagnostic-tools-a-new-way-to-connect-with-consumers-1537902128 [accessed 2021-08-24]
43. Bunniss S, Kelly D. Research paradigms in medical education research. Med Educ 2010 Apr;44(4):358-366. [doi: 10.1111/j.1365-2923.2009.03611.x] [Medline: 20444071]
44. Creswell J, Poth C. Qualitative inquiry and research design: Choosing among five approaches. Thousand Oaks, CA: SAGE Publications Ltd; 2016.
45. Glaser B, Strauss A. The Discovery of Grounded Theory. Chicago, IL: Aldine; 1967.
46. Morse J. "Data were saturated . . . ". Qual Health Res 2015 May;25(5):587-588. [doi: 10.1177/1049732315576699] [Medline: 25829508]
47. Sim J, Saunders B, Waterfield J, Kingstone T. Can sample size in qualitative research be determined a priori? Int J Soc Res Methodol 2018 Mar 27;21(5):619-634. [doi: 10.1080/13645579.2018.1454643]
48. Mason M. Sample size and saturation in PhD studies using qualitative interviews. Forum Qual Soc Res 2010;11(3). [doi: 10.17169/fqs-11.3.1428]
49. Baker S, Edwards R. How many qualitative interviews is enough. National Centre for Research Methods Review Paper. 2012. URL: http://eprints.ncrm.ac.uk/2273/4/how_many_interviews.pdf [accessed 2021-08-24]

50. Lincoln YS, Guba EG. But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. New Dir Eval 1986(30):73-84. [doi: 10.1002/ev.1427]
51. Onwuegbuzie AJ, Leech NL. Validity and qualitative research: an oxymoron? Qual Quant 2006 May 25;41(2):233-249. [doi: 10.1007/s11135-006-9000-3]
52. Choy MA, Sturgiss E, Goodyear-Smith F, Smith GJ. Digital health tools and patients with drug use disorders: qualitative patient experience study of the electronic case-finding and help assessment tool (eCHAT). J Med Internet Res 2020 Sep 14;22(9):e19256 [FREE Full text] [doi: 10.2196/19256] [Medline: 32924959]
53. Archibald MM, Ambagtsheer RC, Casey MG, Lawless M. Using zoom videoconferencing for qualitative data collection: perceptions and experiences of researchers and participants. Int J Qual Methods 2019 Sep 11;18:160940691987459. [doi: 10.1177/1609406919874596]
54. Sundler AJ, Lindberg E, Nilsson C, Palmér L. Qualitative thematic analysis based on descriptive phenomenology. Nurs Open 2019 Jul 07;6(3):733-739 [FREE Full text] [doi: 10.1002/nop2.275] [Medline: 31367394]
55. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods 2017 Oct 02;16(1):160940691773384. [doi: 10.1177/1609406917733847]
56. Forero R, Nahidi S, De Costa J, Mohsin M, Fitzgerald G, Gibson N, et al. Application of four-dimension criteria to assess rigour of qualitative research in emergency medicine. BMC Health Serv Res 2018 Feb 17;18(1):120 [FREE Full text] [doi: 10.1186/s12913-018-2915-2] [Medline: 29454350]
57. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J Qual Methods 2016 Nov 29;5(1):80-92. [doi: 10.1177/160940690600500107]
58. Weiner B. An attributional theory of achievement motivation and emotion. Psychol Rev 1985;92(4):548-573. [doi: 10.1037/0033-295x.92.4.548]
59. Weiner B. An Attributional Theory of Achievement Motivation and Emotion. In: An Attributional Theory of Motivation and Emotion. New York: Springer; 1986.
60. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
61. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
62. Potter WJ, Levine-Donnerstein D. Rethinking validity and reliability in content analysis. J Appl Commun Res 2009 May 21;27(3):258-284. [doi: 10.1080/00909889909365539]
63. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov 01;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
64. Glaser BG. The constant comparative method of qualitative analysis. Social Problems 1965 Apr;12(4):436-445. [doi: 10.1525/sp.1965.12.4.03a00070]
65. Townsend A, Leese J, Adam P, McDonald M, Li LC, Kerr S, et al. Ehealth, participatory medicine, and ethical care: a focus group study of patients' and health care providers' use of health-related internet information. J Med Internet Res 2015 Jun 22;17(6):e155 [FREE Full text] [doi: 10.2196/jmir.3792] [Medline: 26099267]
66. Ataro G. Methods, methodological challenges and lesson learned from phenomenological study about OSCE experience: Overview of paradigm-driven qualitative approach in medical education. Ann Med Surg (Lond) 2020 Jan;49:19-23 [FREE Full text] [doi: 10.1016/j.amsu.2019.11.013] [Medline: 31871678]
67. Pai RR, Alathur S. Assessing mobile health applications with Twitter analytics. Int J Med Inform 2018 May;113:72-84. [doi: 10.1016/j.ijmedinf.2018.02.016] [Medline: 29602436]
68. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
69. Pubic perception of clinical trials. Research!America. 2017. URL: https://www.researchamerica.org/polls-and-publications/polls/public-opinion-polls [accessed 2021-08-24]
70. Rogers E. Diffusion of Innovations, 5th Edition. New York: Free Press; 2003:1-576.
71. Kashdan TB, Rose P, Fincham FD. Curiosity and exploration: facilitating positive subjective experiences and personal growth opportunities. J Pers Assess 2004 Jun;82(3):291-305. [doi: 10.1207/s15327752jpa8203_05] [Medline: 15151805]
72. Hill RJ, Fishbein M, Ajzen I. Belief, attitude, intention and behavior: an introduction to theory and research. Contemp Sociol 1977 Mar;6(2):244. [doi: 10.2307/2065853]
73. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
74. Gelinas L, Pierce R, Winkler S, Cohen IG, Lynch HF, Bierer BE. Using social media as a research recruitment tool: ethical issues and recommendations. Am J Bioeth 2017 Mar 16;17(3):3-14 [FREE Full text] [doi: 10.1080/15265161.2016.1276644] [Medline: 28207365]
75. Mendelson C. Recruiting participants for research from online communities. Comput Inform Nurs 2007;25(6):317-323. [doi: 10.1097/01.NCN.0000299653.13777.51] [Medline: 18000427]

76. Collins M, Shattell M, Thomas SP. Problematic interviewee behaviors in qualitative research. West J Nurs Res 2005 Mar 01;27(2):188-99; discussion 200. [doi: 10.1177/0193945904268068] [Medline: 15695576]

Abbreviations

CHT: consumer health care technologies
FDA: US Food and Drug Administration
Smart-HT: smartphone-supported home tests

Edited by G Eysenbach; submitted 31.12.20; peer-reviewed by Y Man, Y Cai; comments to author 27.01.21; revised version received 13.06.21; accepted 27.07.21; published 14.09.21.

Please cite as:
Dharanikota S, LeRouge CM, Lyon V, Durneva P, Thompson M
Identifying Enablers of Participant Engagement in Clinical Trials of Consumer Health Technologies: Qualitative Study of Influenza Home Testing
J Med Internet Res 2021;23(9):e26869
URL: https://www.jmir.org/2021/9/e26869
doi:10.2196/26869
PMID:34519664

Original Paper

A Chest Patch for Continuous Vital Sign Monitoring: Clinical Validation Study During Movement and Controlled Hypoxia

Carlos Morgado Areia[1,2], MSc, PT; Mauro Santos[2,3], DPhil; Sarah Vollam[1,2], MSc, RN; Marco Pimentel[2,3], DPhil; Louise Young[1,2], RN; Cristian Roman[2,3], PhD; Jody Ede[1,2], MSc, RN; Philippa Piper[1], BSc; Elizabeth King[4], BSc; Owen Gustafson[4], MSc; Mirae Harford[1,2], MD; Akshay Shah[5], FRCA; Lionel Tarassenko[2,3], MA, DPhil; Peter Watkinson[1,2,6], MD

[1]Critical Care Research Group, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom

[2]National Institute for Health Research, Biomedical Research Centre, Oxford, United Kingdom

[3]Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, United Kingdom

[4]Therapies Clinical Service Unit, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom

[5]Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom

[6]Kadoorie Centre for Critical Care Research and Education, Oxford University Hospitals NHS Trust, Oxford, United Kingdom

Corresponding Author:
Carlos Morgado Areia, MSc, PT
Critical Care Research Group, Nuffield Department of Clinical Neurosciences
University of Oxford
Kadoorie Research Centre
John Radcliffe Hospital, Headley Way
Oxford, OX3 9DU
United Kingdom
Phone: 44 1865 231440
Email: carlos.morgadoareia@ndcn.ox.ac.uk

Abstract

Background: The standard of care in general wards includes periodic manual measurements, with the data entered into track-and-trigger charts, either on paper or electronically. Wearable devices may support health care staff, improve patient safety, and promote early deterioration detection in the interval between periodic measurements. However, regulatory standards for ambulatory cardiac monitors estimating heart rate (HR) and respiratory rate (RR) do not specify performance criteria during patient movement or clinical conditions in which the patient's oxygen saturation varies. Therefore, further validation is required before clinical implementation and deployment of any wearable system that provides continuous vital sign measurements.

Objective: The objective of this study is to determine the agreement between a chest-worn patch (VitalPatch) and a gold standard reference device for HR and RR measurements during movement and gradual desaturation (modeling a hypoxic episode) in a controlled environment.

Methods: After the VitalPatch and gold standard devices (Philips MX450) were applied, participants performed different movements in seven consecutive stages: at rest, sit-to-stand, tapping, rubbing, drinking, turning pages, and using a tablet. Hypoxia was then induced, and the participants' oxygen saturation gradually reduced to 80% in a controlled environment. The primary outcome measure was accuracy, defined as the mean absolute error (MAE) of the VitalPatch estimates when compared with HR and RR gold standards (3-lead electrocardiography and capnography, respectively). We defined these as clinically acceptable if the rates were within 5 beats per minute for HR and 3 respirations per minute (rpm) for RR.

Results: Complete data sets were acquired for 29 participants. In the movement phase, the HR estimates were within prespecified limits for all movements. For RR, estimates were also within the acceptable range, with the exception of the sit-to-stand and turning page movements, showing an MAE of 3.05 (95% CI 2.48-3.58) rpm and 3.45 (95% CI 2.71-4.11) rpm, respectively. For the hypoxia phase, both HR and RR estimates were within limits, with an overall MAE of 0.72 (95% CI 0.66-0.78) beats per minute and 1.89 (95% CI 1.75-2.03) rpm, respectively. There were no significant differences in the accuracy of HR and RR estimations between normoxia (≥90%), mild (89.9%-85%), and severe hypoxia (<85%).

Conclusions: The VitalPatch was highly accurate throughout both the movement and hypoxia phases of the study, except for RR estimation during the two types of movements. This study demonstrated that VitalPatch can be safely tested in clinical environments to support earlier detection of cardiorespiratory deterioration.

Trial Registration: ISRCTN Registry ISRCTN61535692; https://www.isrctn.com/ISRCTN61535692

(J Med Internet Res 2021;23(9):e27547) doi:10.2196/27547

KEYWORDS

clinical validation; chest patch; vital signs; remote monitoring; wearable; heart rate; respiratory rate

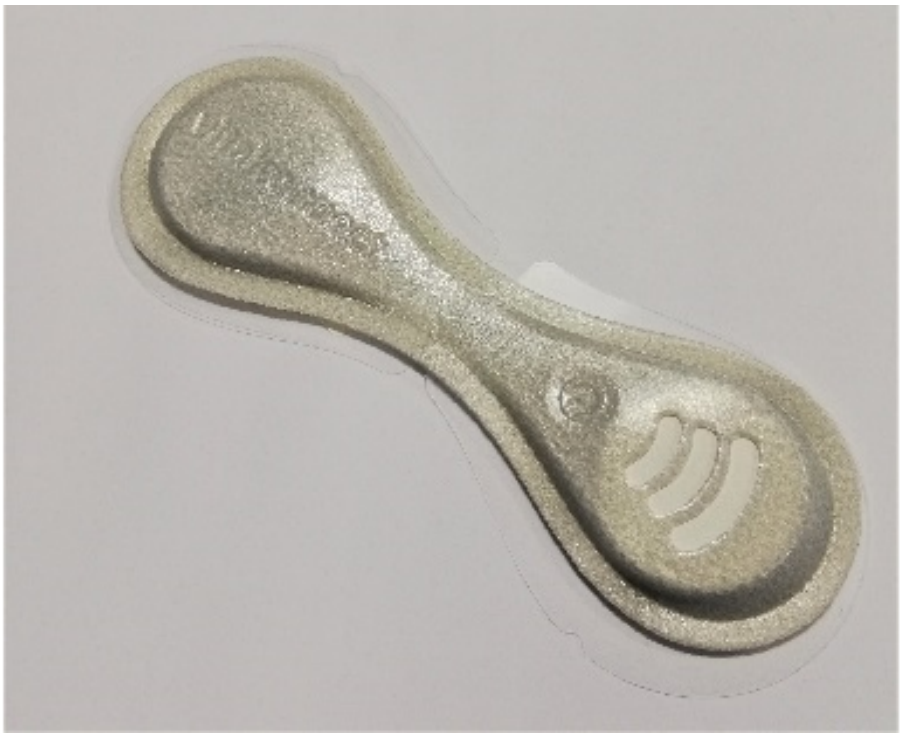
Introduction

Background

Failure to recognize and act on physiological indicators of deterioration in hospital wards is a well-recognized problem [1-3]. Current practice involves manual measurement of vital signs and the use of early warning scores. However, the acquisition of these measurements can be time consuming for health professionals in a busy ward [4]; therefore, the desired monitoring frequency of observations is often not achieved [5]. This is identified as a limitation of periodic monitoring systems, as they are dependent on the frequency of vital sign observations [6]; of more concern is that even when the ideal frequency is achieved, patients might deteriorate between observation sets [7]. Continuous vital sign monitoring using wearable devices may overcome these limitations and improve timely detection of deterioration [8-10]. However, recent observational studies and pilot randomized controlled trials of wearable monitoring devices were inconclusive, and large randomized clinical trials of ambulatory monitoring devices are required to demonstrate improved patient outcomes [1,8,11,12]. This may be because of limitations in currently available ambulatory monitoring devices; for instance, it has been demonstrated that the detection and reduction of noise and artifacts during movement can be challenging in this type of monitoring [13]. Therefore, there is a need for diagnostic accuracy studies addressing motion artifacts to support the development of reliable wearable devices [3,9,14,15].

The VitalPatch [16] is a wireless, battery-operated chest-worn wearable biosensor that records heart rate (HR), electrocardiography (ECG), HR variability, R-R interval (the time elapsed between two successive R-waves of the QRS signal on the electrocardiogram), respiratory rate (RR), body temperature, skin temperature, fall detection, activity (including step count), and posture (body position relative to gravity including fall detection) [17,18]. This device (Figure 1) was well tolerated by users because of its wearability and low impact on activities [19,20]. However, validation studies regarding the VitalPatch showed mixed results [3]; for HR, the mean bias was within acceptable ranges for all validation studies [14,18,21-23] and within acceptable limits of agreement (LoA) in 3 studies [18,21,22]. For RR, the mean bias was also acceptable; however, all studies reported LoA outside of the clinically acceptable range [14,18,21-23].

Figure 1. Picture of a VitalPatch device.



This work is a substudy of our virtual high dependency unit (vHDU) hypoxia study [24]. In the main protocol, our goal was to determine the specificity and sensitivity of currently available ambulatory pulse oximeters for the detection of hypoxemia. In this substudy, the aim is to assess the VitalPatch estimation accuracy during several simulated movements and in controlled hypoxic conditions against gold standard measurements, ECG for HR [25] and capnography for RR [26].

Objectives

The primary outcome is accuracy, defined as the mean absolute error (MAE) of the VitalPatch HR and RR estimation versus gold standard recording of HR (3-lead ECG) and RR (capnography) in healthy volunteers during movement and during induced hypoxia (down to peripheral oxygen saturation [SpO_2] of 80%). In line with the American National Standards Institute/Association for the Advancement of Medical Instrumentation (ANSI/AAMI) EC13:2002 standard [27], we defined the MAE of the patch estimates to be clinically acceptable if the rates are within 5 beats per minute (bpm) and 3 respirations per minute (rpm) [3,21,22,27].

Methods

This manuscript follows the Guidelines for Reporting Reliability and Agreement Studies [28].

Study Design

This was a prospective, observational, cross-sectional cohort study that validated the HR and RR of a single-lead ECG chest patch. The VitalPatch HR is computed via the detection of Q wave, R wave, and S wave complexes from a single-lead ECG waveform. The RR is derived from combined information from three sources, that is, an embedded algorithm uses a weighted average of two characteristics of the ECG signal: (1) Q-, R-, and S-wave amplitude modulation and (2) respiratory sinus arrhythmia (as both ECG-derived signals change during inspiration and expiration), and (3) changes in the accelerometer signal induced by chest movement during respiration [23,29,30]. We compared the HR and RR estimates with the *gold standard* HR derived from 3-lead ECG and RR determined from capnography measurements, respectively. The protocol for the full study has been published [24].

Participants

The screening process and inclusion and exclusion criteria are described in the study protocol [24]. Participants were recruited consecutively from the Oxford area, United Kingdom, between June 18 and August 8, 2019. The study sessions took place at the Cardiovascular Clinical Research Facility, John Radcliffe Hospital, Oxford, United Kingdom.

Device Placement

After the participants provided written informed consent, an arterial line was inserted into the participant's nondominant radial artery. The participants were then fitted with a VitalPatch in a modified lead-II configuration on the left midclavicular line over the intercostal space (as recommended by the vendor), a nonambulatory Philips MX 450 pulse oximeter (dominant arm), a 3-lead ECG (standard configuration) for acquisition of HR, and a nasal cannula for acquisition of RR, both connected to the Philips monitor. The participants were continuously monitored with these devices throughout the study session by trained health care professionals with appropriate resuscitation facilities.

Test Methods

Stage 1: Movement Phase

The participants were considered *at rest* after fitting all the devices (variable duration between participants). The participants then moved into a chair and were asked to complete a series of consecutive motion conditions: 20× sit-to-stand (STS), 2-minute tapping at 2 Hz, 2-minute rubbing at 2 Hz, 20 times drinking from a plastic cup, 50 times turning pages, and a set of predefined table activity tasks [24]. STS was considered a dynamic activity, whereas the remainder were considered *sedentary activities*.

Stage 2: Hypoxia Exposure Phase

Following the movement phase, participants were placed in a semirecumbent, supine position and wore a tight-fitting silicone facemask connected to a hypoxicator unit (Everest Summit Hypoxic Generator). During this phase, oxygen saturations from the *clinical standard* Philips monitor guided the titration of the hypoxicator by a senior anesthetist from the research team. Oxygen (7%) in nitrogen was used to further lower the fraction of inspired oxygen, if required [24]. The inhaled fraction of inspired oxygen was also monitored using an in-line gas analyzer. Participants were gradually desaturated to reach stable prespecified target SpO_2 levels of 95%, 90%, 87%, 85%, 83%, and 80%. These levels were selected to allow sufficient data points for different hypoxia levels. SpO_2 stability was subjective for each target window, that is, a senior anesthetist decided when a stable oxygen level was achieved.

Data Collection

For each participant, the following information was recorded in the case report form: age, sex, height, weight, skin type (Fitzpatrick scale [31]), baseline HR (Philips MX 450 3-lead ECG), and arterial oxygen saturation. Data from all participants were identified using a study number.

All data collection devices (wearables, tablets, laptops, and physiological monitors), and respective software, were synchronized to the same time server at the start of each study session day. Continuous HR and RR data (sampled at 0.25 Hz) from the VitalPatch were sent via Bluetooth Low Energy to an Android tablet–based data capture system, developed in-house. Once the VitalPatch is connected to the tablet, it updates its internal clock and timestamps each sample internally. The Phillips MX 450 3-lead ECG data were collected in real time on a laptop using the ixTrend version 2.1 software (ixcellence GmbH) [32]. For each step of the *movement phase* and prespecified oxygen saturation level of the *hypoxia phase*, the corresponding start and stop time stamps were recorded both in the in-house data capture software and on the respective case report forms [24].

Data Analysis

Sample Size

Our sample size calculation was based on recommendations from the International Organization for Standardization 80601-2-61:2019 (as the primary objective of the original study was the assessment of pulse oximeter accuracy). For our secondary analysis of the VitalPatch, we aimed to have approximately 30 complete patient data sets to ensure sufficient data points for analysis and validation testing using the VitalPatch and gold standard (Philips monitor) continuous data.

Statistical Analysis

To analyze the agreement between the VitalPatch and the gold standard during the movement phase, we compared 40-second median HR and RR estimates from the VitalPatch with simultaneous 40-second median HR and RR estimates from the Phillips monitor from each movement task interval (±20 seconds at the start and at the end, respectively, to accommodate for the 40-second window size). For the hypoxia phase (100% to 80% saturations), 40-second median HR and RR estimates were also sampled for both devices from the start until the end of hypoxia testing. For each window, simultaneous 40-second median SpO_2 estimates from the Philips monitor were also calculated per participant (our pulse oximetry hypoxia study suggested a negligible mean bias between the Philips SpO_2 and arterial oxygen saturation). The 40-second window was chosen as the VitalPatch outputs instantaneous HR and RR estimates every 4 seconds, which gives 10 points per median estimate; the Philips monitor outputs both at 1 Hz. Therefore, this represents the comparison of a 10-point median with a 40-point median centered on the same sample, which is reasonable.

For each phase, descriptive summaries for both devices are shown (number of data points and mean and SD values of HR and RR) and the MAE and root mean square error (RMSE) were computed for these vital signs. However, MAE was the metric used to determine HR and RR accuracy, in accordance with the ANSI/AAMI EC13:2002 standard [27]. It was defined as clinically acceptable if ≤5 bpm and 3 rpm, respectively. Both MAE and RMSE 95% CIs were determined via bootstrap using 10,000 replicates [33].

Bland-Altman plots, the corresponding mean bias, and the 95% LoA (SD 1.96) were also analyzed [34]. Finally, the Pearson correlation coefficient and associated *P* value [35] were reported. All statistical tests were conducted using R v3.6.1 (R Foundation for Statistical Computing) [36].

Missing Values

Participants in whom it was not possible to achieve hypoxemia (SpO_2<90%) were excluded from the analysis to avoid different sample sizes or participants in the movement and hypoxia phases. All the included participants had complete data from the movement phase. If there were any technical issues during an activity, it was repeated until complete data were obtained. For clinical reasons, this was not possible in the hypoxic phase. Participants with intermittent missing data (if not significant) were included. We did not use imputation in this study, given that the median was calculated for consecutive 40-second windows without data on either the VitalPatch or reference device excluded from the analysis.

Results

Participants

The flow diagram of the data set collected in this study is shown in Figure 2. A total of 42 participants were recruited, of which 29 (69%) were included in our analysis. The demographic descriptors are listed in Table 1. Exemplar HR and RR traces collected in the movement and hypoxia phases for one participant are shown below.

Figure 2. Participant flowchart. ABG: arterial blood gas.

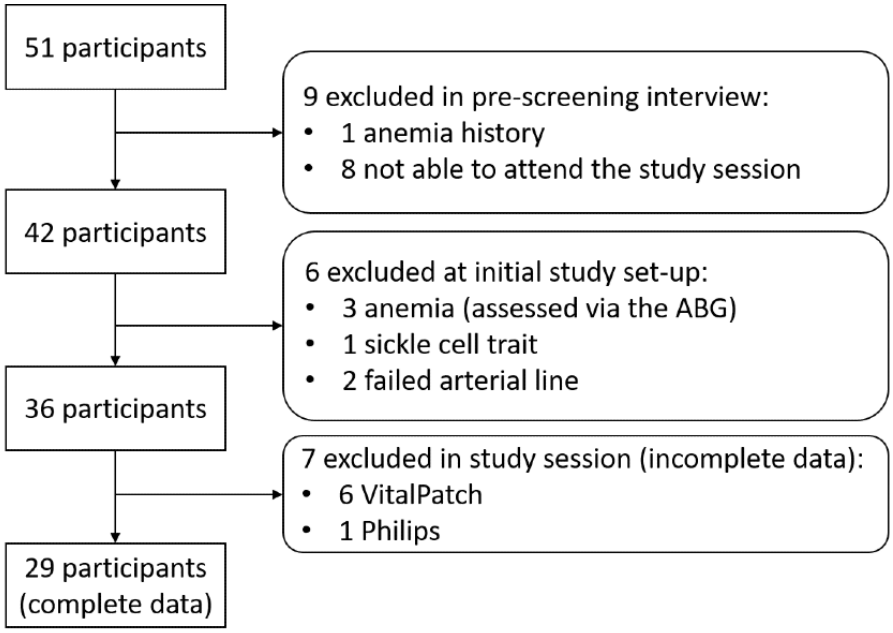


Table 1. Participants' demographics (N=29).

Characteristics	Values
Age (years)	
Mean (SD)	32.5 (10.5)
Median (IQR)	29 (25-38)
Sex (female), n (%)	16 (55)
Height (m)	
Mean (SD)	1.7 (0.1)
Median (IQR)	1.7 (1.6-1.8)
Weight (kg)	
Mean (SD)	70.6 (14.1)
Median (IQR)	68 (59.2-83)
BMI (kg/m^2)	
Mean (SD)	24.3 (3.9)
Median (IQR)	24 (21.5-26.8)
Skin tone, n (%)	
Type 1	7 (24)
Type 2	12 (41)
Type 3	2 (7)
Type 4	8 (28)
Respiratory rate (rpm[a])	
Mean (SD)	16.3 (3.6)
Median (IQR)	17 (13-19)
Heart rate (bpm[b])	
Value, mean (SD)	72.9 (11.9)
Value, median (IQR)	73 (64-82)
SaO_2[c]	
Mean (SD)	99.3 (1.3)
Median (IQR)	100 (99-100)
Systolic BP[d] (mm Hg)	
Mean (SD)	134.1 (20.0)
Median (IQR)	129 (123-146)
Diastolic BP (mm Hg)	
Mean (SD)	78.5 (11.2)
Median (IQR)	75 (70-87)

[a]rpm: respirations per minute.

[b]bpm: beats per minute.

[c]SaO_2: arterial oxygen saturation.

[d]BP: blood pressure.

Movement Phase

The movement phase lasted approximately 13 minutes for each participant. The exemplar HR and RR traces shown in Figure 3 confirm higher variability during the STS movement. The scatter and Bland-Altman plots for HR and RR in Figure 4 also show a higher mean bias and LoA for the dynamic activity (STS) versus one of the sitting movements (using a tablet). Scatter and Bland-Altman plots for all movement tests can be found in Multimedia Appendix 1. Table 2 lists the performance

metrics for the complete movement phase. The VitalPatch HR showed marginal overestimation throughout the sitting activities, with a mean MAE ranging from 0.55-0.84 bpm, RMSE from 0.89-1.32 bpm, bias from 0.06-0.32 bpm, and with a correlation of approximately 1 (P<.001) throughout. In the dynamic activity (STS), there was an increase in the MAE, 2.69 (95% CI 1.58-3.63) bpm, and RMSE 5.75 (95% CI 3.83-7.94) bpm, mean bias 1.92 (95% LoA −8.75 to 12.59) bpm, with a correlation of approximately 0.95 (P<.001).

Figure 3. (A) Exemplar heart rate and (B) respiratory rate data collected during the movement phase for one participant (vHDU-H-029). Red: VitalPatch heart rate; blue: VitalPatch respiratory rate; black: Philips 3-Lead electrocardiography heart rate in part (A) and capnography respiratory rate in part (B); brown horizontal line: movement tests. bpm: beats per minute; HR: heart rate; rpm: respirations per minute; RR: respiratory rate; STS: sit-to-stand.

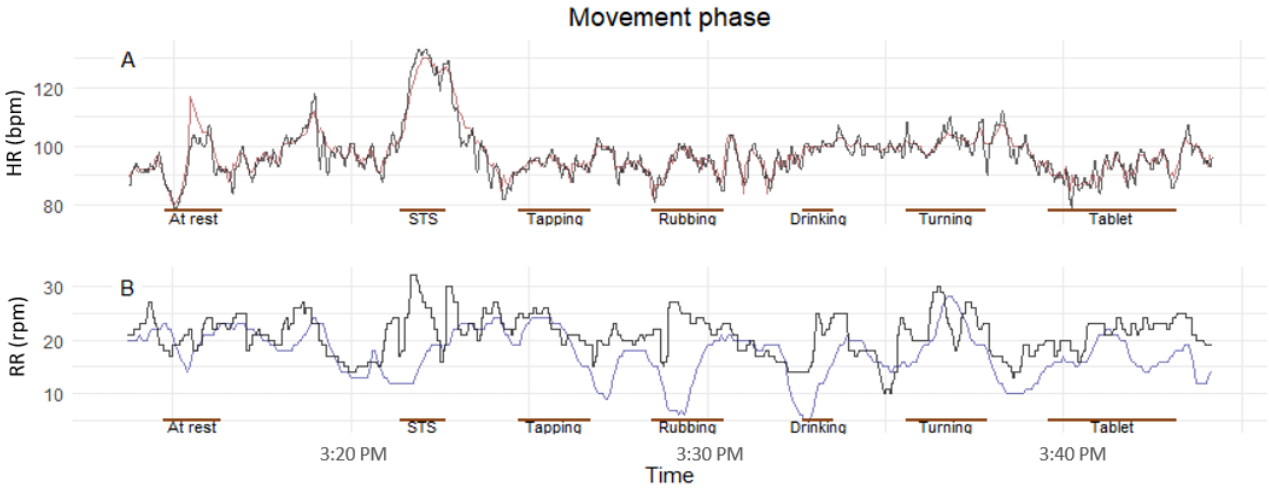


Figure 4. (A and C) Scatter plots and (B and D) Bland-Altman plots for heart rate (A and B in red) and respiratory rate (C and D in blue) in two example movements: (1) repetitive sit-to-stand and (2) using a tablet (while sitting). bpm: beats per minute; HR: heart rate; rpm: respirations per minute; RR: respiratory rate.

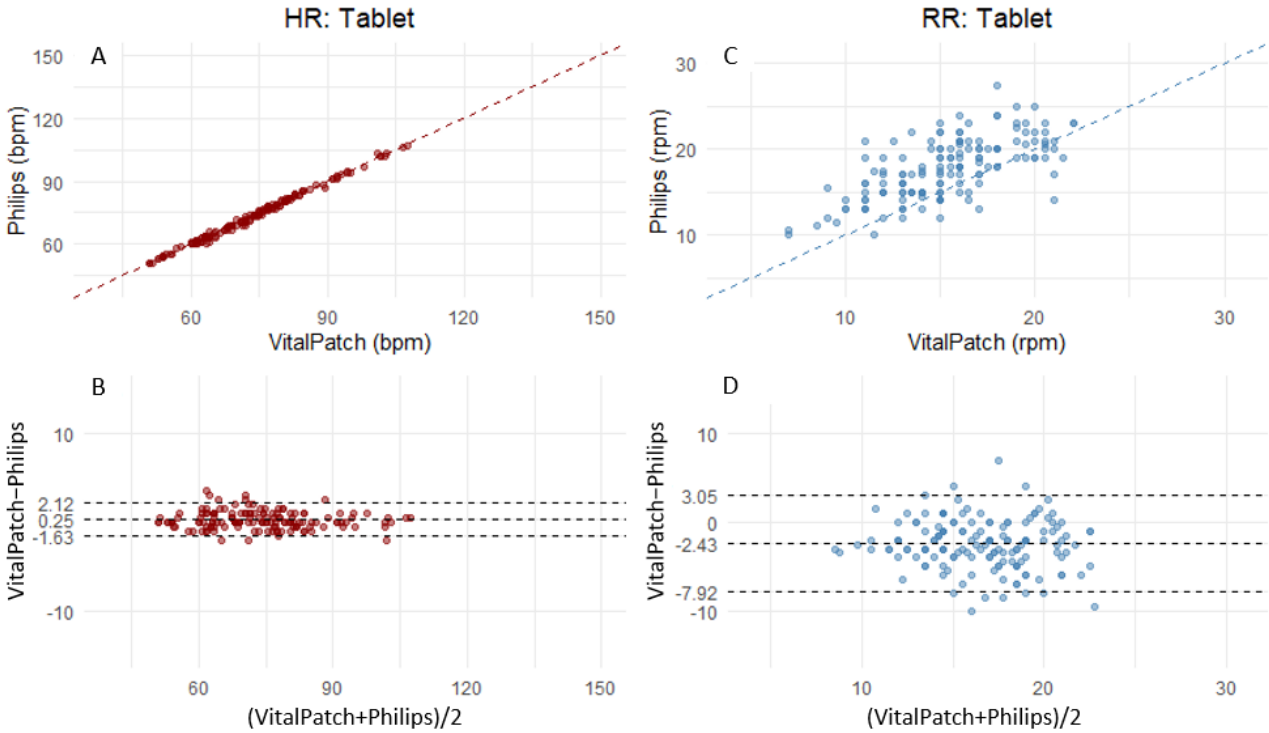


Table 2. Movement phase metrics (N=29).

Movement	Number of data points	Philips, mean (SD)	VitalPatch, mean (SD)	Pearson correlation (R^2)[a]	MAE[b] (95% CI)	RMSE[c] (95% CI)	Mean bias (95% LoA[d])
Heart rate (bpm[e])							
At rest	216	71.87 (12.52)	72.18 (12.86)	0.99	1.16 (0.91 to 1.39)	2.11 (1.62 to 2.62)	0.31 (−3.8 to 4.42)
STS[f]	93	89.94 (17.32)	91.87 (16.38)	0.95	2.69 (1.58 to 3.63)	5.75 (3.83 to 7.94)	1.92 (−8.75 to 12.59)
Tapping	111	70.99 (13.89)	71.07 (14.10)	1	0.55 (0.42 to 0.68)	0.89 (0.66 to 1.12)	0.08 (−1.68 to 1.83)
Rubbing	107	71.90 (12.69)	72.06 (12.69)	1	0.69 (0.51 to 0.85)	1.12 (0.88 to 1.37)	0.15 (−2.03 to 2.34)
Drinking	76	71.93 (13.91)	72.25 (13.68)	1	0.84 (0.61 to 1.05)	1.32 (0.98 to 1.67)	0.32 (−2.21 to 2.84)
Turning pages	98	75.48 (13.71)	75.54 (13.58)	1	0.72 (0.52 to 0.90)	1.21 (0.80 to 1.57)	0.06 (−2.32 to 2.44)
Tablet	152	73.18 (12.45)	73.43 (12.38)	1	0.69 (0.58 to 0.80)	0.99 (0.85 to 1.13)	0.25 (−1.63 to 2.12)
Respiratory rate (rpm[g])							
At rest	127	17.26 (3.29)	15.31 (3.12)	0.72	2.44 (2.11 to 2.77)	3.10 (2.70 to 3.48)	−1.94 (−6.68 to 2.79)
STS	92	18.41 (4.71)	16.83 (5.20)	0.56	3.45 (2.71 to 4.11)	4.89 (3.91 to 5.92)	−1.59 (−10.7 to 7.53)
Tapping	104	18.43 (2.75)	16.75 (3.06)	0.68	2.14 (1.76 to 2.50)	2.86 (2.40 to 3.35)	−1.68 (−6.25 to 2.89)
Rubbing	103	17.50 (3.51)	15.52 (3.13)	0.63	2.44 (1.93 to 2.88)	3.48 (2.45 to 4.38)	−1.97 (−7.63 to 3.69)
Drinking	72	16.90 (3.76)	15.46 (3.48)	0.74	2.30 (1.83 to 2.72)	2.99 (2.36 to 3.62)	−1.44 (−6.62 to 3.75)
Turning pages	97	19.09 (4.08)	17.41 (3.72)	0.53	3.05 (2.48 to 3.58)	4.13 (3.32 to 4.90)	−1.69 (−9.11 to 5.73)
Tablet	152	17.75 (3.51)	15.32 (3.23)	0.66	2.97 (2.61 to 3.31)	3.70 (3.31 to 4.10)	−2.43 (−7.92 to 3.05)

[a]$P<.001$.

[b]MAE: mean absolute error.

[c]RMSE: root mean square error.

[d]LoA: limits of agreement.

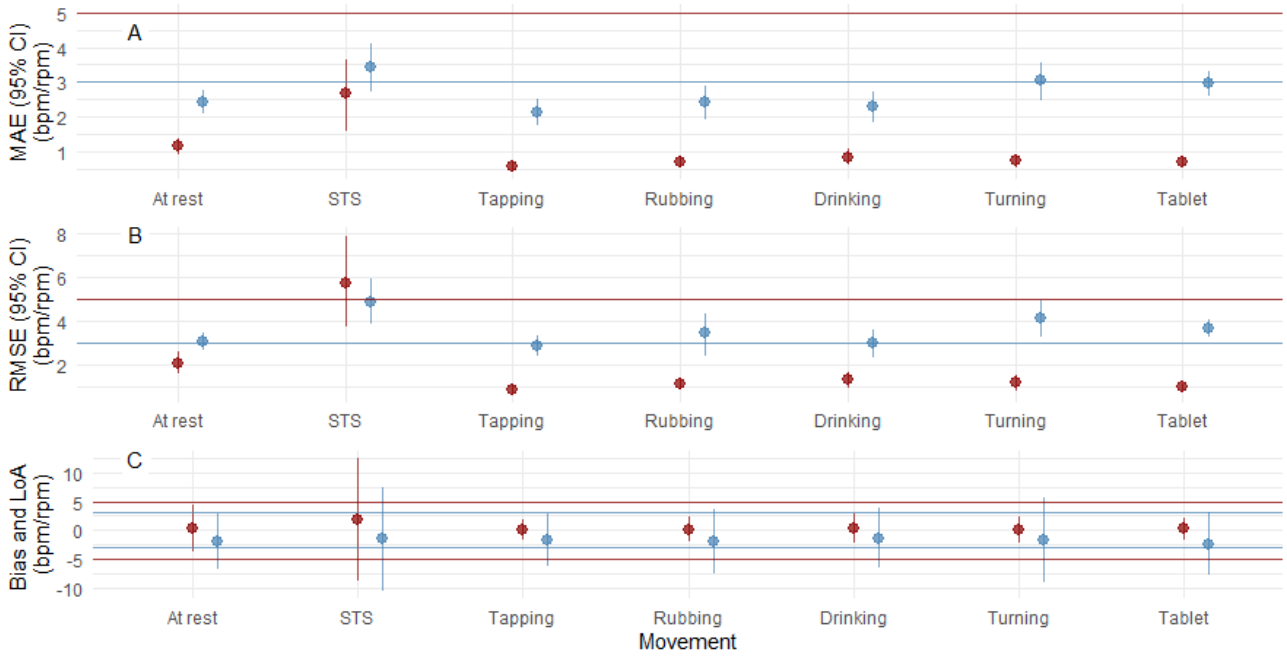
[e]bpm: beats per minute.

[f]STS: sit-to-stand.

[g]rpm: respirations per minute.

In contrast, a consistent underestimation was observed for RR throughout the sitting activities. MAE ranged from 2.14-3.05 rpm, RMSE from 2.86-4.13, mean bias from −1.68 to −2.43 rpm, and the correlation coefficient from 0.53-0.74 ($P<.001$). During the dynamic activity, the MAE was 3.45 (95% CI 2.71-4.11) rpm, RMSE was 4.89 (95% CI 3.91-5.92) rpm, mean bias was −1.59 (95% LoA −10.7 to 7.53) rpm, and the correlation was 0.56 ($P<.001$). An overview of the MAE, RMSE, and mean bias (LoA) for all movements is shown in Figure 5.

Figure 5. Accuracy and Bias (mean bias) plots for all movement tests. Red: heart rate; horizontal red line represents acceptable limits. Blue: respiratory rate; horizontal blue line represents acceptable limits. (A) MAE (95% CI) plot, (B) RMSE (95% CI) plot, (C) bias LOAs. bpm: beats per minute; LoAs: limits of agreement; MAE: mean absolute error; RMSE: root mean square error; rpm: respirations per minute; STS: sit-to-stand.



The VitalPatch HR estimation was within clinically acceptable accuracy, achieving a MAE 0.72 (95% CI 0.66-0.78) bpm and a RMSE 1.16 (95% CI 1.02-1.29) bpm in the hypoxia phase. Figure 6 shows an exemplar participant during the hypoxia phase, and Figure 7 shows the scatter and Bland-Altman plots for HR and RR. The mean bias was 0.49 (95% LoA −1.58 to 2.56), and the correlation was approximately 1 (P<.001). Although more variable, the RR was also within the acceptable accuracy achieving a MAE 1.89 (95% CI 1.75-2.03) rpm and a RMSE 2.92 (95% CI 2.64-3.2) rpm. A minimal mean bias of −0.08 (95% LoA −5.8 to 5.65) and a correlation of approximately 0.72 (P<.001) were observed.

Figure 6. Exemplar heart rate (A) and respiratory rate (B) traces during the hypoxia phase for one participant (vHDU-H-29). red: VitalPatch heart rate; blue: VitalPatch respiratory rate; black: Philips 3-Lead electrocardiography heart rate (A) and capnography respiratory rate (B); brown horizontal line: target hypoxia levels. bpm: beats per minute; HR: heart rate; rpm: respirations per minute; RR: respiratory rate.

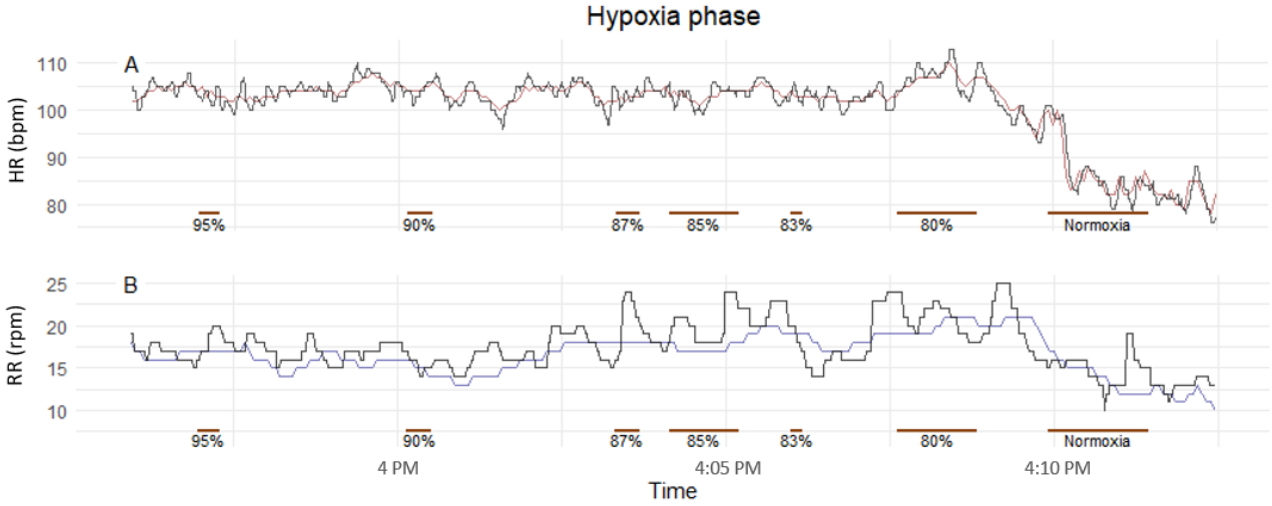
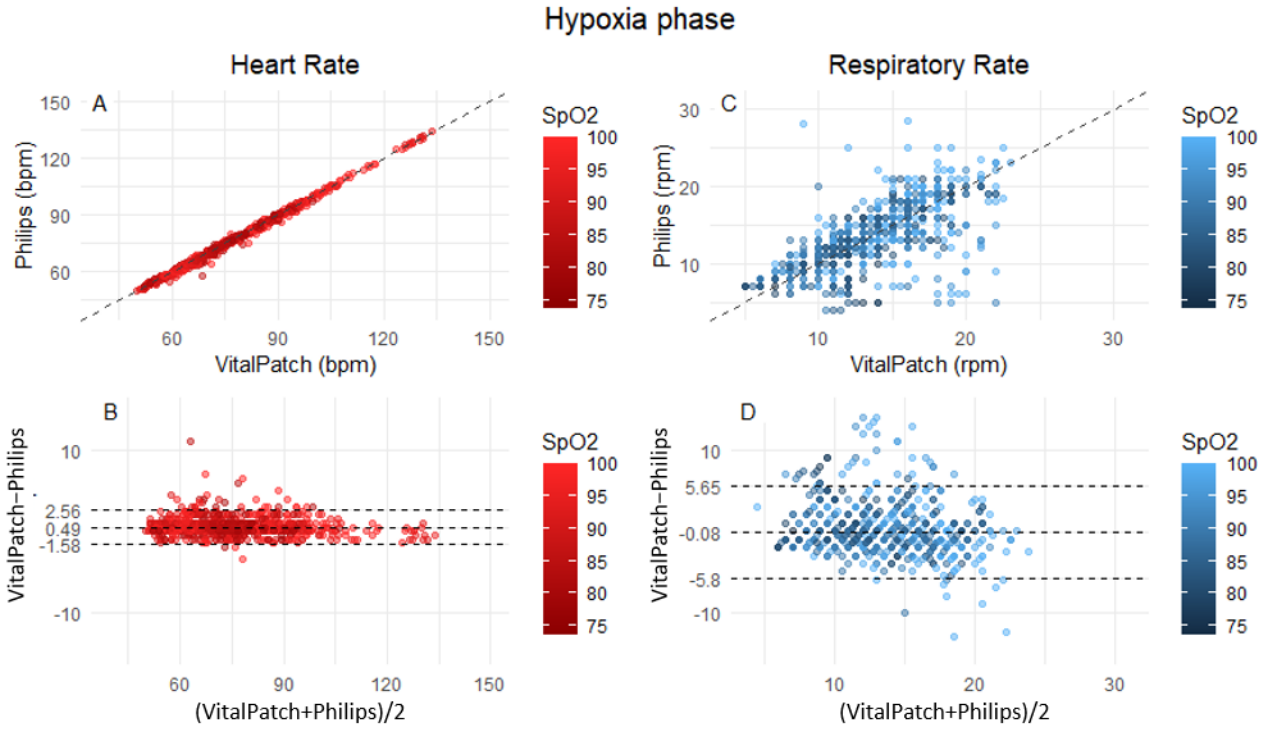
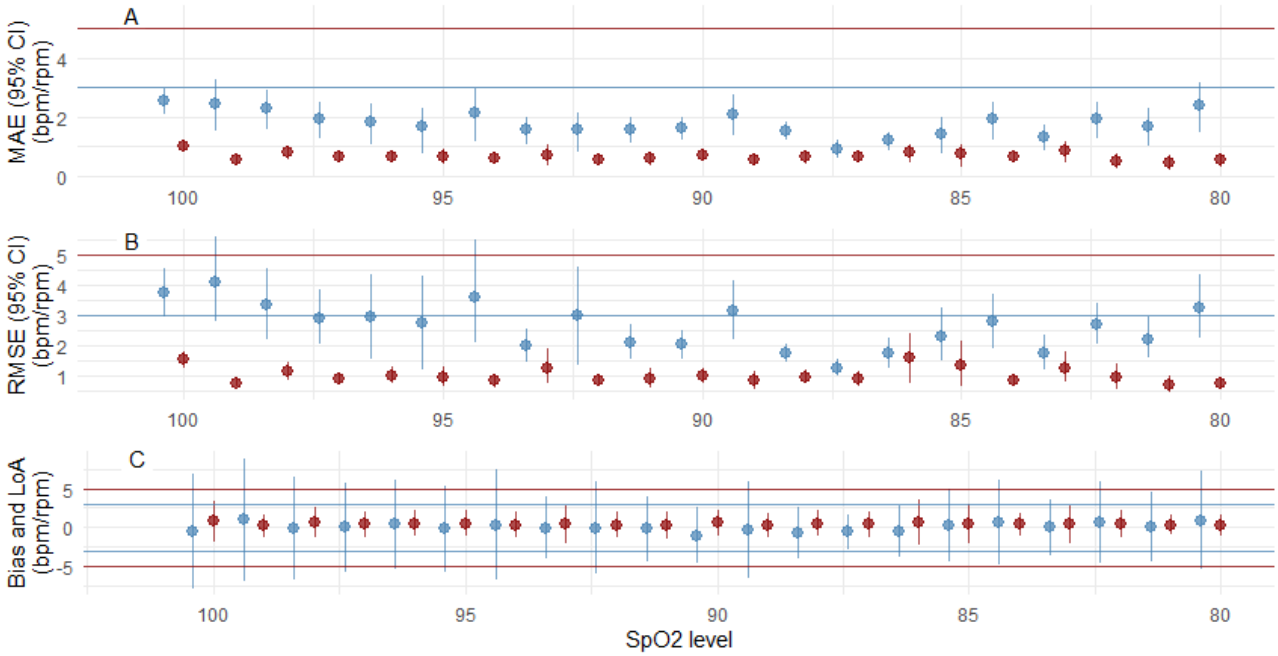


Figure 7. (A and C) Scatter plots and (B and D) Bland-Altman plots for heart rate (A and B in red) and respiratory rate (C and D in blue) during the full hypoxia phase. bpm: beats per minute; rpm: respirations per minute; SpO_2: peripheral oxygen saturation.



There were also no significant differences in the error metrics between the SpO_2 subgroups (normoxia, SpO_2 100% to 90%; mild hypoxia, SpO_2 85%-89.9%; and severe hypoxia, SpO_2<85% [24]), as shown in Multimedia Appendix 2. Finally, Figure 8 shows the MAE, RMSE, and mean bias (LoA) computed for each oxygen saturation level from the hypoxia phase.

Figure 8. Accuracy and Bias (mean bias) plots for the hypoxia phase. Red: heart rate; horizontal red lines represent acceptable limits Blue: respiratory rate; horizontal blue line represents acceptable limits. (A) MAE (95% CI) plot, (B) RMSE (95% CI) plot, (C) bias LOAs. bpm: beats per minute; LoA:s limits of agreement; MAE: mean absolute error; RMSE: root mean square error; rpm: respirations per minute; SpO_2: peripheral oxygen saturation.



Discussion

Principal Findings

The key findings from this study were as follows: (1) The VitalPatch HR estimation was well within acceptable accuracy limits when compared with the gold standard device (Philips 3-lead ECG), (2) The VitalPatch RR estimation was also in agreement with the gold standard device (Philips capnography) for most of the movements and throughout the full hypoxia phase, and (3) there was no impact on the HR and RR estimation performance when participants were exposed to induced hypoxia. These findings support the clinical use of VitalPatch in patients with reduced baseline saturation or *at risk* of hypoxia.

Accuracy testing during motion is important before clinical use, as we aim to implement an ambulatory monitoring system that promotes patient mobility and independence without compromising safety. Our results suggest that the VitalPatch can measure HR with high accuracy and minimal bias during simple movements (mimicking in-hospital patient movements). Reliability was reduced during STS activity but was still within acceptable limits. These results agree with the results of previous clinical validation studies when compared with a reference standard device in healthy volunteers [18,23] and surgical patients [21,22]. When compared with nurse manual measurements in surgical patients [14], VitalPatch meets the accuracy criteria for HR estimation.

The MAE of the VitalPatch RR estimation was comparatively higher, especially during the movement phase. However, it was also mostly within the prespecified limits (except for the STS and turning a page movements), in agreement with previous studies [14,18,21-23]. High variation during STS was expected as it is a dynamic activity, potentially creating several artifacts that could affect VitalPatch accuracy during this test [21,22]; however, a high MAE for RR was not expected for the *turning pages* activity (even if close to the 3-rpm threshold). We postulate that the latter may have created a low-frequency periodic artifact that influenced both the ECG and accelerometer waveforms (both used in the computation of the RR estimate by the VitalPatch).

The most recent study showed that the VitalPatch tended to overestimate RR with a bias of 4.4 rpm [22], but previous studies showed a predominant RR underestimation [14,18,21]. Our study seems to agree with the latter, as VitalPatch mostly underestimated RR throughout the movement phase and varied in the hypoxia phase. However, in contrast to these previous studies, we collected vital signs from healthy participants.

This study also assessed the VitalPatch reliability during a controlled hypoxia phase, and our results suggest that there is no impact on this patch's reliability for HR and RR measurements during periods of reduced SpO_2 levels. The VitalPatch estimates followed the gold standard HR and RR variations caused by the controlled desaturation of the participants during the hypoxia phase, reliably, supporting its use in clinical populations with reduced baseline saturation or *at risk* of hypoxia.

Limitations

This study had some limitations. First, the participants were healthy volunteers in a controlled environment that is not representative of hospitalized patients. In addition, the main study was powered to assess the accuracy of SpO_2 and not that of HR and RR. Although the range required to validate RR was achieved (10 to 30 rpm is required [18] and 4 to 31 rpm was collected), the HR range achieved in this substudy was 40 to 150 bpm, a subset of the range required to validate its accuracy, 30 to 200 bpm [27].

Comparison With Prior Work and Future Research

The ANSI/AAMI EC13:2002 standard [27] only requires that a measure of systematic error (eg, mean bias or MAE) be below the clinically accepted limits. In more recent studies, MAE has been used because it is less susceptible to outliers than the mean bias [18,23,30]. Analyzing the RMSE results, which combines both the systematic and random errors, we note that they follow the trend of those of the MAE metric as the RMSE was also higher during the STS activity for both variables and for the *turning pages* movement for RR. In our study, the LoA (also a measure of random error) seemed to be within the acceptable range when compared with the results of other studies [14,18,21-23] (except for the aforementioned cases). Previous studies reported higher LoA (defined as out of the acceptable range by the authors) for RR when compared with reference standard devices both in healthy volunteers (−10.1 to 6.5) [18] and surgical patients (−15.8 to 11.2) [21] and (−4.4 to 13.3) [22], and when compared with nurse manual measurements in surgical patients (10.32 to 9.04) [14].

In this study, the VitalPatch was subjected to robust testing, during movement and under induced hypoxia, against a gold standard in a controlled environment. Our results suggest that this device can be safely tested in clinical environments (within the collected input range for each vital sign). Our results will inform future phases of our vHDU research project, aiming to develop an accurate wearable system for monitoring ambulatory patients in hospitals.

Conclusions

In conclusion, the VitalPatch was highly accurate throughout movement tests except for the measurement of RR during the STS activity and the *turning pages* movement. This device was also reliable throughout the hypoxia phase, with no significant MAE or bias differences in hypoxic levels. This study demonstrated that VitalPatch can be safely tested in clinical environments to support earlier detection of clinical deterioration.

Acknowledgments

CMA, MS, LY, MH, and SV were funded by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health. AS is currently supported by the NIHR Doctoral Research Fellowship (NIHR-DRF-2017-10-094). The authors would like to thank all the volunteers who participated in this study.

Authors' Contributions

Authorship was determined in accordance with the International Committee of Medical Journal Editors guidelines. CMA, SV, PP, LT, and PW drafted the initial protocol. CMA, SV, EK, JE, LY, OG, CR, MP, and MS conducted the study procedures and data acquisition. AS and MH provided medical coverage on the study days. CMA and MS conducted statistical analyses. CMA drafted the manuscript, and all the authors reviewed and approved the manuscript. The funders had no role in the study protocol design or in the collection, management, analysis, and interpretation of the data or the writing of the final report.

Conflicts of Interest

PW and LT report significant grants from the NIHR, United Kingdom, and the NIHR Biomedical Research Centre, Oxford, during the conduct of the study. PW and LT report modest grants and personal fees from Sensyne Health outside the submitted work. LT work part-time for Sensyne Health and hold shares in the company. PW was previously Chief Medical officer for Sensyne Health and has shares in the company. Sensyne Health have provided research support to his university department.

Multimedia Appendix 1
Scatter and Bland-Altman plots for movement phase.
[DOCX File , 160 KB - jmir_v23i9e27547_app1.docx]

Multimedia Appendix 2
Hypoxia phase subgroups analysis.
[DOCX File , 176 KB - jmir_v23i9e27547_app2.docx]

References

1. Sun L, Joshi M, Khan S, Ashrafian H, Darzi A. Clinical impact of multi-parameter continuous non-invasive monitoring in hospital wards: a systematic review and meta-analysis. J R Soc Med 2020 Jun;113(6):217-224 [FREE Full text] [doi: 10.1177/0141076820925436] [Medline: 32521195]
2. McQuillan P, Pilkington S, Allan A, Taylor B, Short A, Morgan G, et al. Confidential inquiry into quality of care before admission to intensive care. Br Med J 1998 Jun 20;316(7148):1853-1858 [FREE Full text] [doi: 10.1136/bmj.316.7148.1853] [Medline: 9632403]
3. Leenen J, Leerentveld C, van Dijk JD, van Westreenen HL, Schoonhoven L, Patijn G. Current evidence for continuous vital signs monitoring by wearable wireless devices in hospitalized adults: systematic review. J Med Internet Res 2020 Jun 17;22(6):e18636 [FREE Full text] [doi: 10.2196/18636] [Medline: 32469323]
4. Downey CL, Tahir W, Randell R, Brown JM, Jayne DG. Strengths and limitations of early warning scores: a systematic review and narrative synthesis. Int J Nurs Stud 2017 Nov;76:106-119. [doi: 10.1016/j.ijnurstu.2017.09.003] [Medline: 28950188]
5. Clifton DA, Clifton L, Sandu D, Smith GB, Tarassenko L, Vollam SA, et al. 'Errors' and omissions in paper-based early warning scores: the association with changes in vital signs--a database analysis. BMJ Open 2015 Jul 03;5(7):e007376 [FREE Full text] [doi: 10.1136/bmjopen-2014-007376] [Medline: 26141302]
6. Jansen JO, Cuthbertson BH. Detecting critical illness outside the ICU: the role of track and trigger systems. Curr Opin Crit Care 2010 Jun;16(3):184-190. [doi: 10.1097/MCC.0b013e328338844e] [Medline: 20305556]
7. Tarassenko L, Hann A, Young D. Integrated monitoring and analysis for early warning of patient deterioration. Br J Anaesth 2006 Jul;97(1):64-68 [FREE Full text] [doi: 10.1093/bja/ael113] [Medline: 16707529]
8. Downey C, Randell R, Brown J, Jayne DG. Continuous versus intermittent vital signs monitoring using a wearable, wireless patch in patients admitted to surgical wards: pilot cluster randomized controlled trial. J Med Internet Res 2018 Dec 11;20(12):e10802 [FREE Full text] [doi: 10.2196/10802] [Medline: 30538086]
9. Joshi M, Ashrafian H, Aufegger L, Khan S, Arora S, Cooke G, et al. Wearable sensors to improve detection of patient deterioration. Expert Rev Med Devices 2019 Feb;16(2):145-154. [doi: 10.1080/17434440.2019.1563480] [Medline: 30580650]
10. Watkins T, Whisman L, Booker P. Nursing assessment of continuous vital sign surveillance to improve patient safety on the medical/surgical unit. J Clin Nurs 2016 Jan;25(1-2):278-281 [FREE Full text] [doi: 10.1111/jocn.13102] [Medline: 26531215]

11. Weller RS, Foard KL, Harwood TN. Evaluation of a wireless, portable, wearable multi-parameter vital signs monitor in hospitalized neurological and neurosurgical patients. J Clin Monit Comput 2018 Oct;32(5):945-951. [doi: 10.1007/s10877-017-0085-0] [Medline: 29214598]
12. Paul JE, Chong MA, Buckley N, Harsha P, Shanthanna H, Tidy A, et al. Vital sign monitoring with continuous pulse oximetry and wireless clinical notification after surgery (the VIGILANCE pilot study)-a randomized controlled pilot trial. Pilot Feasibility Stud 2019 Feb 26;5:36 [FREE Full text] [doi: 10.1186/s40814-019-0415-8] [Medline: 30858986]
13. Hamilton P, Curley M, Aimi R, Sae-Hau C. Comparison of methods for adaptive removal of motion artifact. In: Proceedings of the Computers in Cardiology Conference. 2000 Presented at: Computers in Cardiology 2000; Sep 24-27, 2000; Cambridge, MA, USA. [doi: 10.1109/cic.2000.898537]
14. Weenk M, van Goor H, Frietman B, Engelen L, van Laarhoven CJ, Smit J, et al. Continuous monitoring of vital signs using wearable devices on the general ward: pilot study. JMIR Mhealth Uhealth 2017 Jul 05;5(7):e91 [FREE Full text] [doi: 10.2196/mhealth.7208] [Medline: 28679490]
15. Verrillo S, Cvach M, Hudson K, Winters B. Using continuous vital sign monitoring to detect early deterioration in adult postoperative inpatients. J Nurs Care Qual 2019;34(2):107-113. [doi: 10.1097/NCQ.0000000000000350] [Medline: 30095509]
16. VitalPatch. VitalConnect. URL: https://vitalconnect.com/solutions/vitalpatch/ [accessed 2021-08-09]
17. Instructions for use, VitalPatch® Biosensor. VitalConnect. URL: https://vitalconnect.com/docs/ifu006/revS/IFU-06 [accessed 2021-08-09]
18. Selvaraj N, Nallathambi G, Moghadam R, Aga A. Fully disposable wireless patch sensor for continuous remote patient monitoring. Annu Int Conf IEEE Eng Med Biol Soc 2018 Jul;2018:1632-1635. [doi: 10.1109/EMBC.2018.8512569] [Medline: 30440706]
19. Tonino R, Larimer K, Eissen O, Schipperus M. Remote patient monitoring in adults receiving transfusion or infusion for hematological disorders using the VitalPatch and accelerateIQ monitoring system: quantitative feasibility study. JMIR Hum Factors 2019 Dec 02;6(4):e15103 [FREE Full text] [doi: 10.2196/15103] [Medline: 31789596]
20. Areia C, Young L, Vollam S, Ede J, Santos M, Tarassenko L, et al. Wearability testing of ambulatory vital sign monitoring devices: prospective observational cohort study. JMIR Mhealth Uhealth 2020 Dec 16;8(12):e20214 [FREE Full text] [doi: 10.2196/20214] [Medline: 33325827]
21. Breteler MJ, Huizinga E, van Loon K, Leenen LP, Dohmen DA, Kalkman CJ, et al. Reliability of wireless monitoring using a wearable patch sensor in high-risk surgical patients at a step-down unit in the Netherlands: a clinical validation study. BMJ Open 2018 Feb 27;8(2):e020162 [FREE Full text] [doi: 10.1136/bmjopen-2017-020162] [Medline: 29487076]
22. Breteler MJ, KleinJan EJ, Dohmen DA, Leenen LP, van Hillegersberg R, Ruurda JP, et al. Vital signs monitoring with wearable sensors in high-risk surgical patients: a clinical validation study. Anesthesiology 2020 Mar;132(3):424-439 [FREE Full text] [doi: 10.1097/ALN.0000000000003029] [Medline: 31743149]
23. Chan AM, Selvaraj N, Ferdosi N, Narasimhan R. Wireless patch sensor for remote monitoring of heart rate, respiration, activity, and falls. Annu Int Conf IEEE Eng Med Biol Soc 2013;2013:6115-6118. [doi: 10.1109/EMBC.2013.6610948] [Medline: 24111135]
24. Areia C, Vollam S, Piper P, King E, Ede J, Young L, et al. Protocol for a prospective, controlled, cross-sectional, diagnostic accuracy study to evaluate the specificity and sensitivity of ambulatory monitoring systems in the prompt detection of hypoxia and during movement. BMJ Open 2020 Jan 12;10(1):e034404 [FREE Full text] [doi: 10.1136/bmjopen-2019-034404] [Medline: 31932393]
25. Nelson BW, Low CA, Jacobson N, Areán P, Torous J, Allen NB. Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. NPJ Digit Med 2020 Jun 26;3:90 [FREE Full text] [doi: 10.1038/s41746-020-0297-4] [Medline: 32613085]
26. Subbe CP, Kinsella S. Continuous monitoring of respiratory rate in emergency admissions: evaluation of the RespiraSense™ sensor in acute care compared to the industry standard and gold standard. Sensors (Basel) 2018 Aug 17;18(8):2700 [FREE Full text] [doi: 10.3390/s18082700] [Medline: 30126085]
27. Cardiac monitors, heart rate meters, alarms. Association for the Advancement of Medical Instrumentation (AAMI). URL: https://global.ihs.com/doc_detail.cfm?document_name=AAMI%20EC13&item_s_key=00134307 [accessed 2021-08-09]
28. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol 2011 Jan;64(1):96-106. [doi: 10.1016/j.jclinepi.2010.03.002] [Medline: 21130355]
29. Yasuma F, Hayano J. Respiratory sinus arrhythmia: why does the heartbeat synchronize with respiratory rhythm? Chest 2004 Feb;125(2):683-690. [doi: 10.1378/chest.125.2.683] [Medline: 14769752]
30. Chan AM, Ferdosi N, Narasimhan R. Ambulatory respiratory rate detection using ECG and a triaxial accelerometer. Annu Int Conf IEEE Eng Med Biol Soc 2013;2013:4058-4061. [doi: 10.1109/EMBC.2013.6610436] [Medline: 24110623]
31. Fitzpatrick T. Soleil et Peau. J Med Esthet 1975;2:33-34 [FREE Full text]
32. ixTrend. ixitos. URL: https://www.ixitos.com/en/products/ixtrend/ [accessed 2021-08-09]
33. Davison AC, Hinkley DV. Bootstrap Methods and their Application. Cambridge: Cambridge University Press; Oct 1997.

34. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. J Biopharm Stat 2007;17(4):571-582. [doi: 10.1080/10543400701329422] [Medline: 17613642]
35. Tiit E. Nonparametric statistical methods. In: Statistics in Medicine. Hoboken, New Jersey: Wiley; May 12, 2000.
36. R: a language and environment for statistical computing. The R Foundation. URL: https://www.r-project.org/ [accessed 2021-08-09]

Abbreviations

ANSI/AAMI: American National Standards Institute/Association for the Advancement of Medical Instrumentation
bpm: beats per minute
ECG: electrocardiography
HR: heart rate
LoA: limits of agreement
MAE: mean absolute error
NIHR: National Institute for Health Research
RMSE: root mean square error
rpm: respirations per minute
RR: respiratory rate
SpO_2: peripheral oxygen saturation
STS: sit-to-stand
vHDU: virtual high dependency unit

Edited by R Kukafka; submitted 28.01.21; peer-reviewed by C Downey, M Breteler; comments to author 16.03.21; revised version received 15.04.21; accepted 21.06.21; published 15.09.21.

Please cite as:
Morgado Areia C, Santos M, Vollam S, Pimentel M, Young L, Roman C, Ede J, Piper P, King E, Gustafson O, Harford M, Shah A, Tarassenko L, Watkinson P
A Chest Patch for Continuous Vital Sign Monitoring: Clinical Validation Study During Movement and Controlled Hypoxia
J Med Internet Res 2021;23(9):e27547
URL: https://www.jmir.org/2021/9/e27547
doi:10.2196/27547
PMID:34524087

Original Paper

A Smartphone-Based App to Improve Adjuvant Treatment Adherence to Multidisciplinary Decisions in Patients With Early-Stage Breast Cancer: Observational Study

Jing Yu[1], MD; Jiayi Wu[1], PhD; Ou Huang[1], PhD; Xiaosong Chen[1], PhD; Kunwei Shen[1], PhD

Department of General Surgery, Comprehensive Breast Health Center, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Corresponding Author:
Xiaosong Chen, PhD
Department of General Surgery
Comprehensive Breast Health Center
Ruijin Hospital, Shanghai Jiao Tong University School of Medicine
197 Ruijin Er Road
Shanghai, 200025
China
Phone: 86 13564497086
Email: chenxiaosong0156@hotmail.com

Abstract

Background: Multidisciplinary treatment (MDT) and adjuvant therapy are associated with improved survival rates in breast cancer. However, nonadherence to MDT decisions is common in patients. We developed a smartphone-based app that can facilitate the full-course management of patients after surgery.

Objective: This study aims to investigate the influence factors of treatment nonadherence and to determine whether this smartphone-based app can improve the compliance rate with MDTs.

Methods: Patients who had received a diagnosis of invasive breast cancer and had undergone MDT between March 2013 and May 2019 were included. Patients were classified into 3 groups: Pre-App cohort (November 2017, before the launch of the app); App nonused, cohort (after November 2017 but not using the app); and App used cohort (after November 2017 and using the app). Univariate and multivariate analyses were performed to identify the factors related to MDT adherence. Compliance with specific adjuvant treatments, including chemotherapy, radiotherapy, endocrine therapy, and targeted therapy, was also evaluated.

Results: A total of 4475 patients were included, with Pre-App, App nonused, and App used cohorts comprising 2966 (66.28%), 861 (19.24%), and 648 (14.48%) patients, respectively. Overall, 15.53% (695/4475) patients did not receive MDT recommendations; the noncompliance rate ranged from 27.4% (75/273) in 2013 to 8.8% (44/500) in 2019. Multivariate analysis demonstrated that app use was independently associated with adherence to adjuvant treatment. Compared with the patients in the Pre-App cohort, patients in the App used cohort were less likely to deviate from MDT recommendations (odds ratio [OR] 0.61, 95% CI 0.43-0.87; *P*=.007); no significant difference was found in the App nonused cohort (*P*=.77). Moreover, app use decreased the noncompliance rate for adjuvant chemotherapy (OR 0.41, 95% CI 0.27-0.65; *P*<.001) and radiotherapy (OR 0.49, 95% CI 0.25-0.96; *P*=.04), but not for anti-HER2 therapy (*P*=.76) or endocrine therapy (*P*=.39).

Conclusions: This smartphone-based app can increase MDT adherence in patients undergoing adjuvant therapy; this was more obvious for adjuvant chemotherapy and radiotherapy.

(J Med Internet Res 2021;23(9):e27576) doi:10.2196/27576

KEYWORDS

breast cancer; adherence; multidisciplinary treatment; adjuvant treatment; smartphone-based app; mobile phone

Introduction

Background

Breast cancer has the highest prevalence rate among malignant diseases in women worldwide [1]. Typically, comprehensive treatment for breast cancer includes locoregional and systemic therapy approaches. The former refers to surgery and radiotherapy, and the latter comprises chemotherapy, targeted therapy, endocrine therapy, and other promising strategies such as immunotherapy [2]. Moreover, multidisciplinary treatment (MDT), which involves physicians from different disciplines with specialized knowledge working as team to discuss the treatment of a given patient, has become a standard care modality that can promote clinical decision-making and improve the overall quality of treatment [3]. To date, nearly 70% to 80% of nonmetastatic breast cancers are curable because of the continuous improvement of therapeutic strategies and MDT discussions [4].

Despite the proven efficacy of adjuvant treatment, patients may have difficulty initiating or pursuing a treatment plan because of the relative complexity of the overall therapy. A cancer registry study demonstrated that the nonadherence rate for adjuvant therapy was approximately 30% for early breast cancer [5]. Meanwhile, it has been reported that nearly one-third of patients discontinued endocrine therapy during 5 years of treatment [6]. Noncompliance was associated with an increased risk of recurrence and poorer clinical outcomes [7,8]. Thus, it is essential to identify the reasons for nonadherence and use effective interventions to improve adherence.

Mobile Health Intervention

Currently, the use of mobile devices to conduct health control and management has become increasingly popular and has the opportunity to affect health behaviors, especially in the condition of chronic disease and cancer setting among adolescents [9]. For example, sharing treatment experiences on the web can help alleviate isolation and emotional distress [10]. Active and regular self-monitoring using the digital app also helps to improve health-related outcomes, such as pain management for cancer [11]. Studies have also argued that the change in actual behavior is modest, and there is a lack of research focusing on effectiveness and acceptability [12,13]. New technologies such as artificial intelligence, virtual reality, and machine learning have also been tailored in health care settings to optimize disease outcomes [14]. Therefore, we launched a smartphone-based app called the *full-course management system* in November 2017, which enabled sustained communication between outpatients and medical workers and provided personalized management according to MDT discussions. Using this app, we aimed to facilitate the postoperative management of patients with breast cancer.

Objective

In this study, we aimed to determine the potential factors that influence patient compliance with MDTs. Meanwhile, we aimed to evaluate whether this smartphone-based app could improve patients' adherence to MDT for early breast cancer patients.

Methods

Data Processing and Cohort Sorting

Patients with a diagnosis of breast cancer who received MDT discussion for adjuvant treatment decisions between March 2013 and May 2019 at Shanghai Ruijin Hospital were retrospectively reviewed. Data on demographic and clinicopathological characteristics as well as follow-up information were retrieved from the Shanghai Jiao Tong University Breast Cancer Database. The inclusion criteria were as follows: (1) female patients with a diagnosis of primary breast cancer and (2) patients who received surgical treatment. Patients were excluded if they met the following criteria: (1) forgone or had missing MDT discussion data, (2) pathologically confirmed without invasive lesions, or (3) missing clinicopathological characteristic data. Patients were then divided into 3 cohorts according to the year of diagnosis and use of the smartphone-based management system. The Pre-App cohort included patients who received a diagnosis between March 2013 and October 2017. The App nonused cohort included patients who received a diagnosis between November 2017 and May 2019 without using the app. The App used cohort referred to patients who received a diagnosis between November 2017 and May 2019 and who had applied the smartphone-based app within the adjuvant treatment setting. It is worth noting that the use of the app was entirely up to patients' willingness, and at least one time of logging record identified from the backstage system would be defined as app used.

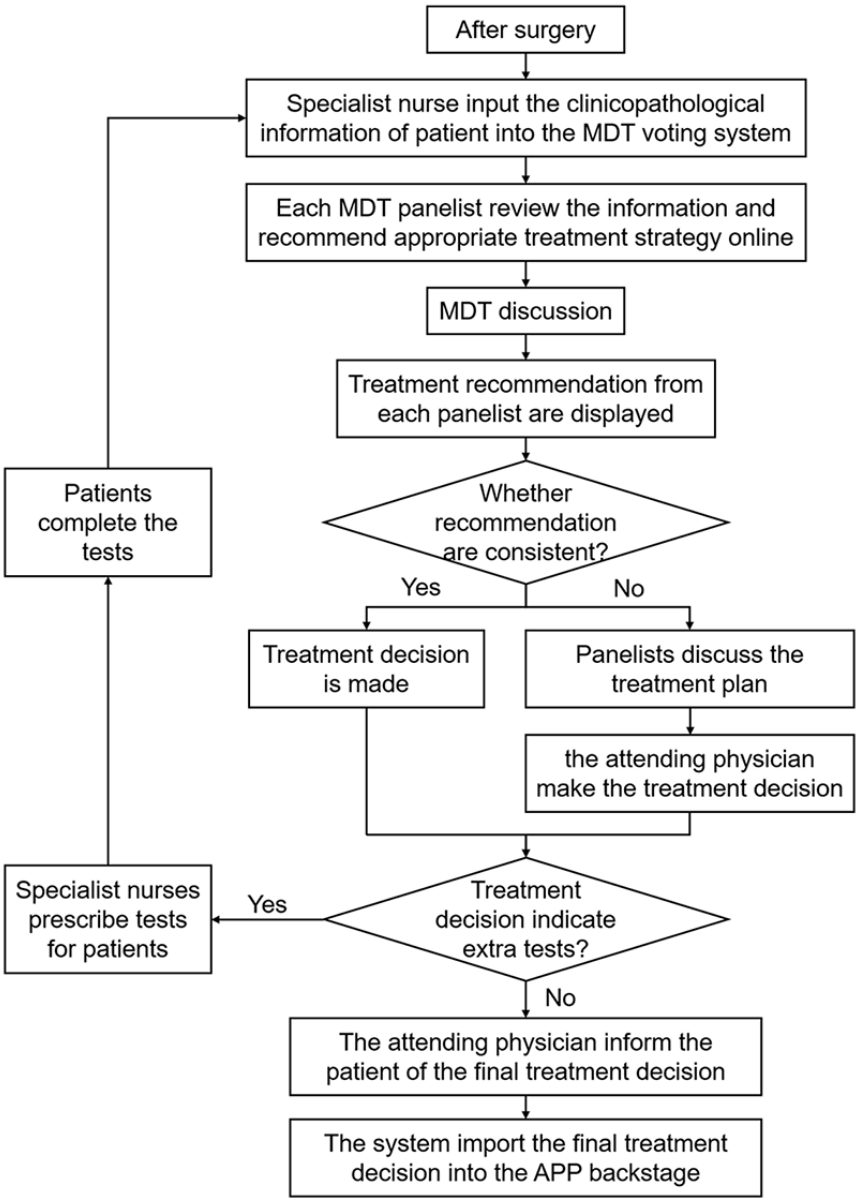
Ethics Approval and Consent to Participate

The study was approved by the Ethical Committees of Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, and informed consent was waived because of the retrospective design of this study. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

MDT Discussion

After breast cancer surgery, an MDT discussion session was held to determine the adjuvant treatment strategy for each patient. In brief, the MDT panel members reviewed the medical history of each patient in advance and selected treatments on a given webpage. The MDT panelists consisted of breast surgeons, medical oncologists, pathologists, radiation oncologists, and oncology nurses. When the MDT meeting was held, a physician reported the cases and displayed the treatment results of each panelist for each patient. If the treatment results were consistent, a treatment decision was made. If the treatment decisions were inconsistent, the panel members discussed the treatment plan, and the attending physician made the final treatment decision for the patient. The flowchart of the MDT is shown in Figure 1.

Figure 1. Flowchart of multidisciplinary treatment. MDT: multidisciplinary treatment.



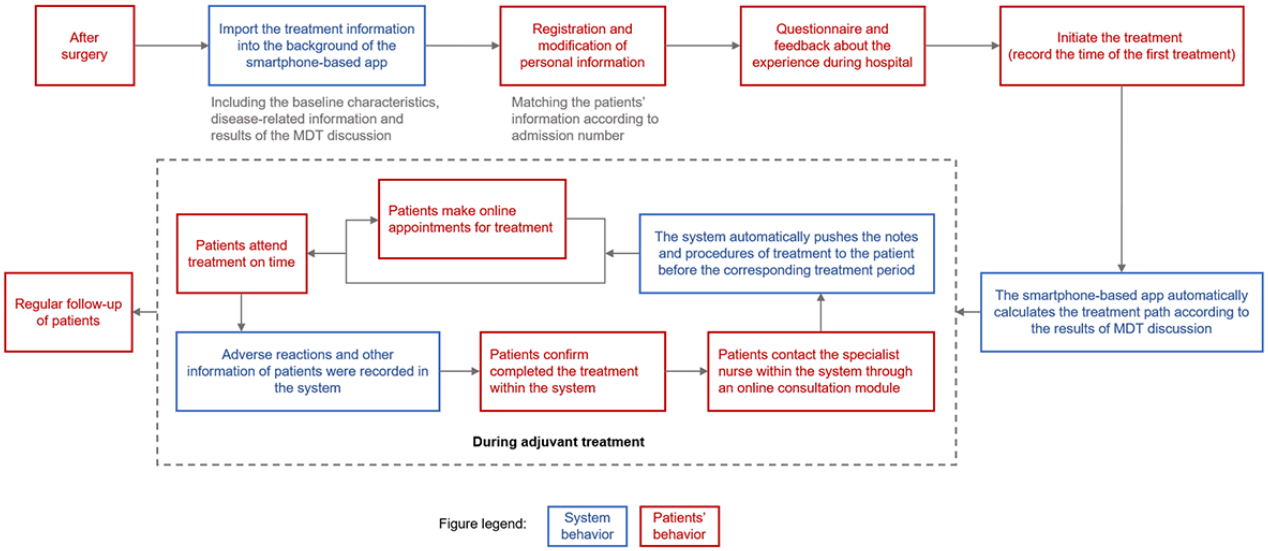
Smartphone-Based Full-Course Management App

The smartphone-based app consists of the following features:

1. Login, logout, setting, and modification of personal information
2. Full-course management included registering for information on the surgical treatment, viewing the results of the MDT discussion, generating the treatment process according to the MDT discussion, receiving the reminder of the treatment, confirming the completion of the treatment, filling in the follow-up information, and consulting a specialist nurse.
3. Questionnaire and feedback

The operational flow of the app is illustrated in Figure 2. In brief, after discharge from the hospital, the baseline characteristics and surgical and pathological information of the patients were uploaded to the system. Then, the patients completed the registration through the smartphone app and initiated the first course of treatment. The system then automatically calculated the treatment schedule according to the MDT recommendation and pushed notifications to patients periodically. After completing each cycle of therapy, patients were required to confirm the treatment-related record on the app. During the entire course of treatment, patients communicated in real time with specialist nurses via this app.

Figure 2. The operation flow of the smartphone-based app. MDT: multidisciplinary treatment.



Measurement of Nonadherence

Treatment adherence was assessed by physicians via outpatient procedures, by follow-up specialists, or by nurses via phone calls. The routine follow-up interval was as follows: (1) every 3 months within the first 2 years after surgery, (2) every 6 months between 3 and 5 years, and (3) once a year after 5 years.

Treatment adherence was evaluated in all patients according to whether their actual treatment was the same as the results of the MDT discussion. Nonadherence was defined as follows: (1) patients refused to receive the prescribed regimen recommended by the MDT, (2) patients received a different regimen with the MDT recommendation, or (3) patients failed to complete the full course of treatment. Adherence was investigated in each therapeutic modality, including chemotherapy, radiotherapy, anti-HER2 (human epidermal growth factor receptor-2) therapy, and endocrine therapy, and then overall compliance was calculated.

Statistical Analysis

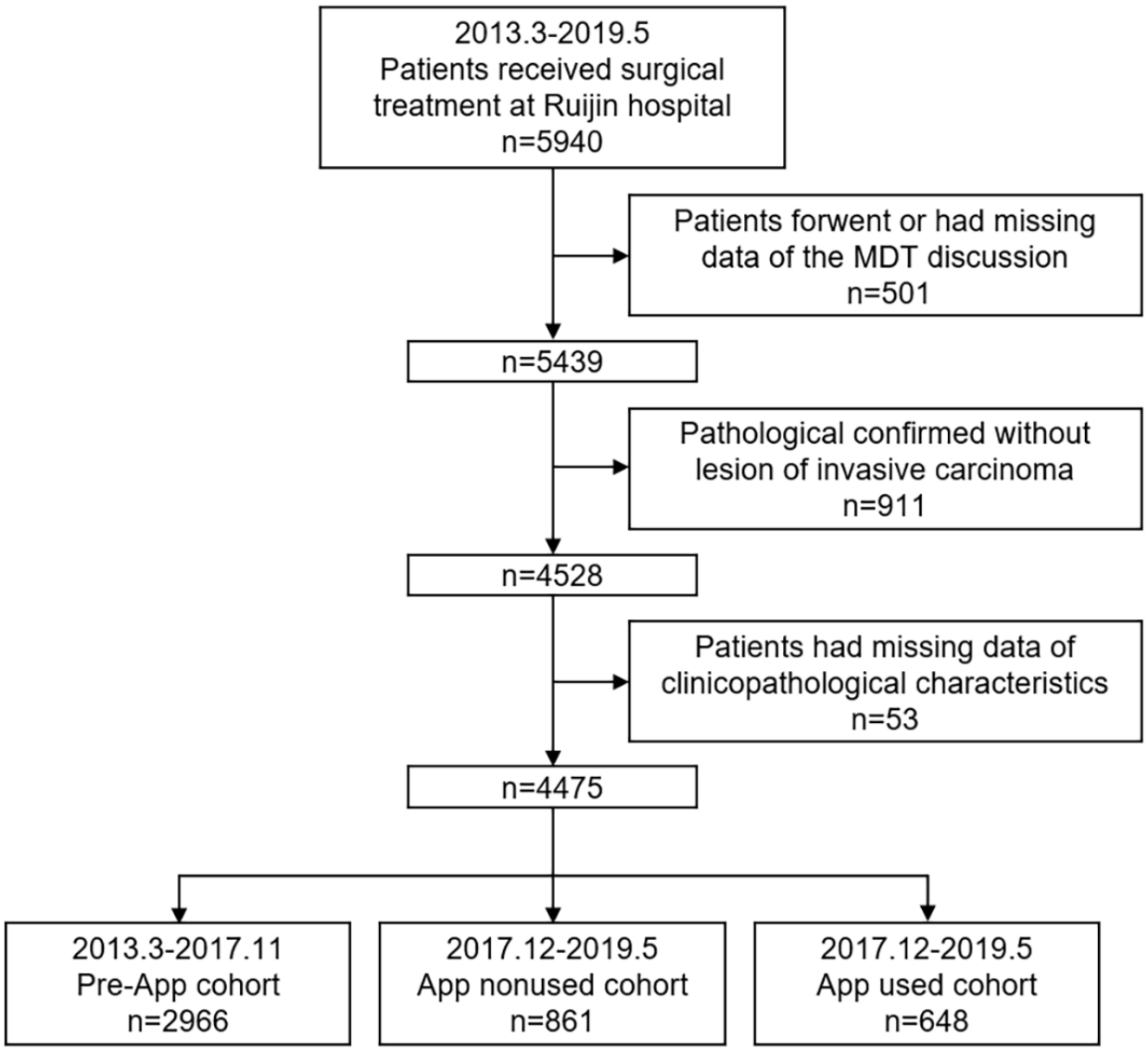
The chi-square test was used to evaluate the differences in patient characteristics among the three app-related groups and between the adherence and nonadherence groups, as well as to compare the difference in treatment adherence among the 3 app-related groups. A multivariate logistic regression model was used to identify the features associated with treatment compliance. Odds ratios (ORs) with 95% CIs were used to evaluate the influence of app use on patient adherence. Interactive analyses were performed to illustrate the effect of app use on compliance among patients with different clinicopathological features. Statistical analyses were performed using IBM SPSS version 25 and GraphPad Prism 8 (GraphPad Prism, Inc).

Results

Baseline Characteristics

A total of 5940 patients were reviewed, and 4475 were included in the study (Figure 3). There were 66.28% (2966/4475), 19.24% (861/4475), and 14.48% (648/4475) of patients in the Pre-App, App nonused, and App used cohorts, respectively.

Figure 3. Flowchart. MDT: multidisciplinary treatment.



Among 4475 patients, 1662 (37.14%) were aged 50 years or older, 2304 (51.49%) were aged between 50 and 70 years, and 11.73% (509/4475) of patients were older than 70 years. A total of 64.31% (2878/4475) of patients underwent mastectomy for the breast, and 66.2% (2940/4475) of patients underwent sentinel lymph node biopsy for the axillary region. Regarding the molecular subtype, there were 20.63% (923/4475), 42.53% (1903/4475), 12.97% (540/4475), 11.17% (500/4475), and 13.61% (609/4475) of tumors classified as luminal-A like, luminal-B like (HER2-negative), luminal-B like (HER2-positive), HER2 positive, and triple negative, respectively (Table 1). When using the American Joint Committee on Cancer staging system to classify the disease, 44.71% (1954/4475) were stage I, 42.08% (1839/4475) were stage II, and 13.2% (577/4475) were stage III. Altogether, 84.46% (3780/4475) of patients received the treatment regimen assigned by MDT, which was defined as adherence.

Table 1. Baseline characteristics of patients who have received surgery and multidisciplinary treatment discussion (N=4475).

Characteristics	Values, n (%)				*P* value
	Total (n=4475)	Pre-App cohort (n=2966)	App nonused cohort (n=861)	App used cohort (n=648)	
Age (years)					<.001
≤50	1662 (37.1)	1100 (37.1)	293 (34)	269 (41.5)	
50-70	2304 (51.5)	1535 (51.8)	435 (50.5)	334 (51.5)	
>70	509 (11.4)	331 (11.2)	133 (15.4)	45 (6.9)	
Educational level[a]					.70
Middle school or lower	1673 (38.2)	1104 (38.1)	334 (39.2)	235 (37.1)	
High school or higher	2709 (61.8)	1792 (61.9)	518 (60.8)	399 (62.9)	
Marital status					.14
Married	4315 (96.4)	2857 (96.3)	825 (95.8)	633 (97.7)	
Others	160 (3.6)	109 (3.7)	36 (4.2)	15 (2.3)	
Menopausal status					.002
Pre	1728 (38.6)	1144 (38.6)	300 (34.8)	284 (43.8)	
Post	2747 (61.4)	1822 (61.4)	561 (65.2)	364 (56.2)	
Benign breast disease history					<.001
Yes	967 (21.6)	572 (19.3)	228 (26.5)	167 (25.8)	
No	3508 (78.4)	2394 (80.7)	633 (73.5)	481 (74.2)	
Malignant disease history					<.001
Yes	209 (4.7)	111 (3.7)	56 (6.5)	42 (6.5)	
No	4266 (95.3)	2855 (96.3)	805 (93.5)	606 (93.5)	
Family history of breast cancer					.002
Yes	340 (7.6)	197 (6.6)	86 (10)	57 (8.8)	
No	4135 (92.4)	2769 (93.4)	775 (90)	591 (91.2)	
Comorbidity					.01
Yes	1766 (39.5)	1184 (39.9)	359 (41.7)	223 (34.4)	
No	2709 (60.5)	1782 (60.1)	502 (58.3)	425 (65.6)	
Breast surgery					<.001
Breast conserving	1597 (35.7)	1004 (33.9)	357 (41.5)	236 (36.4)	
Mastectomy	2878 (64.3)	1962 (66.1)	504 (58.5)	412 (63.6)	
Axillary surgery[b]					<.001
SLNB[c]	2940 (66.2)	1905 (64.6)	626 (74)	409 (63.2)	
ALND[d]	1501 (33.8)	1043 (35.4)	220 (26)	238 (36.8)	
Tumor size[e]					.40
≤2 cm	2542 (58)	1669 (58)	511 (59.3)	362 (55.9)	
>2 cm	1843 (42)	1207 (42)	350 (40.7)	286 (44.1)	
Lymph node status[f]					<.001
Negative	2905 (65.4)	1943 (65.9)	605 (71.5)	357 (55.3)	
Positive	1535 (34.6)	1005 (34.1)	241 (28.5)	289 (44.7)	
Pathological subtype					<.001
IDC[g]	3848 (86)	2578 (86.9)	690 (80.1)	580 (89.5)	
Non-IDC	627 (14)	388 (13.1)	171 (19.9)	68 (10.5)	

Characteristics	Values, n (%)				*P* value
	Total (n=4475)	Pre-App cohort (n=2966)	App nonused cohort (n=861)	App used cohort (n=648)	
Tumor grade					<.001
I or II	2253 (50.3)	1437 (48.4)	469 (54.5)	347 (53.5)	
III	1574 (35.2)	1137 (38.3)	212 (24.6)	225 (34.7)	
Unknown	648 (14.5)	392 (13.2)	180 (20.9)	76 (11.7)	
LVI[h]					<.001
No	3926 (87.7)	2689 (90.7)	722 (83.9)	515 (79.5)	
Yes	549 (12.3)	277 (9.3)	139 (16.1)	133 (20.5)	
Estrogen receptor status					.01
Negative	160 (24.7)	780 (26.3)	184 (21.4)	160 (24.7)	
Positive	3351 (74.9)	2186 (73.7)	677 (78.6)	488 (75.3)	
Progesterone receptor status					<.001
Negative	1629 (36.4)	1153 (38.9)	254 (29.5)	222 (34.3)	
Positive	2846 (63.6)	1813 (61.1)	607 (70.5)	426 (65.7)	
HER2[i] status					.001
Negative	3435 (76.8)	2280 (76.9)	689 (80)	466 (71.9)	
Positive	1040 (23.2)	686 (23.1)	172 (20)	182 (28.1)	
Ki-67					<.001
<14%	1400 (31.3)	1029 (34.7)	245 (28.5)	126 (19.4)	
≥14%	3075 (68.7)	1937 (65.3)	616 (71.5)	522 (80.6)	
Molecular subtype					<.001
Luminal-A like	923 (20.6)	657 (22.2)	184 (21.4)	82 (12.7)	
Luminal-B like (HER2-negative)	1903 (42.5)	1206 (40.7)	399 (46.3)	298 (46)	
Luminal-B like (HER2-positive)	540 (12.1)	331 (11.2)	97 (11.3)	112 (10.8)	
HER2 positive	500 (11.2)	355 (12)	75 (8.7)	70 (10.8)	
Triple negative	609 (13.6)	417 (14.1)	106 (12.3)	86 (13.3)	
TNM[j] stage[k]					.003
Stage I	1954 (44.7)	1289 (44.8)	415 (49.1)	250 (38.7)	
Stage II	1839 (42.1)	1213 (42.1)	329 (38.9)	297 (46)	
Stage III	577 (13.2)	376 (13.1)	102 (12.1)	99 (15.3)	

[a]A total of 993 patients had missing data of educational level.

[b]A total of 34 patients had missing data of axillary surgery procedure.

[c]SLNB: sentinel lymph node biopsy.

[d]ALND: axillary lymph node dissection.

[e]A total of 90 patients had missing data of tumor size.

[f]A total of 35 patients had missing data of lymph node status.

[g]IDC: invasive ductal carcinoma.

[h]LVI: lymphovascular invasion.

[i]HER2: human epidermal growth factor receptor 2.

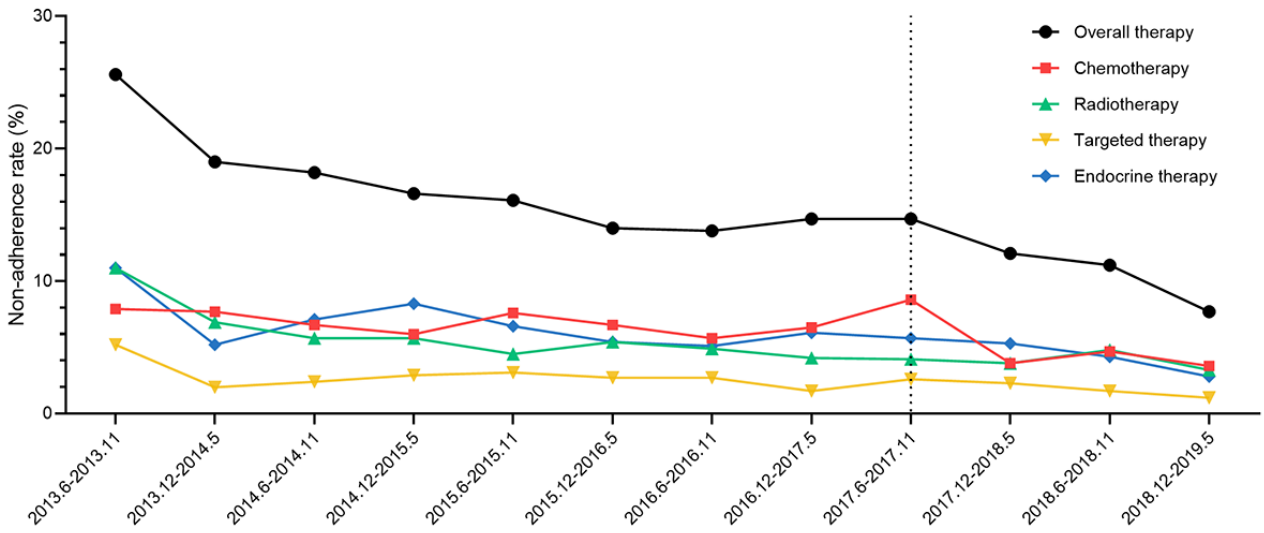
[j]TNM: tumor, lymph node, and metastasis.

[k]A total of 105 patients had missing data of tumor, lymph node, and metastasis stage.

Factors Associated With Treatment Adherence

The rate of nonadherence decreased by year, which was approximately 27.4% in 2013 and 8.8% in 2019 (Figure 4).

Figure 4. Trend of noncompliance rate to multidisciplinary treatment recommendations by year. The dotted line indicates launch of the smartphone-based app.



In the univariate analysis, diagnosis year, demographic characteristics, medical history, surgical type, as well as the staging and characteristics of tumors were significantly associated with adherence to MDT. Moreover, regarding smartphone-based app use, the nonadherence rate was 17.6% in the Pre-App cohort, 13.24% in the App nonused cohort, and 9.1% in the App used cohort (P<.001; Tables 2 and 3; Table S1 in Multimedia Appendix 1).

Table 2. Factors associated with treatment adherence in patients.

Characteristics	Value, n (%)		*P* value
	Adherence (n=3780)	Nonadherence (n=695)	
Year			<.001
2013.3-2014.5	410 (76.9)	123 (23.1)	
2014.6-2015.5	453 (81.2)	105 (18.8)	
2015.6-2016.5	551 (83.6)	108 (16.4)	
2016.6-2017.5	677 (85.2)	118 (14.8)	
2017.6-2018.5	799 (85.5)	136 (14.5)	
2018.6-2019.5	890 (89.4)	105 (10.6)	
Age (years)			<.001
≤50	1444 (86.9)	218 (13.1)	
50-70	1972 (95.6)	332 (14.4)	
>70	364 (71.5)	145 (28.5)	
Educational level			<.001
Middle school or lower	1386 (81.8)	305 (18.2)	
High school or higher	2332 (86.1)	377 (13.9)	
Marital status			.82
Married	3646 (86.7)	669 (13.3)	
Others	134 (83.1)	26 (16.9)	
Menopausal status			.001
Pre	1498 (39.6)	230 (33.1)	
Post	2282 (60.4)	465 (66.9)	
Benign breast disease history			.06
Yes	836 (86.5)	131 (13.5)	
No	2944 (83.9)	564 (16.1)	
Malignant disease history			.08
Yes	167 (79.9)	42 (20.1)	
No	3613 (84.7)	653 (15.3)	
Family history of breast cancer			.008
Yes	304 (89.4)	36 (10.6)	
No	3476 (84.1)	659 (15.9)	
Comorbidity			.08
Yes	1471 (83.3)	295 (16.7)	
No	2309 (85.2)	400 (14.8)	
Breast surgery			.32
Breast conserving	1361 (85.2)	236 (14.8)	
Mastectomy	2419 (84.1)	459 (15.9)	
Axillary surgery			.001
SLNB[a]	2526 (67.2)	414 (60.7)	
ALND[b]	1233 (32.8)	268 (39.3)	
Tumor size (cm)			.001
≤2	2187 (86)	355 (14)	

Characteristics	Value, n (%)		*P* value
	Adherence (n=3780)	Nonadherence (n=695)	
>2	1516 (82.3)	327 (47.9)	
Lymph node status			<.001
Negative	2508 (86.3)	397 (13.7)	
Positive	1250 (81.4)	285 (18.6)	
Pathological subtype			.007
IDC[c]	3228 (83.9)	620 (16.1)	
Non-IDC	552 (88)	75 (12)	
Tumor grade			<.001
I or II	1938 (86)	315 (14)	
III	1279 (81.3)	295 (18.7)	
Unknown	563 (86.9)	85 (13.1)	
LVI[d]			.35
No	3324 (84.7)	602 (15.3)	
Yes	456 (83.1)	93 (16.9)	
ER[e] status			.13
Negative	933 (83)	191 (17.9)	
Positive	2847 (85)	504 (15)	
PR[f] status			.002
Negative	1339 (82.2)	290 (17.8)	
Positive	2441 (85.8)	405 (14.2)	
HER2[g] status			<.001
Negative	2941 (85.6)	494 (14.4)	
Positive	839 (80.7)	201 (19.3)	
Ki-67			.001
<14%	1219 (87.1)	181 (12.9)	
≥14%	2561 (83.3)	514 (16.7)	
Molecular subtype			<.001
Luminal-A like	831 (90)	92 (10)	
Luminal-B like (HER2-negative)	1607 (84.4)	296 (15.6)	
Luminal-B like (HER2-positive)	421 (78)	119 (22)	
HER2 positive	418 (83.6)	82 (16.4)	
Triple negative	503 (82.6)	106 (17.4)	
TNM[h] stage			<.001
Stage I	1713 (87.7)	241 (12.3)	
Stage II	1512 (82.7)	318 (17.3)	
Stage III	459 (79.5)	118 (20.5)	
Group according to app use			<.001
Pre-App cohort	2444 (82.4)	522 (17.6)	
App nonused cohort	747 (86.8)	114 (13.2)	

Characteristics	Value, n (%)		*P* value
	Adherence (n=3780)	Nonadherence (n=695)	
App used cohort	589 (90.9)	59 (9.1)	

[a]SLNB: sentinel lymph node biopsy.

[b]ALND: axillary lymph node dissection.

[c]IDC: invasive ductal carcinoma.

[d]LVI: lymphovascular invasion.

[e]ER: estrogen receptor.

[f]PR: progesterone receptor.

[g]HER2: human epidermal growth factor receptor 2.

[h]TNM: tumor, lymph node, and metastasis.

Table 3. Patients' adherence according to treatment.

Treatment	Value, n (%)			*P* value
	Pre-App cohort (n=2966)	App nonused cohort (n=861)	App used cohort (n=648)	
Chemotherapy				<.001
Adherent	2732 (92.1)	808 (93.8)	625 (96.5)	
Nonadherent	234 (7.9)	53 (6.2)	23 (3.5)	
Radiotherapy				.001
Adherent	2789 (94)	809 (94)	633 (97.7)	
Nonadherent	177 (6)	52 (6)	15 (2.3)	
Targeted therapy[a]				.01
Adherent	594 (86.6)	158 (91.9)	170 (93.4)	
Nonadherent	92 (13.4)	14 (8.1)	12 (6.6)	
Endocrine therapy[b]				<.001
Adherent	2001 (91.2)	637 (93.5)	476 (96.4)	
Nonadherent	194 (8.8)	44 (6.5)	18 (3.6)	
Overall therapy				<.001
Adherent	2444 (82.4)	747 (86.8)	589 (90.9)	
Nonadherent	522 (17.6)	114 (13.2)	59 (9.1)	

[a]In 1028 patients with HER2-positive breast cancer.

[b]In 3337 patients with hormone receptor–positive breast cancer.

Multivariate analysis demonstrated that diagnosis year, age, educational level, axillary surgery, lymph node status, HER2 status, and molecular subtype were significantly associated with patients' adherence to the MDT recommendation. Notably, the use of the smartphone-based app was also an independent factor associated with patient compliance (*P*=.02). Compared with the patients in the Pre-App cohort, patients in the App used cohort were less likely to violate the treatment plan (OR 0.60, 95% CI 0.43-0.87; *P*=.007). However, there was no significant difference in adherence rates between the App nonused cohort and the Pre-App cohort (OR 0.96, 95% CI 0.70-1.29; *P*=.77; Table 4).

Table 4. Multivariate analysis of factors associated with nonadherence.

Characteristics	Odds ratio (95% CI)	*P* value
Diagnosis year (per year)	0.88 (0.81-0.94)	<.001
Age (years)		<.001
50-70 vs ≤50	1.08 (0.88-1.31)	.45
>70 vs ≤50	2.85 (2.19-3.70)	<.001
Educational level (high school or higher vs middle school or lower)	0.76 (0.64-0.91)	.002
Menopausal status (post vs pre)	1.01 (0.72-1.41)	.96
Family history of breast cancer (yes vs no)	0.70 (0.48-1.00)	.05
Axillary surgery (ALND[a] vs SLNB[b])	0.72 (0.55-0.96)	.03
Tumor size (>2 cm vs ≤2 cm)	0.83 (0.62-1.09)	.18
Lymph node status (positive vs negative)	1.40 (1.04-1.88)	.03
Pathological subtype (Non-IDC[c] vs IDC)	0.77 (0.58-1.02)	.07
Tumor grade		.15
III versus I or II	1.19 (0.97-1.45)	.10
Unknown versus I or II	1.36 (0.84-2.20)	.21
PR[d] status (positive vs negative)	0.86 (0.66-1.12)	.28
HER2[e] status (positive vs negative)	2.04 (1.34-3.10)	.001
Ki-67 index (≥14% vs <14%)	0.96 (0.73-1.26)	.75
Molecular subtype		<.001
Luminal-B like (HER2 negative) versus Luminal-A like	1.83 (1.29-2.59)	.001
Luminal-B like (HER2 positive) versus Luminal-A like	1.60 (1.16-2.21)	.005
HER2 positive versus Luminal-A like	N/A[f]	N/A
Triple negative versus Luminal-A like	2.11 (1.41-3.14)	<.001
TNM[g] stage		.02
Stage II versus stage I	1.35 (1.08-1.69)	.008
Stage III versus stage I	1.57 (1.09-2.25)	.01
Group according to app use		.02
App nonused cohort versus Pre-App cohort	0.96 (0.70-1.29)	.77
App used cohort versus Pre-App cohort	0.61 (0.43-0.87)	.007

[a]ALND: axillary lymph node dissection.

[b]SLNB: sentinel lymph node biopsy.

[c]IDC: invasive ductal carcinoma.

[d]PR: progesterone receptor.

[e]HER2: human epidermal growth factor receptor 2.

[f]N/A: not applicable.

[g]TNM: tumor, lymph node, and metastasis.

Patient Adherence According to Treatment

Adjuvant Chemotherapy

Among 4475 patients, 310 (6.93%) patients did not adhere to the adjuvant chemotherapy decision, and the nonadherence rates were 7.9%, 6.2%, and 3.5% in the Pre-App cohort, App nonused, and App used cohorts, respectively (*P*<.001; Table 3; Table S1 in Multimedia Appendix 1). In the multivariate analysis, older age, luminal-B–like tumors, HER2-positive breast cancer, and triple-negative breast cancer were significantly associated with a lower compliance rate. Pathologically diagnosed noninvasive ductal carcinoma was associated with good compliance (*P*=.003). Moreover, compared with the patients in the Pre-App cohort, patients in the App used cohort (OR 0.41, 95% CI 0.27-0.65; *P*<.001) but not in the App nonused cohort (OR 0.74, 95% CI 0.54-1.02; *P*=.07) were less likely to not follow

MDT-recommended chemotherapy (Table S2 in Multimedia Appendix 1).

Adjuvant Radiotherapy

There were 5.45% (244/4475) patients who did not receive adjuvant radiotherapy according to the MDT discussion, and the discordance rates were 6%, 6%, and 2.3% in the 3 groups, respectively (*P*=.001; Table 3; Table S3 in Multimedia Appendix 1). Multivariate analysis demonstrated that later diagnosis year, age, educational level, Ki-67 index, and lymph node status were significantly related to patient adherence (Table S4 in Multimedia Appendix 1). When compared with the patients in the Pre-App cohort, patients in the App used cohort were less likely to not follow the MDT recommendation (OR 0.49, 95% CI 0.25-0.96; *P*=.004). However, there was no significant difference between the App nonused cohort and the Pre-App cohort (OR 1.38, 95% CI 0.86-2.22; *P*=.18) in terms of adjuvant radiotherapy.

Adjuvant Target Therapy

Among 1028 patients with HER2-positive breast cancer, 118 (11.48%) patients violated the anti-HER2 treatment strategy. The nonadherence rates were 13.4%, 8.1%, and 6.6% in the Pre-App, App nonused, and App used cohorts, respectively (*P*=.02; Table 3; Table S5 in Multimedia Appendix 1). In multivariate analysis, only diagnosis year, age >70 years, and educational level were independent factors of treatment adherence (Table S6 in Multimedia Appendix 1). However, app use was not independently associated with MDT adherence to adjuvant targeted therapy (*P*=.76).

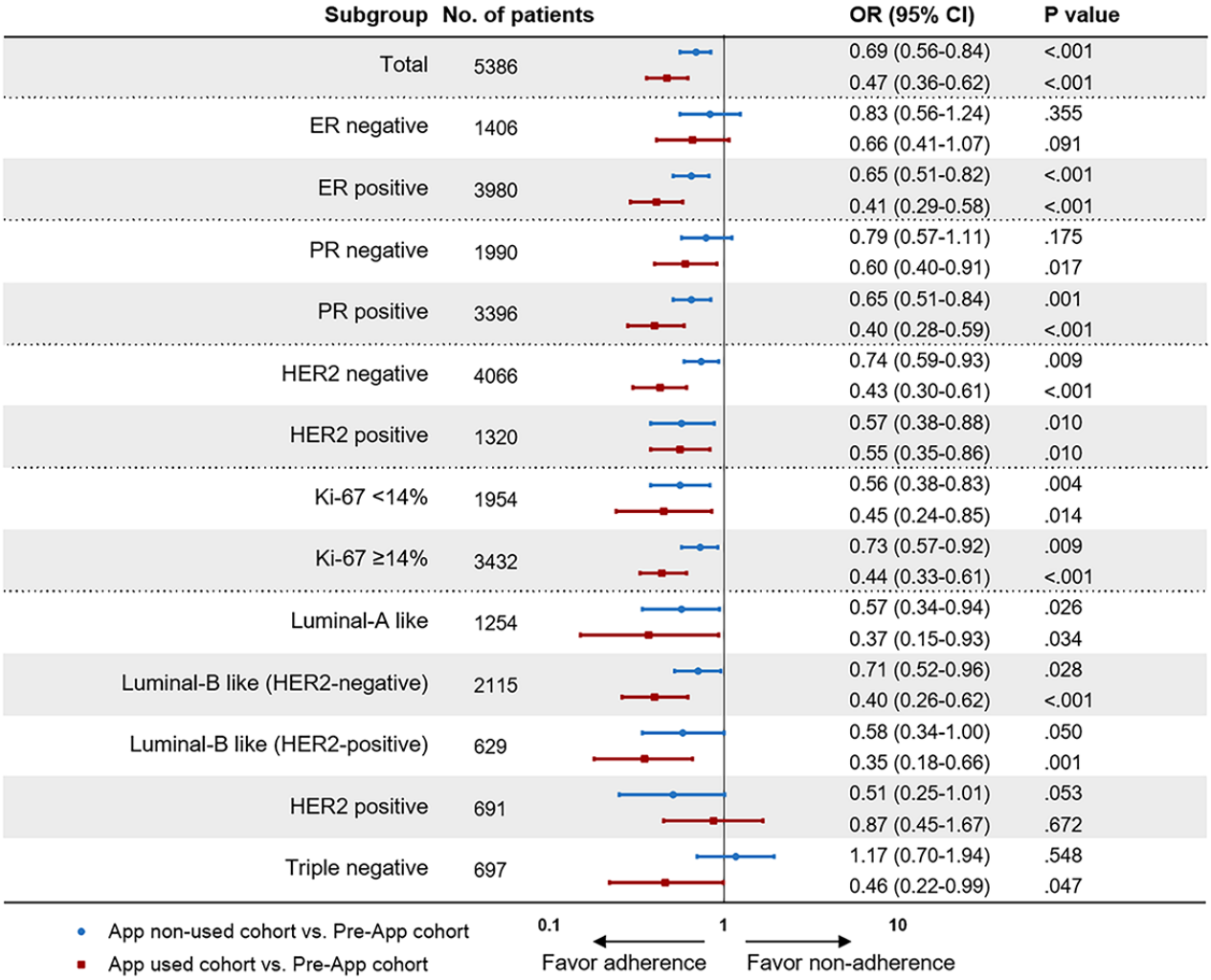
Adjuvant Endocrine Therapy

A total of 3337 patients received a diagnosis of hormone receptor–positive breast cancer, among whom 256 (7.67%) did not follow the doctors' recommendations for endocrine therapy within the follow-up period. The noncompliance rates were 8.8%, 6.5%, and 3.6% in the 3 groups (*P*<.001; Table 3; Table S7 in Multimedia Appendix 1). Multivariate analysis revealed that patients with early diagnosis years, aged >70 years, and with luminal-B–like (HER2-positive) breast cancer were less likely to adhere to the endocrine therapy recommendation (Table S8 in Multimedia Appendix 1). The use of the app was not significantly associated with treatment compliance (*P*=.54).

Subgroup Analysis

Subgroup analyses of MDT-recommended treatment compliance according to clinicopathological parameters were also performed (Figure 5). Age (interaction *P*<.001), menopausal status (interaction *P*<.001), and comorbidity (interaction *P*=.007) were significantly associated with app use and MDT compliance. In patients older than 70 years, those who did not use the app were more likely to be nonadherent (OR 2.01, 95% CI 1.31-3.09; *P*<.001). Meanwhile, use of apps was associated with better adherence in postmenopausal patients (OR 0.54, 95% CI 0.38-0.77; *P*<.001) and patients with comorbidities (OR 0.46, 95% CI 0.28-0.74; *P*<.001). However, there were no significant interactions between the group setting and cancer-related characteristics, including the molecular subtype or tumor, lymph node, and metastasis stage.

Figure 5. Effect of the app use on multidisciplinary treatment compliance in the entire population and stratified according to clinicopathological features. Odds ratios (OR) and 95% CIs for multidisciplinary treatment compliance (an OR<1 indicates a higher compliance rate) are shown. The compliance rate of patients in the Pre-App cohort was used as a reference. CIs have not been adjusted for multiple comparisons. Significant interactions were observed between patient groups and age, menopausal status, as well as comorbidities. ER: estrogen receptor; HER2: human epidermal growth factor receptor-2; OR: odds ratio; PR: progesterone receptor.



Discussion

Principal Findings

In this study, we included 4475 patients with early-stage breast cancer who had undergone an MDT discussion after surgery and found that sociodemographic factors together with clinicopathological factors were associated with MDT compliance in terms of adjuvant treatments. More importantly, the novel smartphone-based app designed for full-course management of patients with breast cancer could improve patients' adherence to MDT recommendations for adjuvant treatment, which will help us better manage breast cancer adjuvant treatment after MDT and improve disease outcomes for these patients.

MDT in Breast Cancer

Currently, MDT has gradually become an integral part of standardized treatment modalities for breast cancer worldwide. It could be conducive to evidence-based decisions for clinicians and provide a convenient medical treatment process for patients [15,16]. Our previous studies demonstrated that MDT discussion could also lower the relative risk of relapse by 16% and the risk of death by 11% in patients with breast cancer [7]. However, a major difficulty after the MDT meeting and during the treatment process was maintaining patient compliance. Data from the Pennsylvania and Florida cancer registries demonstrated that nearly 30% of the patients did not pursue the recommended treatment. The reported nonadherence rate during the 5 years of adjuvant endocrine therapy ranged from 10% to 40% [17-20]. Our study reported a nonadherence rate of 14.6%, and real-world nonadherence to adjuvant therapy may still be underestimated because of loss of patient follow-up and self-reporting methods [8].

Treatment Noncompliance in Breast Cancer

Influence Factors of Noncompliance

Noncompliance is a complex and multifaceted phenomenon that is influenced by demographic, socioeconomic, psychological, and clinical-related factors [21-24]. Several studies including our study have documented that age-specific factors were associated with a lower compliance rate [17,25,26]. This was because weakness and fragility due to old age may

influence tolerance to medical interventions. Regarding disease-related factors, receiving axillary lymph node dissection (ALND), HER2-positive, unfavorable molecular subtype, and advanced tumor, lymph node, and metastasis stage were associated with lower compliance. These patients usually require more complex and aggressive regimens, and their confusion about the therapeutic procedure as well as anxiety about adverse reactions may lead to treatment absence [27]. Moreover, the literature reported that health insurance and employment status were correlated with intravenous chemotherapy adherence [28,29]. Our study demonstrated that educational level was an influencing factor of compliance. These results indicate that economic level was positively related to treatment adherence. Jacobs et al [30] found that cancer-related symptom severity, but not sociodemographic or psychosocial constructs, could influence patients' compliance with oral chemotherapy. The size of enrollment and the delivery method of chemotherapy may have caused differences between the studies. Furthermore, the modifiability of psychological status renders it a key to improving patient compliance [31], but we failed to include this factor in the study because of the difficulty in emotional evaluation during regular follow-up.

Interventions to Improve Patients' Compliance

Withdrawal from treatment was associated with increased recurrence and impaired survival [32]. Our previous study showed that noncompliant patients had 1.8 times the risk of disease relapse and 2.5 times all-cause mortality than those who received the planned treatment [7]. Thus, tailoring enhanced interventions to improve patient adherence is of great importance. Liu et al [33] showed a large difference in patient adherence (59% vs 94%) between those receiving provider-patient communications and those not receiving provider-patient communication. In SWOG S1105, unidirectional text messaging failed to reduce the early discontinuation of adjuvant aromatase inhibitor therapy in women with early-stage breast cancer, suggesting that long-term adherence may call for personalized behavioral interventions and sustained management [34]. A meta-analysis by Lin et al [35] analyzed the influence of psychosocial factors on oral anticancer medication adherence in patients with breast cancer and indicated that using patient-centered interventions and building sustainable relationships may contribute to improved compliance.

Mobile Health Apps

Purpose of Mobile Health Apps

During the treatment process, patients may face complex procedures, unfamiliar treatment-related effects, and numerous changes in psychology and lifestyle. Thus, they need a platform to interact with medical workers, provide feedback on treatment-related problems, and gain informational support. With the popularization of mobile communication, a number of mobile health apps designed for prevention and early diagnosis of disease, management of disease, and survival support and enhancement, greatly facilitate breast cancer health care. As acknowledged, the treatment of breast cancer involves more than surgery; thus, our app was designed for the management of postoperative treatment. Until 2018, approximately 600 apps were designed for breast cancer prevention and management [36]. Those apps designed for providing disease-related information, managing disease, raising awareness, and preventing disease, and the majority of the target population of these apps were patients with breast cancer.

Features and Innovations of Our App

We published an app that engages physicians and oncology nurses in the long-term management of patients after surgery. Our app relies on the most popular social networking platform in China-WeChat, and patients do not need to download and install the software, which ensures convenience and accessibility. Meanwhile, the app was connected to our MDT decision support system [37], and the treatment information of patients can be directly imported into the full-course management app system without manual operation, which could improve efficacy and reduce errors. More than half of the previous apps were developed by nonmedical professors, which may leave concerns regarding the accuracy and validity of the information [36,38]. Our app was designed by health care and computer science professionals, which ensured the credibility of the information and reliability of operation. Several studies have reported that mobile intervention only had modest effects on patients' actual behavior, especially using the most common type of intervention-text messaging [34,39]. We found that after adjusting for the diagnosis year, patients using the smartphone-based full-course management app had a nonadherence rate of only 9.1%, lower than 17.6% in the Pre-App cohort, and 13.24% in the App nonused cohort. The effectiveness of our app may be because, in addition to regular treatment reminders, it supports interactive communication between patients and specialist nurses, which enables the patients to resolve the problem during treatment in a timely manner. The importance of two-way communication has also been demonstrated by Hwang et al [40], who designed an e-monitoring app allowing patients to upload the picture of wounds and consulting doctors on web, and it can decrease the number of clinical visits.

Patient Engagement

Patient engagement was expected to be related to the effectiveness of the digital intervention. Perski et al [41] found that the target behavior, as well as the mechanism of action, may influence engagement. At present, there is a lack of a uniform standard definition of engagement. In this study, we defined the use by logging detected from the backstage system for at least one time, as well as the reasons for patients not to use this app, including not owning a smartphone, older age, being treated in the local hospitals, and refusing to follow the MDT recommendation. Stubbins et al [42] emphasized the importance of providing real-time feedback via the app to improve patients' engagement and adherence. Our app has a feedback function, tending to investigate and improve user experience. A previous review summarized studies using interventions to improve patients' endocrine therapy drug uptake, which should last for years, and all published studies failed to prevent medication discontinuation [43]. Our app was also not effective for targeted or endocrine therapy. Indeed, with the prolongation of treatment, user retention for

smartphone-based apps is another great challenge [44], and further efforts are still needed to minimize user fatigue to enhance patients' long-term engagement.

Economic Factors

There has been growing interest in economic evidence. Although our app is funded by a government project and is completely free for the public to use, the economic status could still potentially influence the use of mobile communication. However, there is a lack of uniform standards to evaluate the economic outcomes, and cost-effectiveness analysis is still insufficient [45,46].

Limitations

In this study, we developed a novel and effective smartphone-based app that can decrease patients' prescription-deviating behaviors during treatment. One strength of this study is the large sample size of the included patients. There are several other potential limitations of this study. First, as a retrospective study, there may be a selection bias among patients. Thus, a prospectively designed study is warranted to validate the influence of this app on patient adherence and long-term survival. Second, there is still a lack of standard methods for evaluating adherence. Further efforts are needed to achieve consensus in the assessment of compliance, including time, frequency, and detection method. Third, the economic status of patients was difficult to assess in this study, which may have greatly influenced the results. Tailoring appropriate criteria to evaluate the financial situation of patients should be considered in the future. Last but not least, the follow-up time is still too short for some of the patients right now because they have not completed adjuvant endocrine therapy for at least 5 years, and the results of continued follow-up are worth expecting. There were also some points that deserve consideration in the development of health care–related apps. First, standardized quality measures for medical apps are lacking, and corresponding scales or rules should be developed. The apps should be comprehensively measured and rated by medical workers, software engineers, and users to improve the quality of apps. Second, the economic impact and long-term effectiveness of apps should be constantly tracked, and novel features warrant consideration to meet patients' needs and ensure user engagement.

Conclusions

We included a large number of patients with early-stage breast cancer within the MDT discussion and found that treatment adherence was independently associated with smartphone-based app use, which can serve as a useful intervention to improve patient compliance with MDT. Prospective studies are needed to validate the effectiveness of smartphone-based full-course management apps in improving patient outcomes to integrate this app into routine MDT clinical practice for breast cancer.

Acknowledgments

The authors would like to thank Yidong Du for her significant contribution to the data input, management, and follow-up. This work was funded by financial support from the National Natural Science Foundation of China (81772797, 82072937) and Shanghai Municipal Education Commission–Gaofeng Clinical Medicine Grant Support (20172007).

Authors' Contributions

JY and XC contributed to the conception and design of the study; JW and OH organized the database; JY performed the statistical analysis and wrote the first draft of the manuscript; KS and XC contributed to manuscript revision and funding acquisition, and they read and approved the submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 72 KB - jmir_v23i9e27576_app1.docx]

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021 May;71(3):209-249 [FREE Full text] [doi: 10.3322/caac.21660] [Medline: 33538338]
2. Waks AG, Winer EP. Breast cancer treatment: a review. J Am Med Assoc 2019 Jan 22;321(3):288-300. [doi: 10.1001/jama.2018.19323] [Medline: 30667505]
3. Saini KS, Taylor C, Ramirez A, Palmieri C, Gunnarsson U, Schmoll HJ, et al. Role of the multidisciplinary team in breast cancer management: results from a large international survey involving 39 countries. Ann Oncol 2012 Apr;23(4):853-859 [FREE Full text] [doi: 10.1093/annonc/mdr352] [Medline: 21821551]
4. Harbeck N, Gnant M. Breast cancer. Lancet 2017 Mar 18;389(10074):1134-1150. [doi: 10.1016/S0140-6736(16)31891-8] [Medline: 27865536]

5. Dean LT, Moss SL, McCarthy AM, Armstrong K. Healthcare system distrust, physician trust, and patient discordance with adjuvant breast cancer treatment recommendations. Cancer Epidemiol Biomarkers Prev 2017 Dec;26(12):1745-1752 [FREE Full text] [doi: 10.1158/1055-9965.EPI-17-0479] [Medline: 28971987]
6. Murphy CC, Bartholomew LK, Carpentier MY, Bluethmann SM, Vernon SW. Adherence to adjuvant hormonal therapy among breast cancer survivors in clinical practice: a systematic review. Breast Cancer Res Treat 2012 Jul;134(2):459-478 [FREE Full text] [doi: 10.1007/s10549-012-2114-5] [Medline: 22689091]
7. Yang X, Huang J, Zhu X, Shen K, Zhu J, Chen X. Compliance with multidisciplinary team recommendations and disease outcomes in early breast cancer patients: an analysis of 4501 consecutive patients. Breast 2020 Aug;52:135-145 [FREE Full text] [doi: 10.1016/j.breast.2020.05.008] [Medline: 32512360]
8. Pistilli B, Paci A, Ferreira AR, Di Meglio A, Poinsignon V, Bardet A, et al. Serum detection of nonadherence to adjuvant tamoxifen and breast cancer recurrence risk. J Clin Oncol 2020 Aug 20;38(24):2762-2772. [doi: 10.1200/JCO.19.01758] [Medline: 32568632]
9. Ramsey WA, Heidelberg RE, Gilbert AM, Heneghan MB, Badawy SM, Alberts NM. eHealth and mHealth interventions in pediatric cancer: a systematic review of interventions across the cancer continuum. Psychooncology 2020 Jan;29(1):17-37. [doi: 10.1002/pon.5280] [Medline: 31692183]
10. Naslund JA, Grande SW, Aschbrenner KA, Elwyn G. Naturally occurring peer support through social media: the experiences of individuals with severe mental illness using YouTube. PLoS One 2014 Oct 15;9(10):e110171 [FREE Full text] [doi: 10.1371/journal.pone.0110171] [Medline: 25333470]
11. Jibb LA, Stevens BJ, Nathan PC, Seto E, Cafazzo JA, Johnston DL, et al. Implementation and preliminary effectiveness of a real-time pain management smartphone app for adolescents with cancer: a multicenter pilot clinical study. Pediatr Blood Cancer 2017 Oct;64(10):e26554. [doi: 10.1002/pbc.26554] [Medline: 28423223]
12. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
13. Badawy SM, Barrera L, Sinno MG, Kaviany S, O'Dwyer LC, Kuhns LM. Text messaging and mobile phone apps as interventions to improve adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2017 May 15;5(5):e66 [FREE Full text] [doi: 10.2196/mhealth.7798] [Medline: 28506955]
14. Radovic A, Badawy SM. Technology use for adolescent health and wellness. Pediatrics 2020 May;145(Suppl 2):186-194 [FREE Full text] [doi: 10.1542/peds.2019-2056G] [Medline: 32358210]
15. Vinod SK, Sidhom MA, Delaney GP. Do multidisciplinary meetings follow guideline-based care? J Oncol Pract 2010 Nov;6(6):276-281 [FREE Full text] [doi: 10.1200/JOP.2010.000019] [Medline: 21358954]
16. Gabel M, Hilton NE, Nathanson SD. Multidisciplinary breast cancer clinics. Do they work? Cancer 1997 Jun 15;79(12):2380-2384. [Medline: 9191526]
17. Wulaningsih W, Garmo H, Ahlgren J, Holmberg L, Folkvaljon Y, Wigertz A, et al. Determinants of non-adherence to adjuvant endocrine treatment in women with breast cancer: the role of comorbidity. Breast Cancer Res Treat 2018 Nov;172(1):167-177 [FREE Full text] [doi: 10.1007/s10549-018-4890-z] [Medline: 30030708]
18. Kroenke CH, Hershman DL, Gomez SL, Adams SR, Eldridge EH, Kwan ML, et al. Personal and clinical social support and adherence to adjuvant endocrine therapy among hormone receptor-positive breast cancer patients in an integrated health care system. Breast Cancer Res Treat 2018 Aug;170(3):623-631 [FREE Full text] [doi: 10.1007/s10549-018-4774-2] [Medline: 29671113]
19. Brier MJ, Chambless DL, Gross R, Chen J, Mao JJ. Perceived barriers to treatment predict adherence to aromatase inhibitors among breast cancer survivors. Cancer 2017 Jan 01;123(1):169-176 [FREE Full text] [doi: 10.1002/cncr.30318] [Medline: 27570979]
20. Lundgren C, Lindman H, Rolander B, Ekholm M. Good adherence to adjuvant endocrine therapy in early breast cancer - a population-based study based on the Swedish Prescribed Drug Register. Acta Oncol 2018 Jul;57(7):935-940. [doi: 10.1080/0284186X.2018.1442932] [Medline: 29493327]
21. Lima MP, Machado WD, Irigaray TQ. Predictive factors of treatment adherence in cancer outpatients. Psychooncology 2018 Dec;27(12):2823-2828. [doi: 10.1002/pon.4897] [Medline: 30239056]
22. Puts MT, Tu HA, Tourangeau A, Howell D, Fitch M, Springall E, et al. Factors influencing adherence to cancer treatment in older adults with cancer: a systematic review. Ann Oncol 2014 Mar;25(3):564-577 [FREE Full text] [doi: 10.1093/annonc/mdt433] [Medline: 24285020]
23. Kondryn HJ, Edmondson CL, Hill J, Eden TO. Treatment non-adherence in teenage and young adult patients with cancer. Lancet Oncol 2011 Jan;12(1):100-108. [doi: 10.1016/S1470-2045(10)70069-3] [Medline: 20580606]
24. Ruddy K, Mayer E, Partridge A. Patient adherence and persistence with oral anticancer treatment. CA Cancer J Clin 2009;59(1):56-66 [FREE Full text] [doi: 10.3322/caac.20004] [Medline: 19147869]
25. Rajan S, Foreman J, Wallis MG, Caldas C, Britton P. Multidisciplinary decisions in breast cancer: does the patient receive what the team has recommended? Br J Cancer 2013 Jun 25;108(12):2442-2447 [FREE Full text] [doi: 10.1038/bjc.2013.267] [Medline: 23736032]

26. Mislang AR, Wildes TM, Kanesvaran R, Baldini C, Holmes HM, Nightingale G, et al. Adherence to oral cancer therapy in older adults: the International Society of Geriatric Oncology (SIOG) taskforce recommendations. Cancer Treat Rev 2017 Jun;57:58-66. [doi: 10.1016/j.ctrv.2017.05.002] [Medline: 28550714]
27. Paranjpe R, John G, Trivedi M, Abughosh S. Identifying adherence barriers to oral endocrine therapy among breast cancer survivors. Breast Cancer Res Treat 2019 Apr;174(2):297-305. [doi: 10.1007/s10549-018-05073-z] [Medline: 30523459]
28. Wells JS, Strickland OL, Dalton JA, Freeman S. Adherence to intravenous chemotherapy in African American and white women with early-stage breast cancer. Cancer Nurs 2015;38(2):89-98 [FREE Full text] [doi: 10.1097/NCC.0000000000000139] [Medline: 24831041]
29. Gallups SF, Connolly MC, Bender CM, Rosenzweig MQ. Predictors of adherence and treatment delays among African American women recommended to receive breast cancer chemotherapy. Womens Health Issues 2018;28(6):553-558. [doi: 10.1016/j.whi.2018.08.001] [Medline: 30241793]
30. Jacobs JM, Ream ME, Pensak N, Nisotel LE, Fishbein JN, MacDonald JJ, et al. Patient experiences with oral chemotherapy: adherence, symptoms, and quality of life. J Natl Compr Canc Netw 2019 Mar 01;17(3):221-228 [FREE Full text] [doi: 10.6004/jnccn.2018.7098] [Medline: 30865917]
31. Wassermann J, Gelber SI, Rosenberg SM, Ruddy KJ, Tamimi RM, Schapira L, et al. Nonadherent behaviors among young women on adjuvant endocrine therapy for breast cancer. Cancer 2019 Sep 15;125(18):3266-3274 [FREE Full text] [doi: 10.1002/cncr.32192] [Medline: 31120571]
32. Font R, Espinas JA, Barnadas A, Izquierdo A, Galceran J, Saladie F, et al. Influence of adherence to adjuvant endocrine therapy on disease-free and overall survival: a population-based study in Catalonia, Spain. Breast Cancer Res Treat 2019 Jun;175(3):733-740. [doi: 10.1007/s10549-019-05201-3] [Medline: 30887164]
33. Liu Y, Malin JL, Diamant AL, Thind A, Maly RC. Adherence to adjuvant hormone therapy in low-income women with breast cancer: the role of provider-patient communication. Breast Cancer Res Treat 2013 Feb;137(3):829-836 [FREE Full text] [doi: 10.1007/s10549-012-2387-8] [Medline: 23263740]
34. Hershman DL, Unger JM, Hillyer GC, Moseley A, Arnold KB, Dakhil SR, et al. Randomized trial of text messaging to reduce early discontinuation of adjuvant aromatase inhibitor therapy in women with early-stage breast cancer: SWOG S1105. J Clin Oncol 2020 Jul 01;38(19):2122-2129 [FREE Full text] [doi: 10.1200/JCO.19.02699] [Medline: 32369401]
35. Lin C, Clark R, Tu P, Bosworth HB, Zullig LL. Breast cancer oral anti-cancer medication adherence: a systematic review of psychosocial motivators and barriers. Breast Cancer Res Treat 2017 Sep;165(2):247-260. [doi: 10.1007/s10549-017-4317-2] [Medline: 28573448]
36. Giunti G, Giunta DH, Guisado-Fernandez E, Bender JL, Fernandez-Luque L. A biopsy of breast cancer mobile applications: state of the practice review. Int J Med Inform 2018 Feb;110:1-9 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.10.022] [Medline: 29331247]
37. Zhu N, Cao J, Shen K, Chen X, Zhu S. A decision support system with intelligent recommendation for multi-disciplinary medical treatment. ACM Trans Multimed Comput Commun Appl 2020 Apr;16(1s):1-23. [doi: 10.1145/3352573]
38. Mobasheri MH, Johnston M, King D, Leff D, Thiruchelvam P, Darzi A. Smartphone breast applications - what's the evidence? Breast 2014 Oct;23(5):683-689 [FREE Full text] [doi: 10.1016/j.breast.2014.07.006] [Medline: 25153432]
39. Livingston PM, Heckel L, Orellana L, Ashley D, Ugalde A, Botti M, et al. Outcomes of a randomized controlled trial assessing a smartphone application to reduce unmet needs among people diagnosed with CancEr (ACE). Cancer Med 2020 Jan;9(2):507-516 [FREE Full text] [doi: 10.1002/cam4.2718] [Medline: 31769226]
40. Hwang H. Electronic wound monitoring after ambulatory breast cancer surgery: improving patient care and satisfaction using a smartphone app. BC Med J 2016 Oct;58(8):448-453 [FREE Full text]
41. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
42. Stubbins R, He T, Yu X, Puppala M, Ezeana CF, Chen S, et al. A behavior-modification, clinical-grade mobile application to improve breast cancer survivors' accountability and health outcomes. JCO Clin Cancer Inform 2018 Dec;2:1-11 [FREE Full text] [doi: 10.1200/CCI.18.00054] [Medline: 30652617]
43. Ekinci E, Nathoo S, Korattyil T, Vadhariya A, Zaghloul HA, Niravath PA, et al. Interventions to improve endocrine therapy adherence in breast cancer survivors: what is the evidence? J Cancer Surviv 2018 Jun;12(3):348-356. [doi: 10.1007/s11764-017-0674-4] [Medline: 29396760]
44. Yu C, Liu C, Du J, Liu H, Zhang H, Zhao Y, MISSION-2 Collaborative Group. Smartphone-based application to improve medication adherence in patients after surgical coronary revascularization. Am Heart J 2020 Oct;228:17-26. [doi: 10.1016/j.ahj.2020.06.019] [Medline: 32745732]
45. Badawy SM, Kuhns LM. Economic evaluation of text-messaging and smartphone-based interventions to improve medication adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2016 Oct 25;4(4):e121 [FREE Full text] [doi: 10.2196/mhealth.6425] [Medline: 27780795]
46. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017 Feb 2;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]

Abbreviations

ALND: axillary lymph node dissection
HER2: human epidermal growth factor receptor-2
MDT: multidisciplinary treatment
OR: odds ratio

Edited by R Kukafka; submitted 28.01.21; peer-reviewed by L Qin, L Dayer, S Badawy, MD, MS; comments to author 30.03.21; revised version received 24.05.21; accepted 21.07.21; published 16.09.21.

Please cite as:
Yu J, Wu J, Huang O, Chen X, Shen K
A Smartphone-Based App to Improve Adjuvant Treatment Adherence to Multidisciplinary Decisions in Patients With Early-Stage Breast Cancer: Observational Study
J Med Internet Res 2021;23(9):e27576
URL: https://www.jmir.org/2021/9/e27576
doi:10.2196/27576
PMID:34528890

Original Paper

Understanding Preferences for Lifestyle-Focused Visual Text Messages in Patients With Cardiovascular and Chronic Respiratory Disease: Discrete Choice Experiment

Michael Choi[1], BMedSci; Rebecca Raeside[1,2], BBioMedSci, MPH; Karice Hyun[1,2,3], BSc, MAppStat, PhD; Stephanie R Partridge[1,2,4], BSc (Hons1), PhD, APD; Aravinda Thiagalingam[1,5,6], FCSANZ, MBChB, PhD; Julie Redfern[1,2,6,7], BSc, BAppSc (Physio Hons1), PhD

[1]Westmead Applied Research Centre, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[2]Consumer Engagement and Codesign Research Hub, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[3]Department of Cardiology, Concord Repatriation General Hospital, Sydney, Australia

[4]Prevention Research Collaboration, Charles Perkins Centre, University of Sydney, Sydney, Australia

[5]Department of Cardiology, Westmead Hospital, Sydney, Australia

[6]George Institute for Global Health, University of New South Wales, Sydney, Australia

[7]Research Education Network, Western Sydney Local Health District, Sydney, Australia

Corresponding Author:
Julie Redfern, BSc, BAppSc (Physio Hons1), PhD
Westmead Applied Research Centre
Faculty of Medicine and Health
University of Sydney
Level 6, Block K, Westmead Hospital, Westmead
Sydney, 2154
Australia
Phone: 61 88909214
Email: julie.redfern@sydney.edu.au

Abstract

Background: Supporting healthy lifestyle changes is a key aim of cardiovascular and pulmonary rehabilitation programs. SMS text messaging programs have demonstrated effectiveness in cardiovascular disease risk reduction, weight loss, increasing physical activity, and smoking cessation. The optimization of SMS text messaging programs may deliver greater population benefits as mobile phone use becomes ubiquitous. Visual messaging (ie, image-based messages) has the potential to communicate health messages via digital technology and result in enhanced engagement.

Objective: This study aims to determine and understand patient preferences for lifestyle-focused visual text messages that support cardiovascular and pulmonary rehabilitation.

Methods: A discrete choice experiment was conducted in a 4-stage iterative process to elicit patient preferences for visual message features. Attribute and level development yielded 3 attributes (purpose, image type, and web address), and 16 choice sets were subsequently constructed according to a full factorial design. Patients participating in cardiovascular and pulmonary rehabilitation were surveyed (on the web) for their preferences regarding the visual message choice sets. Respondents were asked to choose among 16 pairs of visual messages regarding key lifestyle behaviors, namely, physical activity and nutrition. The data were analyzed using a conditional logit model.

Results: There was a total of 1728 observations from 54 unique respondents. Two factors that were associated with patient preference were gain-framed purpose compared with no purpose (odds ratio [OR] 1.93, 95% CI 1.40-2.65) and real images compared with cartoon images (OR 1.26, 95% CI 1.04-1.54). A loss-framed purpose was less preferred than no purpose (OR 0.55, 95% CI 0.42-0.74). Overall, patients preferred positive images that were colorful and engaged with text that supported the image and had a preference for images of real people rather than cartoons.

Conclusions: A discrete choice experiment is a scientific method for eliciting patient preferences for a visual messaging intervention that is designed to support changes in lifestyle behaviors. SMS text messaging programs that use visual aids may result in greater patient satisfaction by using a gain frame, using real images, and avoiding a loss frame. Further research is needed

to explore the feasibility of implementation and the health and behavioral outcomes associated with such visual messaging programs.

(J Med Internet Res 2021;23(9):e26224) doi:10.2196/26224

KEYWORDS

mHealth; cardiovascular disease; respiratory disease; visual communication; lifestyle change; consumer preferences; secondary prevention; rehabilitation; persuasive health technology

Introduction

Background

Noncommunicable diseases (NCDs) are the leading cause of disease burden worldwide [1]. Two major NCDs—cardiovascular disease (CVD) and chronic respiratory disease—have negative impacts on quality of life [1], health care costs [2], and worker productivity [3,4]. NCDs are generally defined by their persistent effects on health, which require long-term management [5]. NCD disease burden is largely manageable by addressing the following four key modifiable risk factors: tobacco use, alcohol misuse, physical inactivity, and diet [5]. Chronic disease management programs, also known as rehabilitation, offer group exercise and education sessions with the aim of improving key modifiable risk factors [6,7]. The benefits of both cardiac and pulmonary rehabilitation are well evidenced, which include reduced mortality, improved quality of life, and reduced hospital admissions [7,8]. However, uptake and attendance rates are poor, estimated to be between 10% and 30% globally [9,10] because of logistical barriers of time and distance [11,12].

The rapid advancement of mobile technology presents an opportunity to increase participation in existing rehabilitation programs and support self-management. There are >4.7 billion mobile phone users worldwide [13], and SMS text messaging has become an inexpensive and simple way for people to communicate. Strong evidence supports the effectiveness of SMS text messaging interventions to promote smoking cessation [14], weight loss [15], and increased physical activity [16]. A randomized controlled trial is currently underway to evaluate the effectiveness of a text message support program for patients with CVD and chronic respiratory disease [17]. This is based on a previous randomized controlled trial, the Text Me study, which showed that a lifestyle-focused SMS text messaging intervention was effective in improving cardiovascular risk factors in patients with coronary heart disease [18]. Given that visual aids have been found to enhance the engagement and understanding of health messages [19], it is possible that a visual messaging program could demonstrate similar, if not better, outcomes to previous SMS text messaging interventions for different populations; however, this is a relatively unexplored area. Researchers have explored the application of visual aids in technologies aimed at healthy behavior change, such as websites [20,21] and mobile apps [22-24]. However, there is a paucity of research investigating how visual aids may look like in a mobile messaging format. To the best of our knowledge, only one other study [25] has examined whether visual-based mobile phone messages could be used to support healthy lifestyle behavior change. However, this was tested in a cohort of young, healthy university students with no chronic disease, and the messages were stylistically simple. Therefore, we explored the feasibility of an infographic-style visual message containing both images and text. As an infographic-style visual message is made up of many characteristics (eg, color, type of image, and intention), we also investigated patient preferences with regard to the differing characteristics of this specific format. This project, which explored the use of visual aids in mobile messaging, represents a novel expansion in the growing field of persuasive health technology, which aims to use technology to support positive behavior change [26].

Understanding patient preferences regarding various aspects of SMS text messaging interventions is important for improving them. Qualitative methods such as focus groups are commonly used to gather patients' perspectives, as used by the authors of the Text Me trial to investigate why participants found the intervention useful [27]. These methods are robust; however, when trying to understand what patients like and dislike about certain features of infographic-style visual messaging (ie, whether they prefer this type of image over another), alternative methods may be better suited. For example, one SMS text messaging intervention used a quantitative method called a discrete choice experiment (DCE) to elicit patient preferences for various aspects, such as the frequency of messaging [28]. DCEs are a quantitative survey method and have been used to measure patient preferences for health care services and products [29]. DCEs define health care products and services by their attributes (ie, characteristics), where each attribute has various levels. For example, one attribute of a health care product could be price, and its levels could be US $1, US $2, and US $5. DCEs then present respondents with a choice set containing 2 or more alternatives to choose from, where each alternative is described by a different attribute-level combination. When a respondent chooses an option, their preferences for certain attributes can be estimated, and the relative weighting of attributes can be calculated [30]. Although this is a robust preference elicitation method, no DCEs to date have been used to explore the preferences for visual messaging in a population of patients with CVD and chronic respiratory disease.

Objectives

Therefore, the overall aim of this study is to understand the preferences of patients with CVD and chronic respiratory disease for lifestyle-focused visual-based messages.

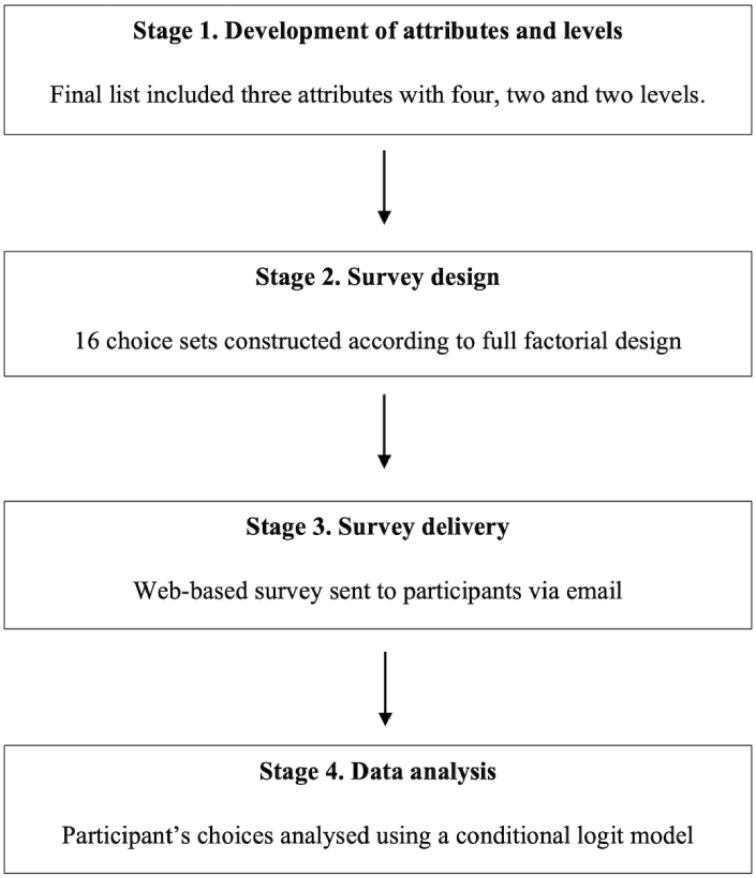
The first objective is to determine which visual messaging attribute levels were preferred by patients with CVD and chronic respiratory disease using a DCE. The second objective is to explore the reasons why patients preferred or not preferred certain visual messages using a single open-ended question.

Methods

Study Overview

This study was conducted as part of the process evaluation component of an integrated text messaging (ITM) study [17]. Ethics approval was obtained from the Ethics Review Committee of the Sydney Local Health District (HREC/16/RPAH/362), Sydney, Australia. To address the first objective of determining which visual messaging attribute levels influenced participants' choices, a DCE was conducted. This DCE followed the ISPOR (International Society for Pharmacoeconomics and Outcomes Research) guidelines on conjoint analysis methods [31]. The following stages were conducted and are described in the following sections: (1) the development of attributes and levels, (2) survey design, (3) survey delivery, and (4) data analysis (Figure 1). For addressing the second objective of exploring the reasons why certain visual messages were preferred or not preferred over others, a single open-ended question was asked after each choice set. This is described in the final section of the *Methods* section.

Figure 1. The four stages of the discrete choice experiment.



Stage 1: Development of Attributes and Levels

The development of attributes and attribute levels was carried out in two steps. First, a list of potential attributes was generated by examining the literature and discussing with experts from SMS text messaging interventions (AT, JR, and RR) [32,33]. Eight attributes were identified on the basis that patients might find them preferable: education (simple steps on how to perform a behavior) [28], framing (explaining the positive or negative outcomes of performing a behavior) [34-36], call to action (a command to perform a behavior) [37], testimonial (stories of similar people performing certain behaviors), color, mood (the overall emotional feeling), URL (a link to a web address), and a credibility indicator (acknowledgment of information as coming from a credible source) [28]. After the attributes were identified, they were reviewed independently by the research team (AT, JR, and RR), and their suggestions were then incorporated into the revised list of attributes and levels. Following DCE guidelines [29], which recommend no more than 6 attributes to minimize survey burden, we combined similar defined attributes. For example, the *URL* and *credibility* attributes were combined into a new attribute, *URL to a credible source*, because the internet site used was of high quality. Where attributes could not be combined together, they were redefined as attribute levels. For example, the attributes *color*, *mood*, and *framing* were redefined as levels *gain-frame* and *loss-frame* underneath a new attribute called *purpose*. A visual message with the *gain-frame* attribute level was specified by bright, saturated colors to create a positive mood and contained text explaining the benefits of performing the health behavior. On the other hand, a visual message with the *loss-frame* attribute level was specified by dark, unsaturated colors to create a negative mood and contained text explaining the losses of not performing the health behavior. The final list of attributes and levels derived from this iterative process is outlined (Textbox 1).

Textbox 1. Included attributes and levels of this discrete choice experiment.

Message purpose

- None (no additional information)
- Educational (simple instructions for performing health behaviors)
- Gain-frame (benefits of health action are explained)
- Loss-frame (losses of not performing health action are explained)

Image type

- Not real-life images (eg, cartoons, icons, and stick figures)
- Real-life images (eg, real-life people, objects, and places)

URL link

- Without (no web address provided)
- With (web address provided)

Stage 2: Survey Design

Given the small number of attributes and levels (3 attributes at 4, 2, and 2 levels), we decided to use a full factorial design consisting of 16 choice sets. However, no orthogonal array was available to fit this design, and so, the *idefix* package (Rstudio Inc v3.6.0) [38] was used to generate an appropriate survey design that randomized and balanced the various attribute-level profiles. The result was a DCE design with a DB-error (Bayesian D-error) efficiency of 0.4%. Once the survey design was specified, visual messages were crafted according to the required attribute-level combinations of each choice set. The content of the visual messages was sourced from the original SMS text messages developed for the ITM study (developed based on the published process) [17,39] along with stakeholder websites. The ITM SMS text messages themselves were developed based on the behavior change taxonomy of Michie et al [40]. Therefore, our visual messages were directly adapted from these original ITM SMS text messages. The first 8 choice sets centered around the message *go walking*, and the latter 8 centered around *eat healthy*. Visual messages were constructed according to the level specified for each attribute. For example, for the attribute levels *educational*, *nonreal*, and *with URL*, the actions required to follow the proposed health behavior would be explained, a cartoon image would be inserted, and a URL link would be attached (Figures 2 and 3). Similarly, for the attribute levels *gain-frame*, *real image*, and *without URL*, a real-life photograph would be inserted, and bright saturated colors would be used to emphasize the benefits of taking that certain health action (Figures 2 and 3). Different variations in font, sizing, and layout were considered for each visual message. After an iterative design process for refinement, 32 visual messages were selected. Each choice set was preceded by the question stem "If you were going to improve your diet or your level of physical activity, which visual message would you prefer to receive?"

Figure 2. Example of a choice set from the discrete choice experiment survey.

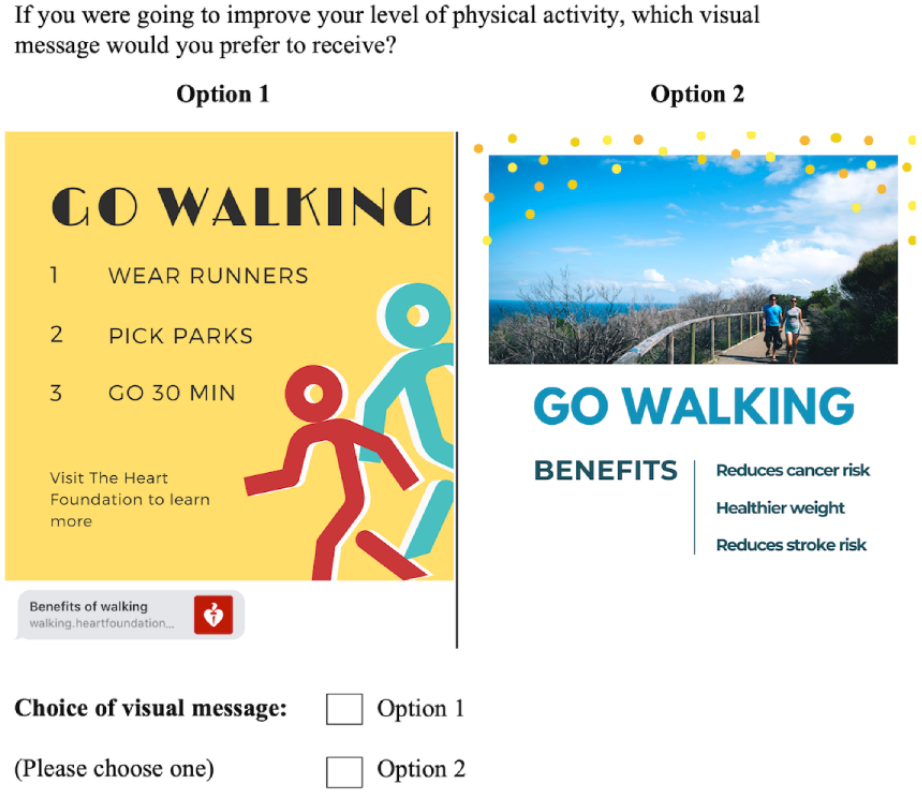
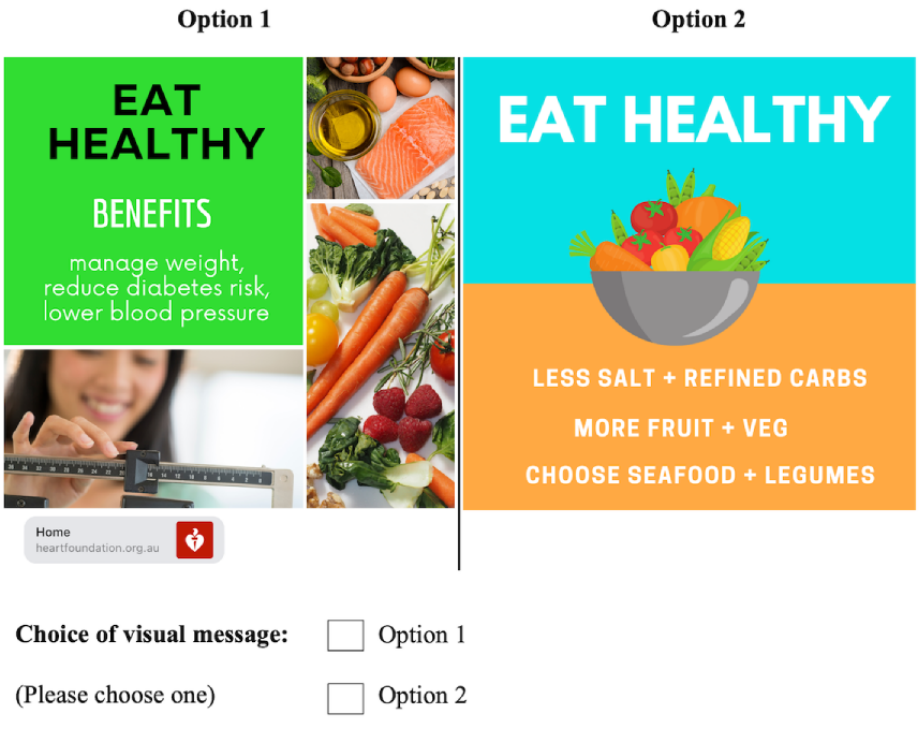


Figure 3. Example of a "go walking" choice set from the discrete choice experiment survey.



Stage 3: Survey Delivery

Before the delivery of the web-based survey to participants, it was pilot tested among the research team. The aim was to explore the acceptability of the DCE survey, the time taken to complete it, and the ease of understanding. Once refinements were made, a link to the web-based survey was sent to 141 participants via their nominated email address. These participants had already been recruited as part of the ITM study [17]. Eligible respondents consisted of adults with a medical history of CVD (including coronary heart disease, cardiomyopathy, peripheral arterial disease, and stroke) or chronic respiratory disease (chronic obstructive pulmonary disease, chronic bronchitis, emphysema, chronic asthma, and bronchiectasis). These participants were recruited from 6 rehabilitation clinics across Sydney, Australia. Eligible respondents were identified by clinical rehabilitation staff and invited to participate in the ITM study. Participants provided informed consent to participate in message feedback and improvement. Of the 141 participants, 54 (38.3%) people responded to the survey. Out of the 54 respondents, 6 (11%) respondents did not fully complete all questions.

Stage 4: Data Analysis

Overview

Data analysis was informed by the ISPOR guidelines on appropriate statistical methods for DCEs [41]. Analyses were performed using SAS v9.4 (SAS Institute). Patient characteristics were summarized as mean and SD for continuous variables and number and percentage for categorical variables. To analyze the DCE, a conditional logit model, also known as McFadden's choice model [42], was used, and odds ratios (ORs) and 95% CIs were estimated. The conditional logit model is based on the random utility theory, which states that choices are made because they bring utility to the individual [43]. It was assumed that the patients would choose the visual message (1 or 0) that maximizes their utility depending on the attributes. This was formulated as follows:

$$U\ (1,\ 0) = \beta_1\ educational + \beta_2\ gain\text{-}frame + \beta_3\ loss\text{-}frame + \beta_4\ real\ image + \beta_5\ URL + \quad \textbf{(1)}$$

Single Open-ended Question

For addressing the second objective, participants were asked an open-ended question to explore the reasons why certain visual messages were chosen or not chosen over others. After each choice set, participants were asked, "What did you like or dislike about the visual message you have chosen?" Participants typed their responses in a free text box (Figures 2 and 3). An inductive thematic analysis was conducted to analyze these responses [44]. First, the research team familiarized themselves with the data by reading through the responses. Second, a set of initial codes was generated, and each response was tagged with one or more of these codes. As this was an iterative method, new codes were added as new elements came up in the data. Next, the codes were grouped under themes, which were defined as patterns that describe something significant or interesting about the data. These themes were then reviewed, and definitions were developed. Overlapping themes were separated, and subthemes were developed. The interactions between themes were also examined.

Results

Overview

We received 54 surveys that resulted in a total of 1728 observations.

Respondent Characteristics

Most of the study sample (mean age 69.7 years, SD 8.7) was male (34/54, 62%), was of White ethnicity (43/54, 84%), and was diagnosed with CVD (32/54, 59%; Table 1). The median level of education in the cohort was diploma certification.

Table 1. Baseline demographic and clinical characteristics (N=54).

Demographic	Values
Age (years), mean (SD)	69.7 (8.7)
Gender, n (%)	
Men	34 (62)
Women	20 (37)
Ethnicity[a,b], n (%)	
White	43 (84)
Aboriginal or Torres Strait Islander, Chinese, Japanese, Malay, South Asian[c], Arab or Persian, Black African, Sub-Saharan African, or Native American Indian	6 (12)
Other	2 (4)
Education[a,d], median (IQR)	5.0 (4.0-7.0)
Clinical disease type, n (%)	
Diagnosed with CVD[e]	32 (59)
Diagnosed with chronic respiratory disease	22 (41)

[a]As baseline data were collected from previous medical records, we collected all available data at the time of extraction.

[b]Ethnicity data were only available for 51 participants.

[c]Bangladesh, India, Nepal, Pakistan, or Sri Lanka.

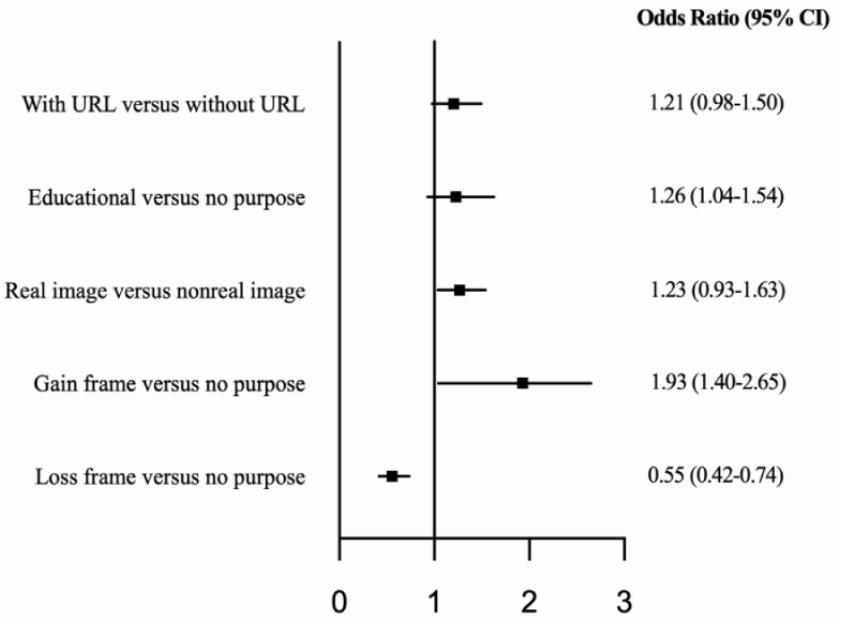
[d]Education data were only available for 51 participants.

[e]CVD: cardiovascular disease.

Visual Messaging Attribute Levels That Influenced Preferences

As per the first objective, a DCE was used to determine which visual messaging attribute levels influenced preferences. According to the conditional logit model, the following three attribute levels significantly influenced participant preferences: a gain frame, a loss frame, and real images. Visual messages with a gain-framed purpose were approximately 93% more likely to be chosen over messages with no specific purpose (OR 1.93, 95% CI 1.40-2.65; Figure 4). Conversely, visual messages with a loss-framed purpose were approximately 45% less likely to be chosen than messages compared with no purpose (OR 0.55, 95% CI 0.42-0.74). Visual messages with a real image were 26% more likely to be chosen than messages with a nonreal image (OR 1.26, 95% CI 1.04-1.54).

Figure 4. Attribute levels of visual messages, which influenced participant choice.



Reasons for Choosing Certain Visual Messages

As per the second objective, an open-ended question was asked after each choice set to explore the reasons why certain visual messages were preferred over others. The inductive thematic analysis resulted in the following six themes and subthemes: positive emotions (positivity, healthy, and cheerfulness), visual appeal (colorful, engaging image, high brightness, and high contrast), ease of understanding (clarity, large font, relationship between image and message, and concise), usefulness (benefits explained, informative, encouraging, and URL link), image reality (natural beauty, real people, and real food), and disliked features (forced to make a choice, irrelevant, condescending, negative emotions, and ignoring mental health). Illustrative quotations are provided in Table 2, and the conceptual patterns and relationships among the themes are shown in Figure 5.

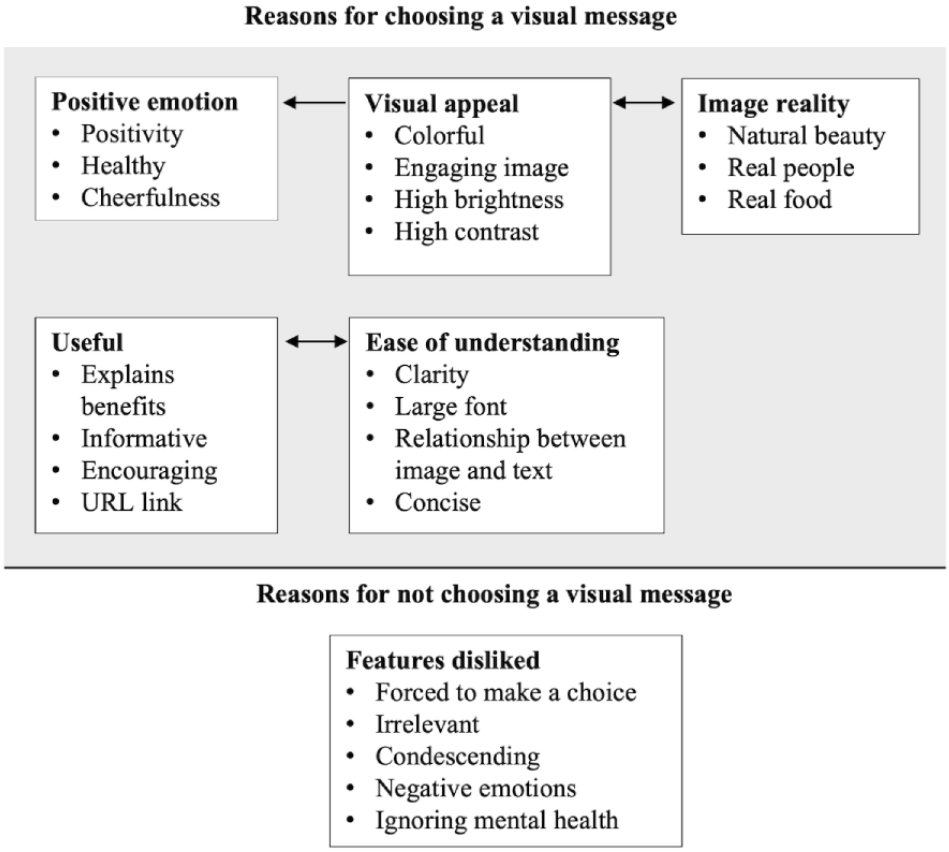
Table 2. Illustrative quotations of what participants liked and disliked about the visual messages.

Theme	Quotations	Participant characteristics (gender, age, and disease type)
Positive emotions		
Positivity	"Stands out with a more positive vibe."	Male, 56 years, COPD[a]
Cheerful	"Much cheerier–colour probably."	Female, 66 years, CVD[b]
Visual appeal		
Colorful	"Bright colours encourage people to look, read and absorb idea."	Male, 53 years, CVD
Engaging image	"Photo of good-looking food makes me want to eat better."	Female, 57 years, COPD
Ease of understanding		
Simplicity	"Clear. Easy to read."	Male, 29 years, CVD
Relationship between image and message	"Photos supported text clearly and effectively. Good clear message."	Male, 48 years, CVD
Usefulness		
Benefits explained	"It explains what the things [benefits] that you can get from walking."	Male, 43 years, COPD
URL link	"Links to Heart Foundation [website]."	Male, 36 years, CVD
Image reality		
Nature images	"The real outdoor picture instantly draws you to a feeling of health, fresh air and sunshine."	Female, 42 years, COPD
Real people	"I'd rather see photos of people than cartoonish drawings. The other option is very unattractive!"	Female, 66 years, CVD
Features disliked		
Forced to make a choice	"Neither is that good. Need to give reason rather than lecture or order exercise."	Male, 56 years, CVD
Condescending or offensive tone	"Both are terrible assuming people are idiots."	Female, 57 years, CVD
Negative emotions	"Yet again try to stay away from negative images."	Male, 40 years, CVD

[a]COPD: chronic obstructive pulmonary disease.

[b]CVD: cardiovascular disease.

Figure 5. Illustrative quotations of what participants liked and disliked about the visual messages.



Discussion

Principal Findings

To our knowledge, this is the first study to examine the preferences of patients with CVD and chronic respiratory disease for varying characteristics of lifestyle-focused visual-based messages to be sent digitally. As per the first objective, this study used a DCE to determine whether 3 visual messaging attribute levels significantly influenced participant preferences: gain frame, loss frame, and real images. As per the second objective, this study analyzed the responses to a single open-ended question and discovered the reasons for choosing or not choosing certain visual messages. The reasons for choosing a visual message included positive emotion, visual appeal, image reality, ease of understanding, and usefulness. The reasons for not choosing a visual message included irrelevant information, condescending tone, and negative emotions. These findings have important design implications for infographic-style health messages, suggesting that gain-framed messages with lifelike photos are likely to be well received by patients with CVD and chronic respiratory disease.

Comparison With Prior Work

This study found that framing had a strong influence on participants' preferences; gain-framed messages were liked, and loss-framed messages were disliked. This proclivity toward gain framing and aversion to loss framing is a well-known phenomenon reported in the health communication literature. It is unclear why framing effects occur. Therefore, understanding how framing can effect preferences may inform future implementation of visual messaging programs. It has been suggested that the framing effect depends on the type of behavior being communicated. For example, Rothman and Salovey [45] discovered that gain-framed messages are more effective than loss-framed messages in promoting disease-preventing behaviors. In support of this idea, a meta-analytic review of 94 studies demonstrated that gain-framed messages were slightly more effective than loss-framed messages in promoting disease-preventing behaviors [46]. Rothman and Salovey [45] also discovered that loss-framed messages were more effective than gain-framed messages in promoting disease-detecting behaviors. They proposed that this occurs because of prospect theory [47], which states that people will be risk averse when considering actions with guaranteed benefits and risk seeking when considering actions with potential losses. In this study, two disease-preventing behaviors (walking and healthy eating) were communicated to participants, and it was found that gain-framed messages were preferred. In line with the work of Rothman and Salovey [45], we suggest that the observed framing effect is because of the fact that the behaviors being promoted are disease preventing and that if they were disease-detecting behaviors, then loss-framed messages would be preferred over gain-framed messages. Therefore, when incorporating framing appeals into health messages, it is important to consider what type of behavior is being advocated so that the message is well received by participants.

Another line of research has proposed that framing effects may depend on the individual's disposition. For example, Covey [48] suggested that framed messages that are congruent with

an individual's motivation are likely to be preferred, that is, gain-framed messages are more persuasive when the recipient is reward oriented, and loss-framed messages are more persuasive when the recipient is threat aversive. It is unknown what the motivational status of this population is, although some evidence suggests that the patients undergoing cardiac rehabilitation are likely to be threat aversive as they have undergone a major coronary event and would want to avoid it again in the future [19]. In that case, the findings of this study contradict the findings of Covey [48]; however, this is difficult to tease out, as we do not know whether this population is threat aversive or reward oriented. Therefore, it is unknown whether the framing effects in this study are mediated by the disposition of the participants. Future research is needed to test whether the motivational status of people with chronic diseases might mediate framing effects.

In discussing framing effects, one study found little impact because of message framing. Niu et al [25] examined the impact of framed SMS text messages on encouraging healthy eating practices. They discovered that participants' cognitive responses did not differ between gain-framed and loss-framed messages, that is, there was no observable difference because of framing. However, there are at least 3 factors that may explain the discrepancy in findings between this study and the study by Niu et al [25]. First, the studies had different aims; this study sought to understand one's choices, and the other examined one's cognitive processes. Second, different methodologies were used; this study used a DCE, and the other study used a questionnaire. Finally, the samples were different demographically and clinically; this study examined an older cohort with chronic disease, and the other study examined a group of young, healthy undergraduates. These three factors may explain why this paper discovered a strong response to framing and why Niu et al [25] saw virtually no response to framing.

Although framing was an influential design attribute, there were other factors that may have accounted for the large difference in preferences for gain-framed and loss-framed messages. For example, certain colors may have attracted some respondents to that particular visual message. In this DCE, bright saturated colors were used when designing gain-framed messages. In contrast, dark colors were used when designing loss-framed messages. The purpose of the color was to accentuate the respective gains and losses that the supposed action would bring. Color, hue, and saturation preferences have been demonstrated [19,49], which may have influenced the participants' choices. The findings of the thematic analysis support the idea that messages with positive colors are preferred. Therefore, the strong preference for gain-framed messages was likely influenced by the use of positive colors.

Our results highlight that adapting messages into a digital visual format requires the consideration of several design features. For example, the use of images is a key design feature. As the digital age advances, capturing and communicating photographs will become easier and can be used in mobile health (mHealth) interventions. Although it is understood that visual media can impact the processes leading to behavior change [19,25], no studies to date have examined whether the type of image might have a different impact. This study found preferences for real images over nonreal ones in a population of cardiac and pulmonary rehabilitation patients. It is known that photos of people are powerful, persuasive tools as viewers can establish a relationship with the subject [50,51]. They elicit immediate emotional reactions that can affect their behavior. The effects of visual images on behavior change have also been reported when there are demographic similarities between the subject in the photograph and the viewer [52]. It is unknown whether this effect also applies to patients with chronic diseases. Future studies can examine whether visual image tailoring may influence the reception of visual messages.

Text is another important element of visual messages. In this study, we paired brief textual information with images. In general, participants preferred these messages to messages with no text as the text explained the visual message of the photo in a manner that was simple and easy to understand. This finding adds to the current understanding of effective infographic-style health messages. In fact, simplicity as a design principle has been identified as an important factor influencing user acceptability of mHealth technologies in patients with CVD [53]. However, messages that were too simplistic and contained no textual information were generally disliked. Instead, what patients seemed to prefer was a textual message that gave meaning to the visual aid in a clear and concise way. Factors such as font sizing, colors, layout, and positioning all contribute to how the text in a visual message is perceived and should be carefully considered to maximize clarity.

An interesting feature (further research to evaluate) used in this study's visual messages was a URL link. The purpose of this was to act as a means for people to pursue more information. The presence of the URL link also serves as a credibility indicator. The addition of a URL link is a benefit of SMS text messaging as a communication modality. Not only can respondents see the website address, but those with internet capabilities are also able to follow the link immediately and view the website's information. The findings from this study were aligned with those of past studies examining the role of visual aids in supporting chronic disease education. Visual aids have been used across a variety of formats, including print materials, websites, and mobile apps, to support the education of patients with CVD and chronic respiratory disease [20,22,24,54-57]. However, it is unknown whether visual-based messages can be used. The findings from this study suggest that there is great potential for visual messages that combine real images with a simple lifestyle-focused message to be well received by patients undergoing chronic disease rehabilitation.

The strength of this study lies in its DCE methodology. DCEs have many advantages over traditional surveys in eliciting preferences. First, DCE is an accepted survey method. The benefit is that best practice guidelines were able to be followed both in experimental design and also in statistical analysis [31,41]. Although the DCE originated in economics [30] and has been widely used to determine preferences for health services and products [29], this is the first instance of a DCE being used to explore the preferences for visual messaging. This only highlights the versatility of a DCE and its future potential to be used more widely in examining patient preferences in the field of mHealth interventions.

Another key advantage of DCE is its quantitative analysis. Rather than using a traditional questionnaire that yields qualitative results, DCE can measure preferences [29]. Of course, learning about patient preferences is enhanced by both qualitative and quantitative methods. In fact, qualitative methods such as focus groups are highly recommended when developing an initial idea of attributes, which the target population may care about [33]. The full power of the DCE is realized when the attributes are numeric. This allows for powerful analyses, where choice can be graphed as a dependent variable of attribute variables. However, in this study, the attributes were categorical variables (message purpose, image type, and URL link); thus, this type of analysis was not possible. Nevertheless, quantitative results were obtained in the form of an OR, allowing the relative size of the attributes to be ascertained.

DCE is also a robust method, as it indirectly measures preferences. Participants were asked to choose the message they preferred but not directly told of the different attributes that comprised each message. This reduced participants' bias, as they have to consider which message they actually prefer by comparing the two. Alternative preference elicitation methods could have also been used. For example, multiprofile best-worst scaling is a method in which the participant is asked to rank several choices from best to worst [58]. A variant of this method is single-profile best-worst scaling, in which a participant is presented with a single option and asked to state the best and worst aspects. Therefore, the DCE is a robust quantitative method that holds great potential for future studies on patient preferences for mHealth interventions.

Limitations

One of the limitations of this study was the lack of a no-choice alternative. Participants were offered to choose between two visual messages with no choice to opt out. This is an important limitation as thematic analysis of the free text responses revealed that some participants did not like either visual message, suggesting that they would have likely chosen the opt out if it was offered. In choice sets where there was no clear preference for the participant, an opt out would improve the validity of the results and better reflect their preferences. Therefore, future DCEs exploring patient preferences for visual messaging should use a no-choice alternative.

Another limitation of this study was that some attributes might have been missed in the DCE. Attribute selection is an important step in running a DCE [31]. However, as DCEs work best when focusing on only a few attributes, other attributes may have been excluded from the survey. Therefore, there may have been attributes that influenced patient preferences but were not tested in the DCE. To ensure that all attributes are captured in the survey, it is useful to use qualitative methods, such as focus groups and interviews [33]. Owing to COVID-19, these face-to-face qualitative methods were not possible, limiting the scope of possible attributes that this DCE could have examined. In the future, given the uncertainties of the current health situation, we would like to conduct qualitative interviews over the phone or via videoconferencing to obtain valuable participant input in the development of visual messages.

In relation to the design of the visual messages, although they were developed using a systematic and iterative process, we did not pilot test them before seeking consumer feedback. Therefore, we have identified this as a limitation and, in the future, would like to test different variations of font, sizing, and layout to determine optimal combinations. In relation to the 87 participants who were sent the survey link but did not open it, there is no way of knowing their preferences, so we cannot assume that they disliked the messages from the fact that they did not open the survey.

The fourth limitation was the number of choice sets. The DCE guidelines state that the ideal number of choice tasks is 8-16 [29]. This survey was constructed with 16 choice tasks, which were thought to be an acceptable length for respondents. However, there were a number of partially completed surveys, suggesting that there were too many choice tasks. Therefore, a shorter survey may have resulted in more completions. Instead of using a full fractional design, future DCEs should use a fractional factorial design, which will allow the number of choice tasks to be reduced so that the survey is less burdensome to participants.

The strength of the investigation was limited by the small sample size. Therefore, the results are not generalizable to a wider population of people undergoing both cardiac and pulmonary rehabilitation. The small sample size also meant that data analysis was limited. In this study, a conditional logit model was used to examine the associations between the choice outcome and the visual message attributes [42]. With a larger sample size, a different model could have been used, namely, a multinomial logit model, which examines associations between the choice outcome and demographic characteristics as well as the visual message attributes. Other DCEs have used a multinomial model to examine whether preferences are influenced by age, sex, ethnicity, income, and education [28]. Future DCEs with sufficiently large sample sizes should investigate the effect of demographics on preferences. Finally, a limitation arises from the fact that this research was an additional substudy with recruitment from the ITM trial, which was not specifically designed to test visual messages [17].

Conclusions

This study explored the preferences for lifestyle-focused visual messages in patients with CVD and chronic respiratory disease and found that message framing and image type were important characteristics. Framing was the most important factor that influenced preferences. A strong preference for gain-framed messages with real images was shown. As such, when designing visual messaging interventions for patients with chronic conditions, these attributes may influence the acceptability of messages. Owing to COVID-19 restrictions, the DCE was delivered on the web. Further research is needed to deliver visual media content using digital technology and determine whether the identified preferences actually hold. Future studies can examine the implementation of these visual messages and their effects on chronic disease rehabilitation health and behavioral outcomes.

Acknowledgments

The authors thank all of the staff and participants who supported and participated in this study and the integrated text messaging study. This work was supported by a National Heart Foundation 2015 New South Wales Cardiovascular Research Network Research Development Project grant (grant 101134). JR was supported by the National Health and Medical Research Council Career Development Fellowship (Application ID: 1143538). SRP was supported by the National Health and Medical Research Council of Australia and the National Heart Foundation Early Career Fellowship (Application ID: 1157438). KH was supported by a National Heart Foundation Postdoctoral Fellowship (Application ID: 102138).

Authors' Contributions

MC and RR managed the data collection. MC and KH conducted the data analysis and interpretation. MC, SRP, AT, and JR contributed to the conception of the study, the study design, the interpretation of the data, and the drafting and critical revision of the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1736-1788 [FREE Full text] [doi: 10.1016/S0140-6736(18)32203-7] [Medline: 30496103]
2. Vandenberghe D, Albrecht J. The financial burden of non-communicable diseases in the European Union: a systematic review. Eur J Public Health 2020 Aug 01;30(4):833-839. [doi: 10.1093/eurpub/ckz073] [Medline: 31220862]
3. Carter HE, Schofield D, Shrestha R. Productivity costs of cardiovascular disease mortality across disease types and socioeconomic groups. Open Heart 2019 Feb 16;6(1):e000939 [FREE Full text] [doi: 10.1136/openhrt-2018-000939] [Medline: 30997129]
4. Patel JG, Coutinho AD, Lunacsek OE, Dalal AA. COPD affects worker productivity and health care costs. Int J Chron Obstruct Pulmon Dis 2018 Jul 30;13:2301-2311 [FREE Full text] [doi: 10.2147/COPD.S163795] [Medline: 30104870]
5. Noncommunicable diseases. World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2021-08-10]
6. Dalal HM, Doherty P, Taylor RS. Cardiac rehabilitation. BMJ 2015 Sep 29;351:h5000 [FREE Full text] [doi: 10.1136/bmj.h5000] [Medline: 26419744]
7. McCarthy B, Casey D, Devane D, Murphy K, Murphy E, Lacasse Y. Pulmonary rehabilitation for chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2015 Feb 23(2):CD003793. [doi: 10.1002/14651858.CD003793.pub3] [Medline: 25705944]
8. Anderson L, Taylor RS. Cardiac rehabilitation for people with heart disease: an overview of Cochrane systematic reviews. Cochrane Database Syst Rev 2014 Dec 12;2014(12):CD011273 [FREE Full text] [doi: 10.1002/14651858.CD011273.pub2] [Medline: 25503364]
9. Clark RA, Conway A, Poulsen V, Keech W, Tirimacco R, Tideman P. Alternative models of cardiac rehabilitation: a systematic review. Eur J Prev Cardiol 2015 Jan;22(1):35-74. [doi: 10.1177/2047487313501093] [Medline: 23943649]
10. Bjarnason-Wehrens B, McGee H, Zwisler A, Piepoli MF, Benzer W, Schmid J, Cardiac Rehabilitation Section European Association of Cardiovascular PreventionRehabilitation. Cardiac rehabilitation in Europe: results from the European Cardiac Rehabilitation Inventory Survey. Eur J Cardiovasc Prev Rehabil 2010 Aug;17(4):410-418. [doi: 10.1097/HJR.0b013e328334f42d] [Medline: 20300001]
11. Keating A, Lee A, Holland AE. What prevents people with chronic obstructive pulmonary disease from attending pulmonary rehabilitation? A systematic review. Chron Respir Dis 2011;8(2):89-99 [FREE Full text] [doi: 10.1177/1479972310393756] [Medline: 21596892]
12. Rose M, Timmons SM, Amerson R, Reimels E, Pruitt RH. Facilitators and barriers in cardiac rehabilitation participation: an integrative review. J Nurse Pract 2011 May;7(5):399-408. [doi: 10.1016/j.nurpra.2011.02.003]
13. Number of mobile phone users worldwide from 2015 to 2020 (in billions). Statista. URL: http://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide/ [accessed 2016-09-03]
14. Scott-Sheldon LA, Lantini R, Jennings EG, Thind H, Rosen RK, Salmoirago-Blotcher E, et al. Text messaging-based interventions for smoking cessation: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016 May 20;4(2):e49 [FREE Full text] [doi: 10.2196/mhealth.5436] [Medline: 27207211]
15. Siopis G, Chey T, Allman-Farinelli M. A systematic review and meta-analysis of interventions for weight management using text messaging. J Hum Nutr Diet 2015 Feb;28 Suppl 2:1-15. [doi: 10.1111/jhn.12207] [Medline: 24480032]
16. Stephens J, Allen J. Mobile phone interventions to increase physical activity and reduce weight: a systematic review. J Cardiovasc Nurs 2013;28(4):320-329 [FREE Full text] [doi: 10.1097/JCN.0b013e318250a3e7] [Medline: 22635061]

17. Redfern J, Hyun K, Singleton A, Hafiz N, Raeside R, Spencer L, et al. ITM support for patients with chronic respiratory and cardiovascular diseases: a protocol for a randomised controlled trial. BMJ Open 2019 Mar 01;9(3):e023863 [FREE Full text] [doi: 10.1136/bmjopen-2018-023863] [Medline: 30826759]
18. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of lifestyle-focused text messaging on risk factor modification in patients with coronary heart disease: a randomized clinical trial. JAMA 2015 Sep;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
19. Houts PS, Doak CC, Doak LG, Loscalzo MJ. The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. Patient Educ Couns 2006 May;61(2):173-190. [doi: 10.1016/j.pec.2005.05.004] [Medline: 16122896]
20. Melholt C, Joensson K, Spindler H, Hansen J, Andreasen JJ, Nielsen G, et al. Cardiac patients' experiences with a telerehabilitation web portal: implications for eHealth literacy. Patient Educ Couns 2018 May;101(5):854-861. [doi: 10.1016/j.pec.2017.12.017] [Medline: 29305064]
21. Molan N, Emmanuel S, Langley T, Holloway CJ. Evaluating the effectiveness of an online cardiac rehabilitation resource (www.svhhearthealth.com.au) in improving knowledge and confidence for patients with newly diagnosed cardiac conditions: a pre-experimental pilot study. Heart Lung Circ 2019 May;28(5):761-770. [doi: 10.1016/j.hlc.2018.03.013] [Medline: 29691158]
22. Varnfield M, Karunanithi M, Lee C, Honeyman E, Arnold D, Ding H, et al. Smartphone-based home care model improved use of cardiac rehabilitation in postmyocardial infarction patients: results from a randomised controlled trial. Heart 2014 Nov;100(22):1770-1779 [FREE Full text] [doi: 10.1136/heartjnl-2014-305783] [Medline: 24973083]
23. Forman DE, LaFond K, Panch T, Allsup K, Manning K, Sattelmair J. Utility and efficacy of a smartphone application to enhance the learning and behavior goals of traditional cardiac rehabilitation: a feasibility study. J Cardiopulm Rehabil Prev 2014;34(5):327-334. [doi: 10.1097/HCR.0000000000000058] [Medline: 24866355]
24. Farmer A, Williams V, Velardo C, Shah SA, Yu L, Rutter H, et al. Self-management support using a digital health system compared with usual care for chronic obstructive pulmonary disease: randomized controlled trial. J Med Internet Res 2017 May 03;19(5):e144. [doi: 10.2196/jmir.7116] [Medline: 28468749]
25. Niu Z, Jeong DC, Brickman J, Nam Y, Liu S, Stapleton JL. A picture worth a thousand texts? Investigating the influences of visual appeals in a text message-based health intervention. J Health Commun 2020 Mar 03;25(3):204-213. [doi: 10.1080/10810730.2020.1731631] [Medline: 32096455]
26. Orji R, Moffatt K. Persuasive technology for health and wellness: state-of-the-art and emerging trends. Health Informatics J 2018 Mar;24(1):66-91 [FREE Full text] [doi: 10.1177/1460458216650979] [Medline: 27245673]
27. Redfern J, Santo K, Coorey G, Thakkar J, Hackett M, Thiagalingam A, et al. Factors influencing engagement, perceived usefulness and behavioral mechanisms associated with a text message support program. PLoS One 2016 Oct 14;11(10):e0163929 [FREE Full text] [doi: 10.1371/journal.pone.0163929] [Medline: 27741244]
28. Ramirez M, Wu S, Beale E. Designing a text messaging intervention to improve physical activity behavior among low-income Latino patients with diabetes: a discrete-choice experiment, Los Angeles, 2014-2015. Prev Chronic Dis 2016 Dec 22;13:E171 [FREE Full text] [doi: 10.5888/pcd13.160035] [Medline: 28005532]
29. Ryan M, Gerard K, Amaya-Amaya M. Discrete choice experiments in a nutshell. In: Using Discrete Choice Experiments to Value Health and Health Care. Dordrecht, Netherlands: Springer Nature; 2008.
30. Lancaster KJ. A new approach to consumer theory. J Polit Econ 1966 Apr;74(2):132-157. [doi: 10.1086/259131]
31. Bridges JF, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value Health 2011 Jul;14(4):403-413 [FREE Full text] [doi: 10.1016/j.jval.2010.11.013] [Medline: 21669364]
32. Coast J, Horrocks S. Developing attributes and levels for discrete choice experiments using qualitative methods. J Health Serv Res Policy 2007 Jan;12(1):25-30. [doi: 10.1258/135581907779497602] [Medline: 17244394]
33. Coast J, Al-Janabi H, Sutton EJ, Horrocks SA, Vosper AJ, Swancutt DR, et al. Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. Health Econ 2012 Jun;21(6):730-741. [doi: 10.1002/hec.1739] [Medline: 21557381]
34. Latimer AE, Brawley LR, Bassett RL. A systematic review of three approaches for constructing physical activity messages: what messages work and what improvements are needed? Int J Behav Nutr Phys Act 2010 May 11;7:36 [FREE Full text] [doi: 10.1186/1479-5868-7-36] [Medline: 20459779]
35. Zhao X, Villagran MM, Kreps GL, McHorney C. Gain versus loss framing in adherence-promoting communication targeting patients with chronic diseases: the moderating effect of individual time perspective. Health Commun 2012;27(1):75-85. [doi: 10.1080/10410236.2011.569002] [Medline: 21745039]
36. O'Keefe DJ, Jensen JD. The relative persuasiveness of gain-framed and loss-framed messages for encouraging disease prevention behaviors: a meta-analytic review. J Health Commun 2007;12(7):623-644. [doi: 10.1080/10810730701615198] [Medline: 17934940]
37. Thrasher JF, Anshari D, Lambert-Jessup V, Islam F, Mead E, Popova L, et al. Assessing smoking cessation messages with a discrete choice experiment. Tob Regul Sci 2018 Mar;4(2):73-87 [FREE Full text] [doi: 10.18001/TRS.4.2.7] [Medline: 30828595]

38. Traets F, Sanchez DG, Vandebroek M. Generating optimal designs for discrete choice experiments in R: the idefix package. J Stat Softw 2020 Nov 29;96(3):1-41. [doi: 10.18637/jss.v096.i03]
39. Redfern J, Thiagalingam A, Jan S, Whittaker R, Hackett ML, Mooney J, et al. Development of a set of mobile phone text messages designed for prevention of recurrent cardiovascular events. Eur J Prev Cardiol 2014 Apr;21(4):492-499. [doi: 10.1177/2047487312449416] [Medline: 22605787]
40. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
41. Hauber AB, González JM, Groothuis-Oudshoorn CG, Prior T, Marshall DA, Cunningham C, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR conjoint analysis good research practices task force. Value Health 2016 Jun;19(4):300-315 [FREE Full text] [doi: 10.1016/j.jval.2016.04.004] [Medline: 27325321]
42. McFadden D. Conditional logit analysis of qualitative choice behaviour. In: Frontiers in Econometrics. New York: Academic Press; 1974.
43. Manski CF. The structure of random utility models. Theor Decis 1977 Jul 1;8(3):229-254. [doi: 10.1007/bf00133443]
44. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods 2017 Oct 02;16(1):160940691773384. [doi: 10.1177/1609406917733847]
45. Rothman AJ, Salovey P. Shaping perceptions to motivate healthy behavior: the role of message framing. Psychol Bull 1997 Jan;121(1):3-19. [doi: 10.1037/0033-2909.121.1.3] [Medline: 9000890]
46. Gallagher KM, Updegraff JA. Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. Ann Behav Med 2012 Feb;43(1):101-116. [doi: 10.1007/s12160-011-9308-7] [Medline: 21993844]
47. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. Science 1981 Jan 30;211(4481):453-458. [doi: 10.1126/science.7455683] [Medline: 7455683]
48. Covey J. The role of dispositional factors in moderating message framing effects. Health Psychol 2014 Jan;33(1):52-65. [doi: 10.1037/a0029305] [Medline: 22924446]
49. Hurlbert A, Ling Y. Understanding colour perception and preference. In: Colour Design (Second Edition). Sawston, United Kingdom: Woodhead Publishing Series in Textiles; 2017.
50. Joffe H. The power of visual material: persuasion, emotion and identification. Diogenes 2008 Feb 01;55(1):84-93. [doi: 10.1177/0392192107087919]
51. Richards AS, Hample D. Facial similarity mitigates the persuasive effects of source bias: an evolutionary explanation for kinship and susceptibility to influence. Commun Monogr 2015 Feb 26;83(1):1-24. [doi: 10.1080/03637751.2015.1014822]
52. Buller MK, Bettinghaus EP, Fluharty L, Andersen PA, Slater MD, Henry KL, et al. Improving health communication with photographic images that increase identification in three minority populations. Health Educ Res 2019 Apr 01;34(2):145-158 [FREE Full text] [doi: 10.1093/her/cyy054] [Medline: 30726902]
53. Neubeck L, Lowres N, Benjamin EJ, Freedman SB, Coorey G, Redfern J. The mobile revolution--using smartphone apps to prevent cardiovascular disease. Nat Rev Cardiol 2015 Jun;12(6):350-360. [doi: 10.1038/nrcardio.2015.34] [Medline: 25801714]
54. Clark M, Kelly T, Deighan C. A systematic review of the Heart Manual literature. Eur J Cardiovasc Nurs 2011 Mar;10(1):3-13. [doi: 10.1016/j.ejcnurse.2010.03.003] [Medline: 20451459]
55. DeWalt DA, Malone RM, Bryant ME, Kosnar MC, Corr KE, Rothman RL, et al. A heart failure self-management program for patients of all literacy levels: a randomized, controlled trial [ISRCTN11535170]. BMC Health Serv Res 2006 Mar 13;6:30 [FREE Full text] [doi: 10.1186/1472-6963-6-30] [Medline: 16533388]
56. Bucknall CE, Miller G, Lloyd SM, Cleland J, McCluskey S, Cotton M, et al. Glasgow supported self-management trial (GSuST) for patients with moderate to severe COPD: randomised controlled trial. BMJ 2012 Mar 06;344:e1060 [FREE Full text] [doi: 10.1136/bmj.e1060] [Medline: 22395923]
57. Sartori AC, Rodrigues Lucena TF, Lopes CT, Picinin Bernuci M, Yamaguchi MU. Educational intervention using on medication adherence in hypertension and diabetes patients: a randomized clinical trial. Telemed J E Health 2020 Dec;26(12):1526-1532. [doi: 10.1089/tmj.2019.0305] [Medline: 32155382]
58. Mühlbacher AC, Kaczynski A, Zweifel P, Johnson FR. Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview. Health Econ Rev 2016 Dec;6(1):2 [FREE Full text] [doi: 10.1186/s13561-015-0079-x] [Medline: 26743636]

Abbreviations

CVD: cardiovascular disease
DCE: discrete choice experiment
ISPOR: International Society for Pharmacoeconomics and Outcomes Research
ITM: integrated text messaging
mHealth: mobile health
NCD: noncommunicable disease
OR: odds ratio

Edited by R Kukafka; submitted 07.12.20; peer-reviewed by R Krukowski, T Ndabu, J Job, A McLean; comments to author 20.01.21; revised version received 15.02.21; accepted 21.07.21; published 20.09.21.

Please cite as:
Choi M, Raeside R, Hyun K, Partridge SR, Thiagalingam A, Redfern J
Understanding Preferences for Lifestyle-Focused Visual Text Messages in Patients With Cardiovascular and Chronic Respiratory Disease: Discrete Choice Experiment
J Med Internet Res 2021;23(9):e26224
URL: https://www.jmir.org/2021/9/e26224
doi:10.2196/26224
PMID:34542413

Original Paper

Determinants of the Use of Health and Fitness Mobile Apps by Patients With Asthma: Secondary Analysis of Observational Studies

Ana Luísa Neves[1,2,3*], MD, MSc, PhD; Cristina Jácome[1,2*], PT, MSc, PhD; Tiago Taveira-Gomes[1,2,4,5], MD, PhD; Ana Margarida Pereira[1,2,6], MD; Rute Almeida[1,2], PhD; Rita Amaral[1,2,7,8], PhD; Magna Alves-Correia[6], MD; Sandra Mendes[2], MSc; Cláudia Chaves-Loureiro[9], MD, PhD; Margarida Valério[9], MD; Cristina Lopes[10,11], MD, PhD; Joana Carvalho[12], MD; Ana Mendes[13], MD; Carmelita Ribeiro[14], MD; Sara Prates[15], MD; José Alberto Ferreira[16], MD; Maria Fernanda Teixeira[17], MD; Joana Branco[18], MD; Marta Santalha[19], MD; Maria João Vasconcelos[20], MD; Carlos Lozoya[21], MD, PhD; Natacha Santos[22], MD; Francisca Cardia[23], MD; Ana Sofia Moreira[24], MD; Luís Taborda-Barata[25,26], MD, PhD; Cláudia Sofia Pinto[27], MD; Rosário Ferreira[28], MD; Pedro Morais Silva[29], MD; Tania Monteiro Ferreira[30], MD; Raquel Câmara[31], MD; Rui Lobo[32], MD; Diana Bordalo[33], MD; Cristina Guimarães[34], MD; Maria Espírito Santo[35], MD; José Ferraz de Oliveira[36], MD; Maria José Cálix Augusto[37], MD; Ricardo Gomes[38], MD; Inês Vieira[39], MD; Sofia da Silva[40], MD; Maria Marques[41], MD; João Cardoso[42], MD, PhD; Ana Morete[6,43], MD; Margarida Aroso[44], MD; Ana Margarida Cruz[45], MD; Carlos Nunes[46], MD; Rita Câmara[47], MD; Natalina Rodrigues[48], MD; Carmo Abreu[49], MD; Ana Luísa Albuquerque[50], MD; Claúdia Vieira[51], MD; Carlos Santos[52], MD; Rosália Páscoa[1,2,53], MD; Carla Chaves-Loureiro[54], MD, MSc; Adelaide Alves[55], MD; Ângela Neves[56], MD; José Varanda Marques[57], MD; Bruno Reis[58], MD; Manuel Ferreira-Magalhães[2,17], MD, PhD; João Almeida Fonseca[1], MD, PhD

[1]Department of Community Medicine, Information and Health Decision Sciences (MEDCIDS), Faculty of Medicine, University of Porto, Porto, Portugal

[2]Center for Health Technology and Services Research (CINTESIS), Faculty of Medicine, University of Porto, Porto, Portugal

[3]Imperial NIHR Patient Safety Translational Research Centre, Institute of Global Health Innovation, Imperial College London, London, United Kingdom

[4]Instituto de Investigação e Formação Avançada em Ciências e Tecnologias da Saúde, Institute of Research and Advanced Training in Health Sciences and Technologies, University Institute of Health Sciences, Cooperativa de Ensino Superior Politécnico e Universitário, CRL, Gandra, Portugal

[5]Faculdade de Ciências da Saúde, Universidade Fernando Pessoa, Porto, Portugal

[6]Allergy Unit, Instituto and Hospital CUF, Porto, Portugal

[7]Department of Cardiovascular and Respiratory Sciences, Porto Health School, Polytechnic Institute of Porto, Porto, Portugal

[8]Department of Women's and Children's Health, Paediatric Research, Uppsala University, Uppsala, Sweden

[9]Serviço Pneumologia, Hospitais da Universidade de Coimbra, Coimbra, Portugal

[10]Unidade de Imunoalergologia, Hospital Pedro Hispano, Unidade Local de Saúde de Matosinhos, Matosinhos, Portugal

[11]Imunologia Básica e Clínica, Faculdade de Medicina, Universidade do Porto, Porto, Portugal

[12]Serviço de Pediatria, Hospital Pedro Hispano, Unidade Local de Saúde de Matosinhos, Matosinhos, Portugal

[13]Serviço de Imunoalergologia, Hospital de Santa Maria, Centro Hospitalar Lisboa Norte, Lisboa, Portugal

[14]Serviço de Imunoalergologia, Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal

[15]Serviço de Imunoalergologia, Hospital de Dona Estefânia, Centro Hospitalar Universitário de Lisboa Central, Lisboa, Portugal

[16]Serviço de Imunoalergologia, Unidade I, Centro Hospitalar Vila Nova de Gaia/Espinho, Vila Nova de Gaia, Portugal

[17]Serviço de Pediatria, Centro Materno Infantil do Norte, Centro Hospitalar Universitário do Porto, Porto, Portugal

[18]Serviço de Pneumologia, Hospital Beatriz Ângelo, Loures, Portugal

[19]Serviço de Pediatria, Hospital da Senhora da Oliveira, Guimarães, Portugal

[20]Serviço de Imunoalergologia, Centro Hospitalar Universitário de São João, Porto, Portugal

[21]Serviço de Imunoalergologia, Hospital Amato Lusitano, Unidade Local de Saúde de Castelo Branco, Castelo Branco, Portugal

[22]Serviço de Imunoalergologia, Centro Hospitalar Universitário do Algarve, Portimão, Portugal

[23]Unidade de Saúde Familiar Terras de Azurara, Agrupamento de Centros de Saúde Dão Lafões, Mangualde, Portugal

[24]Unidade de Imunoalergologia, Hospital do Divino Espirito Santo, Ponta Delgada, Portugal

[25]Department of Allergy & Clinical Immunology, Cova da Beira University Hospital Centre, Covilhã, Portugal

[26]Centro de Investigação em Ciências da Saúde - Health Sciences Research Centre & NuESA –Environment & Health Study Group, Faculty of Health Sciences, University of Beira Interior, Covilhã, Portugal

[27]Serviço de Pneumologia, Hospital São Pedro de Vila Real, Centro Hospitalar de Trás-os-Montes e Alto Douro, Vila Real, Portugal
[28]Departamento de Pediatria, Hospital de Santa Maria, Centro Hospitalar de Lisboa Norte, Lisboa, Portugal
[29]Imunoalergologia, Grupo HPA Saúde, Portimão, Portugal
[30]Unidade de Saúde Familiar Progresso e Saúde, Agrupamento de Centros de Saúde Baixo Mondego, Tocha, Portugal
[31]Serviço de Pneumologia, Hospital Nossa Senhora do Rosário, Centro Hospitalar Barreiro Montijo, Barreiro, Portugal
[32]Unidade de Saúde Familiar João Semana, Agrupamento de Centros de Saúde Baixo Vouga, Ovar, Portugal
[33]Serviço de Pediatria, Unidade Hospitalar de Famalicão, Centro Hospitalar do Médio Ave, Vila Nova de Famalicão, Portugal
[34]Unidade de Saúde Familiar Caminhos do Cértoma, Agrupamento de Centros de Saúde Baixo Mondego, Pampilhosa, Portugal
[35]Unidade de Saúde Familiar Arte Nova, Agrupamento de Centros de Saúde Baixo Vouga, Oliveirinha, Portugal
[36]Imunoalergologia, Hospital Privado de Alfena, Trofa Saúde, Alfena, Portugal
[37]Serviço de Pediatria, Hospital de São Teotónio, Centro Hospitalar Tondela–Viseu, Viseu, Portugal
[38]Serviço de Pneumologia, Hospital Garcia de Orta, Almada, Portugal
[39]Unidade de Cuidados Saúde Personalizados Arnaldo Sampaio, Agrupamento de Centros de Saúde Pinhal Litoral, Leiria, Portugal
[40]Unidade de Saúde Familiar Cuidarte, Unidade Local de Saúde do Alto Minho, Portuzelo, Portugal
[41]Serviço de Imunoalergologia, Centro Hospitalar Universitário do Porto, Porto, Portugal
[42]Serviço de Pneumologia, Hospital Santa Marta, Centro Hospitalar Universitário de Lisboa Central, Lisboa, Portugal
[43]Serviço de Imunoalergologia, Hospital Infante D Pedro, Centro Hospitalar Baixo Vouga, Aveiro, Portugal
[44]Unidade de Saúde Familiar Pedras Rubras, Agrupamento de Centros de Saúde do Grande Porto III - Maia/Valongo, Maia, Portugal
[45]Unidade de Saúde Familiar Bom Porto, Agrupamento de Centros de Saúde do Grande Porto V - Porto Ocidental, Porto, Portugal
[46]Imunoalergologia, Centro de Imunoalergologia do Algarve, Portimão, Portugal
[47]Serviço de Imunoalergologia, Serviço de Saúde da Região Autónoma da Madeira, Funchal, Portugal
[48]Unidade de Saúde Familiar Mondego, Agrupamento de Centros de Saúde Baixo Mondego, Coimbra, Portugal
[49]Serviço de Imunoalergologia, Hospital São Pedro de Vila Real, Centro Hospitalar De Trás-Os-Montes E Alto Douro, Vila Real, Portugal
[50]Unidade de Saúde Familiar Coimbra Centro, Agrupamento de Centros de Saúde Baixo Mondego, Coimbra, Portugal
[51]Unidade de Saúde Familiar Corgo, Agrupamentos de Centros de Saúde Douro I - Marão e Douro Norte, Vila Real, Portugal
[52]Unidade de Saúde Familiar Santo António, Agrupamento de Centros de Saúde do Cávado III - Barcelos/Esposende, Barcelos, Portugal
[53]Unidade de Saúde Familiar Abel Salazar, Agrupamento de Centros de Saúde do Gaia, Vila Nova de Gaia, Portugal
[54]Serviço Pediatria Ambulatória, Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal
[55]Serviço de Pneumologia, Unidade I, Centro Hospitalar Vila Nova de Gaia/Espinho, Vila Nova de Gaia, Portugal
[56]Unidade de Saúde Familiar Araceti, Agrupamento de Centros de Saúde Baixo Mondego, Arazede, Portugal
[57]Unidade de Saúde Familiar Viseu-Cidade, Agrupamento de Centros de Saúde do Dão Lafões, Viseu, Portugal
[58]Unidade de Cuidados Saúde Personalizados Sicó, Agrupamento de Centros de Saúde Pinhal Litoral, Leiria, Portugal
[*]these authors contributed equally

Corresponding Author:
Cristina Jácome, PT, MSc, PhD
Department of Community Medicine, Information and Health Decision Sciences (MEDCIDS)
Faculty of Medicine
University of Porto
Praça de Gomes Teixeira
Porto, 4099-002
Portugal
Phone: 351 225513622
Email: cristinajacome.ft@gmail.com

Abstract

Background: Health and fitness apps have potential benefits to improve self-management and disease control among patients with asthma. However, inconsistent use rates have been reported across studies, regions, and health systems. A better understanding of the characteristics of users and nonusers is critical to design solutions that are effectively integrated in patients' daily lives, and to ensure that these equitably reach out to different groups of patients, thus improving rather than entrenching health inequities.

Objective: This study aimed to evaluate the use of general health and fitness apps by patients with asthma and to identify determinants of usage.

Methods: A secondary analysis of the INSPIRERS observational studies was conducted using data from face-to-face visits. Patients with a diagnosis of asthma were included between November 2017 and August 2020. Individual-level data were collected,

including age, gender, marital status, educational level, health status, presence of anxiety and depression, postcode, socioeconomic level, digital literacy, use of health services, and use of health and fitness apps. Multivariate logistic regression was used to model the probability of being a health and fitness app user. Statistical analysis was performed in R.

Results: A total of 526 patients attended a face-to-face visit in the 49 recruiting centers and 514 had complete data. Most participants were ≤40 years old (66.4%), had at least 10 years of education (57.4%), and were in the 3 higher quintiles of the socioeconomic deprivation index (70.1%). The majority reported an overall good health status (visual analogue scale [VAS] score>70 in 93.1%) and the prevalence of anxiety and depression was 34.3% and 11.9%, respectively. The proportion of participants who reported using health and fitness mobile apps was 41.1% (n=211). Multivariate models revealed that single individuals and those with more than 10 years of education are more likely to use health and fitness mobile apps (adjusted odds ratio [aOR] 2.22, 95%CI 1.05-4.75 and aOR 1.95, 95%CI 1.12-3.45, respectively). Higher digital literacy scores were also associated with higher odds of being a user of health and fitness apps, with participants in the second, third, and fourth quartiles reporting aORs of 6.74 (95%CI 2.90-17.40), 10.30 (95%CI 4.28-27.56), and 11.52 (95%CI 4.78-30.87), respectively. Participants with depression symptoms had lower odds of using health and fitness apps (aOR 0.32, 95%CI 0.12-0.83).

Conclusions: A better understanding of the barriers and enhancers of app use among patients with lower education, lower digital literacy, or depressive symptoms is key to design tailored interventions to ensure a sustained and equitable use of these technologies. Future studies should also assess users' general health-seeking behavior and their interest and concerns specifically about digital tools. These factors may impact both initial engagement and sustained use.

(J Med Internet Res 2021;23(9):e25472) doi:10.2196/25472

KEYWORDS

mobile apps; smartphone; patient participation; self-management; asthma

Introduction

Smart mobile technology has revolutionized how we communicate, share, and consume content, seeping into many different sectors of society, including health care [1]. With the democratization of smartphone use, with 3.8 billion smartphone users worldwide [2], the market of specific apps has experienced a boom. Often free, easy to download, and easy to use, mobile apps have an extensive application in social, educational, and entertainment fields and naturally in the fields of self-management and health behavior change [3]. According to the software application industry, around 500 million smartphone users worldwide were using a health and fitness app in 2015; and by 2018, an estimated 50% of the 3.4 billion smartphone and tablet users, including health care professionals, consumers, and patients, would have downloaded one [4]. The total global mHealth market is predicted to reach the US $100 billion mark in 2021, which constitutes a 5-fold increase from 2016 [5]. In this context, it is hypothesized that apps may become ubiquitous solutions impacting a large number of patients, often capitalizing on gamification strategies and social interaction [6]. In particular, health and fitness apps are a promising approach for improving self-management behaviors in patients with asthma, a prevalent long-term condition with potential social and economic impacts [7,8], which requires a range of self-management skills in everyday life [9]. Indeed, around 1500 mobile apps are targeting patients with asthma in both the Apple App Store and the Google Play Store [10]. A systematic review published by Unni et al [11] suggests that the use of mobile apps by patients with asthma may have benefits across a range of outcomes, including medication adherence and asthma control.

However, the current use of smart devices and apps among patients with asthma remains unexplored, as emphasized by a position paper of the European Academy of Allergy and Clinical Immunology, highlighting the lack of published studies on the use of mHealth in allergic diseases [12]. While there are more than 100 papers published over the last 5 years, they either evaluate the characteristics of specific apps (rather than their use) or focus on the impact of asthma-specific apps.

A recent study reported that smart device ownership levels in patients with asthma are similar to those of the general population, that three-quarters of patients had downloaded/used a general app, yet only one-third had ever used a health and fitness app [13]. A significant variability exists in usage among different racial/ethnic and sociodemographic groups. Nonetheless, this evidence comes mainly from studies conducted in the United States and does not address specific disease contexts [14-18]. The use of health and fitness apps in the asthma context can be explored through the lens of the conceptual model developed by Andersen et al [19], which proposes that the use of health services is driven by three dynamics: predisposing factors (eg, age and gender), enabling factors (eg, socioeconomic level, education, and literacy), and need (eg, clinical characteristics and severity of disease).

The purpose of this study is to evaluate the use of health and fitness apps by patients with asthma and to identify determinants of usage. Specifically, we will investigate the following: (1) the proportion of patients with asthma using health and fitness apps and (2) the relationships among predisposing, enabling and need factors, and using mHealth apps.

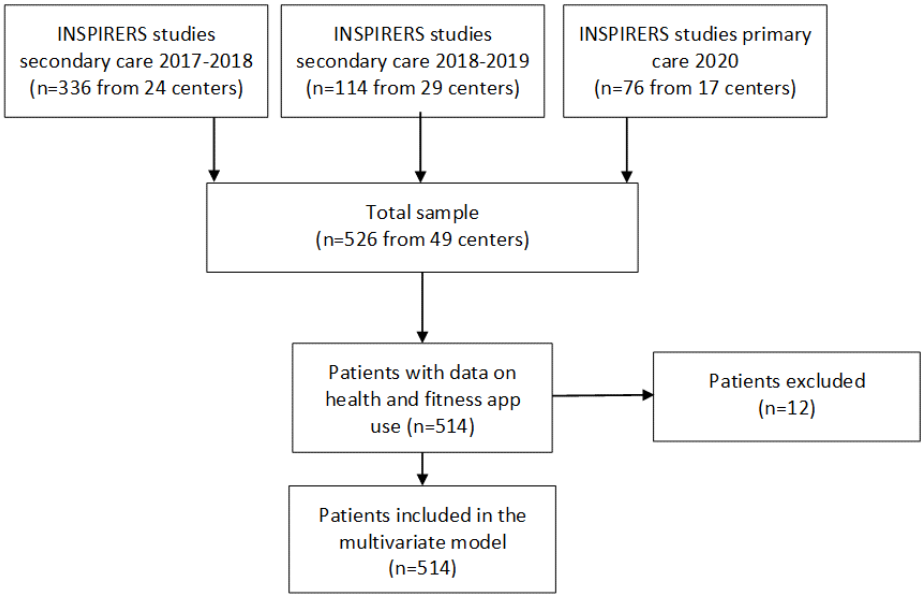
Methods

Study Design

A secondary analysis of INSPIRERS observational studies involving 32 secondary care centers (allergy, pulmonology, and pediatrics departments) and 17 primary care centers in Portugal was performed (Figure 1), as part of the INSPIRERS project. The design of the INSPIRERS observational studies was

disseminated through email contacts, social networks, and oral communications at national meetings/conferences, and physicians/centers interested in being part of the study contacted the research team. A convenience sample of adolescents and adults with persistent asthma was recruited for the INSPIRERS studies between November 2017 and August 2020. Depending on the study, each center was asked to recruit a minimum of 2-10 patients. The 3 INSPIRERS observational studies address the topic of adherence to asthma inhalers among adolescents and adults with persistent asthma (Figure 1). Further details on the project setting and methods have been previously published [20,21]. This study is reported in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines [22].

Figure 1. INSPIRERS studies flowchart.



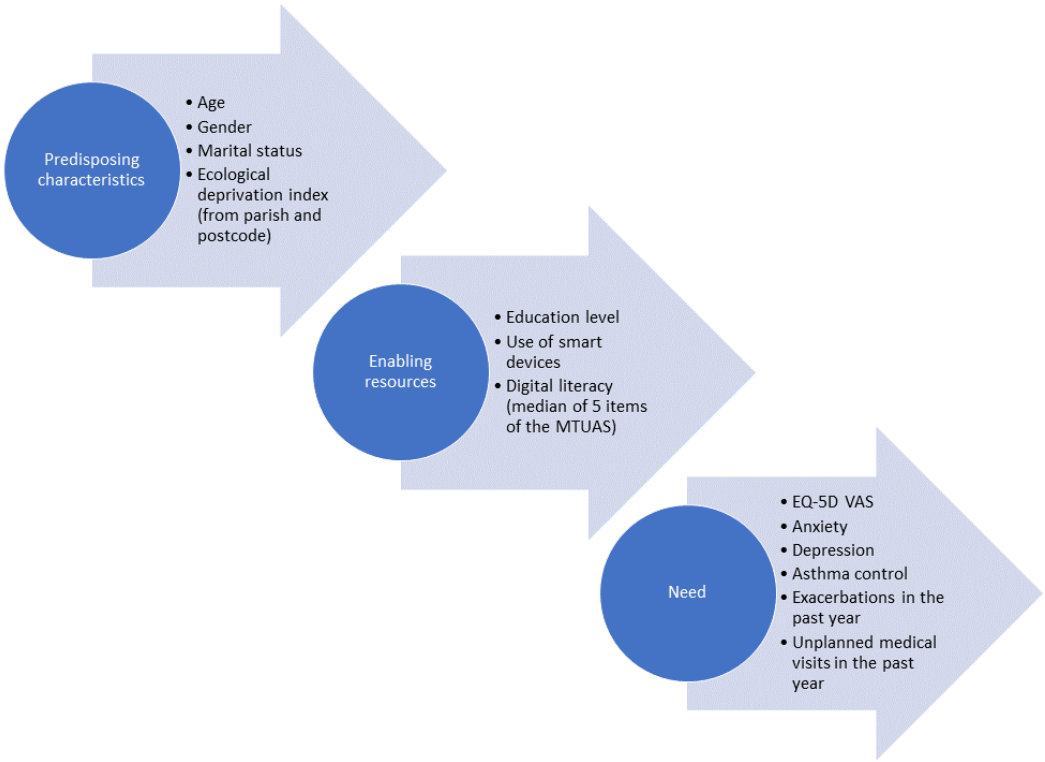
Participants and Data Collection

Data were collected using a questionnaire during a face-to-face visit. The questionnaire had a section to be completed by the physician addressing patients' asthma characteristics and a set of questions to be self-completed by the patient, as detailed below. Patients were included if they had a previous medical diagnosis of persistent asthma, were at least 13 years old, and had an active prescription for an inhaled controller medication for asthma. Patients were excluded if they had a diagnosis of a chronic lung disease other than asthma or a diagnosis of another significant chronic condition with possible interference with the study aims.

Users and nonusers were defined as individuals who answered "yes" or "no," respectively, to the question "Have you ever downloaded and used a health and fitness app?" Health and fitness apps were defined as a range of apps related to personal fitness, workout tracking, diet and nutritional tips, health and safety, etc. In accordance with the conceptual model proposed by Andersen et al [19], variables collected included predisposing factors, enabling factors, and need (Figure 2). Predisposing factors included demographic data (ie, age, gender, marital status, parish, and postcode), and enabling factors included education level, use of smart devices, and digital literacy, both collected from patients. Digital literacy was defined as the median of 5 items of the Media and Technology Usage and Attitudes Scale (MTUAS, ie, use of the GPS, browsing the web, taking pictures, gaming, and checking social networks) rated by frequency of use in a 10-point Likert scale (1=never to 10=all the time) [23]. Additionally, socioeconomic level was explored as an enabling factor, which was defined as the Portuguese ecological deprivation index, extracted from the patient residence information (civil parish/postcode), and categorized into 5 quintiles (Q1=least deprived to Q5=most deprived) [24]. Need variables included smoking status, patients' perceived overall health status (from EQ-5D Visual Analog Scale [VAS], ranging from 0 [worst imaginable health state] to 100 [best imaginable health state]) [25], the presence of anxiety or depression (cut-off≥8 in the Hospital Anxiety and Depression Subscales) [26], and physicians' input, including asthma control level (uncontrolled, partially controlled, or well controlled according to the classification of the Global Initiative for Asthma [27]), number of exacerbations (episodes of progressive increase in shortness of breath, cough, wheezing, or chest tightness, requiring a change in maintenance therapy) in the past year [28], and the number of unplanned appointments in the past year.

Age was categorized into age bands (13-18,18-30, 30-40, 40-50, 50-65, and ≥65 years). Other continuous variables (socioeconomic level, median digital literacy, and overall health status) were categorized into quartiles.

Figure 2. Variables collected included predisposing factors, enabling factors, and need according to the Andersen et al [19]. MTUAS: Media and Technology Usage and Attitudes Scale; VAS: Visual Analog Scale.



Data Analysis

Counts and proportions were calculated for each variable. Multivariate logistic regression with the Enter method was used to ascertain the determinants of being a user of a health and fitness mobile app (dependent variable) in accordance with Andersen et al's [19] conceptual model. Categorical variables, such as gender, age, marital status, educational level, socioeconomic level, use of smart devices, digital literacy, overall health status, smoking status, presence of anxiety or depression, and asthma status were explored as independent variables, irrespective of significance in preliminary univariate logistic regressions. We assessed model fit using the pseudo-R2 Nagelkerke method and tested for evidence for poor model fit using the Hosmer–Lemeshow test. In addition, to assess the increasing contribution of each covariate to the model, we adopted a stepwise approach to build models starting from 1 covariate until including all covariates in the full models. For each model, we computed the described goodness of fit statistics. The quality of the final model was also assessed using the Aikake Information Criterion. Adjusted odds ratios and 95% CIs were calculated. Statistical analyses were conducted using R and the "glm" package. The map of Portugal was created using Paintmaps [29].

Ethics Approval

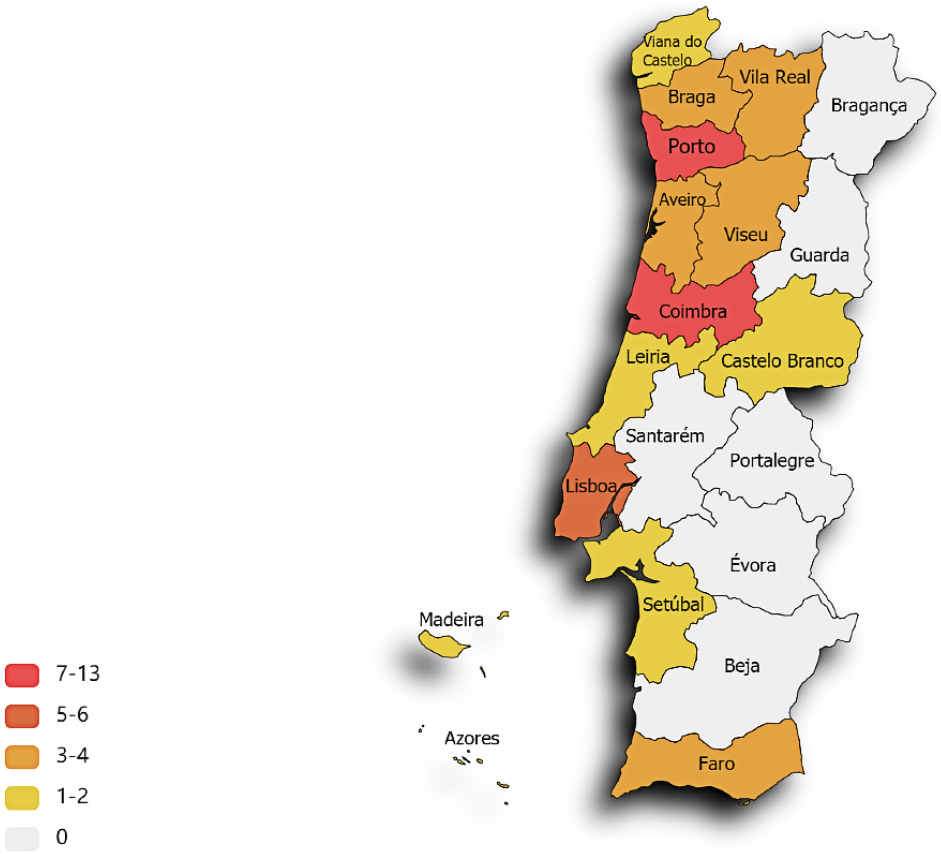
The studies were approved by the ethics committees of all participating centers. The studies were conducted in accordance with the ethical standards established in the Declaration of Helsinki. Eligible patients were approached by physicians during medical visits and invited to participate. Written informed consent was obtained before enrollment. Adult patients signed a consent form; adolescents signed an assent form and a parental consent form was also obtained.

Results

A total of 526 patients attended a face-to-face visit between November 2017 and August 2020 at the 49 recruiting centers. Of those, 12 did not answer the question "Have you ever downloaded and used a health and fitness app?" and were excluded (Figure 1). The recruiting centers included 12 of the 18 Portuguese districts, which represented 9,189,723 inhabitants (89% of the total national population) [30]. A detailed overview of the distribution of the participating centers by district is provided in Figure 3.

The majority of the subjects were ≤40 years old (66.4%, n=341) and 63.4% (n=326) were female. Most participants were single (58.0%, n=298) and had at least 10 years of education (57.4%, n=295). Approximately one-third were in the 2 lower quintiles of the socioeconomic deprivation index (29.9%, n=154). Regarding general health status, as assessed by EQ-5D VAS, 69.8% (n=359) of the participants reported a score of ≥70. Most of the subjects were never smokers (75.5%, n=388). The prevalence of anxiety and depression symptoms in the sample was, respectively, 34.4% (n=177) and 11.9% (n=61).

Figure 3. Number of participating centres per district.



Asthma was well controlled among 51.0% (n=262) of participants. While 50.8% (n=261) of participants had 1 or more asthma exacerbations during the last year, most of the participants did not have any unplanned appointments (66.1%, n=340) or inpatient admissions (94.2%, n=484). The proportion of participants who reported using health and fitness mobile apps was 41.1% (n=211). A full description of the sample, as well as the characteristics of the nonuser and user groups, is provided in Table 1.

Characteristics of both users and nonusers were explored using multivariate logistic regression. The aORs show that single individuals and those with more than 10 years of education are more likely to use health and fitness mobile apps (aOR 2.22, 95%CI 1.05-4.75, and aOR 1.95 95%CI 1.12-3.45, respectively). Higher digital literacy scores were also associated with higher odds of being a user of health and fitness apps, with participants in the second, third, and fourth quartiles showing, respectively, aORs of 6.74 (95%CI 2.90-17.40), 10.30 (95%CI 4.28-27.56), and 11.52 (95%CI 4.78-30.87). Participants with depression symptoms had lower odds of using health and fitness apps (aOR 0.32, 95%CI 0.12-0.83). No significant associations were found with gender, age, socioeconomic level, general health status, smoking status, anxiety, and asthma control (including level of control, number of inpatient admissions, or number of exacerbations). A detailed overview of the multivariate analysis is provided in Table 2.

Table 1. Characteristics of the participants according to their use of health and fitness mobile apps (N=514).

Characteristics	Nonusers (n=303), n (%)	Users (n=211), n (%)	Total, n (%)
Sociodemographic			
Gender			
Female	189 (62.4)	137 (64.9)	326 (63.4)
Male	114 (37.6)	74 (35.1)	188 (36.6)
Age band (years)[a]			
13-18	92 (30.4)	62 (29.4)	154 (30.0)
18-30	41 (13.5)	74 (35.1)	115 (22.4)
30-40	40 (13.2)	32 (15.2)	72 (14.0)
40-50	51 (16.8)	31 (14.7)	82 (16.0)
50-65	51 (16.8)	7 (3.3)	58 (11.3)
≥65	23 (7.6)	3 (1.4)	26 (5.1)
Marital status[b]			
Married	128 (42.2)	49 (23.2)	177 (34.4)
Separated	19 (6.3)	11 (5.2)	30 (5.8)
Single	149 (49.2)	149 (70.6)	298 (58.0)
Widow	6 (2.0)	2 (1.0)	8 (1.6)
Education level (years)			
0-10	159 (52.5)	60 (28.4)	219 (42.6)
>10	144 (47.5)	151 (71.6)	295 (57.4)
Socioeconomic level[c]			
Q1 (least deprived)	32 (10.6)	19 (9.0)	51 (9.9)
Q2	55 (18.2)	48 (22.7)	103 (20.0)
Q3	72 (23.8)	37 (17.5)	109 (21.2)
Q4	72 (23.8)	61 (28.9)	133 (25.9)
Q5 (most deprived)	64 (21.1)	40 (19.0)	104 (20.2)
Digital use and literacy			
Use of smart devices	262 (86.5)	211 (100)	473 (92.0)
Mean digital literacy[d]			
Q1 (0-4.17)	82 (27.1)	10 (4.7)	92 (17.9)
Q2 (4.17-5.67)	76 (25.1)	66 (31.3)	142 (28.6)
Q3 (5.67-6.83)	51 (16.8)	66 (31.3)	117 (22.8)
Q4 (6.83-10.00)	54 (17.8)	69 (32.7)	123 (23.9)
General health status			
Overall health[e]			
Q1 (0-70)	92 (30.4)	54 (25.6)	146 (28.4)
Q2 (70-80)	68 (22.4)	53 (25.1)	121 (23.5)
Q3 (80-90)	87 (28.7)	64 (30.3)	151 (29.4)
Q4 (90-100)	49 (16.2)	38 (18.0)	87 (16.9)
Smoking status[b]			
Never smokers	237 (78.2)	151 (71.6)	388 (75.5)
Former smokers	40 (13.2)	46 (21.8)	86 (16.7)

Characteristics	Nonusers (n=303), n (%)	Users (n=211), n (%)	Total, n (%)
Current smokers	25 (8.3)	14 (6.6)	39 (7.6)
Anxiety symptoms[f]	110 (36.3)	67 (31.8)	177 (34.4)
Depression symptoms[f]	52 (17.2)	9 (4.3)	61 (11.9)
Asthma status			
Asthma control[g]			
Well-controlled	152 (50.2)	110 (52.1)	262 (51.0)
Partially/uncontrolled	150 (49.5)	98 (46.4)	248 (48.2)
≥1 asthma exacerbation in the past year[h]	160 (52.8)	101 (47.9)	261 (50.8)
≥1 unplanned appointment in the past year[c]	105 (34.7)	55 (26.1)	160 (31.1)
≥1 inpatient admission in the past year[c]	11 (3.6)	5 (2.4)	16 (3.1)

[a]7 patients with missing data.
[b]1 patient with missing data.
[c]14 patients with missing data.
[d]40 patients with missing data.
[e]9 patients with missing data.
[f]2 patients with missing data.
[g]4 patients with missing data.
[h]15 patients with missing data.

Table 2. Multivariate analysis to explain the use of health and fitness apps.

Predictor	Odds ratio (5%-95% CI)	*P* value
Gender		
Male	Reference[a]	
Female	1.37 (0.86-2.20)	.19
Age band (years)		
13-18	Reference	
18-30	1.93 (0.97-3.89)	.06
30-40	1.53 (0.59-4.05)	.38
40-50	1.41 (0.53-3.77)	.49
50-65	0.63 (0.12-2.86)	.57
≥65	1.68 (0.19-11.47)	.62
Marital status		
Married	Reference	
Separated	1.64 (0.55-4.96)	.37
Single	2.22 (1.05-4.75)	*.04*[b]
Widow	1.17 (0.03-33.55)	.93
Education level (years)		
0-10	Reference	
>10	1.95 (1.12-3.45)	*.02*
Socioeconomic level		
Q1 (least deprived)	Reference	
Q2	2.02 (0.84-4.92)	.12
Q3	1.08 (0.45-2.64)	.86
Q4	1.42 (0.61-3.33)	.41
Q5 (most deprived)	1.34 (0.56-3.24)	.52
Mean digital literacy		
Q1 (0-4.17)	Reference	
Q2 (4.17-5.67)	6.74 (2.90-17.40)	*<.001*
Q3 (5.67-6.83)	10.30 (4.28-27.56)	*<.001*
Q4 (6.83-10)	11.52 (4.78-30.87)	*<.001*
Overall health		
Q1 (0-70)	Reference	
Q2 (70-80)	1.05 (0.56-1.99)	.87
Q3 (80-90)	0.87 (0.46-1.63)	.67
Q4 (90-100)	0.86 (0.42-1.76)	.69
Smoking status		
Never smokers	Reference	
Former smokers	1.83 (0.97-3.53)	.07
Current smokers	0.84 (0.34-2.05)	.70
Anxiety symptoms		
No	Reference	
Yes	1.12 (0.64-1.95)	.69

Predictor	Odds ratio (5%-95% CI)	*P* value
Depression symptoms		
No	Reference	
Yes	0.32 (0.12-0.83)	*.02*
Asthma control		
Well-controlled	Reference	
Partially/uncontrolled	0.89 (0.54-1.44)	.62
Asthma exacerbation in the past year		
0	Reference	
≥1	0.99 (0.56-1.75)	.96
Unplanned appointment in the past year		
0	Reference	
≥1	1.01 (0.52-1.91)	.99
Inpatient admission in the past year		
0	Reference	
≥1	1.73 (0.37-8.03)	.48

[a]Confidence intervals could not be calculated. Reference means the category used as reference (ie, to which other categories are compared). The Hosmer–Lemeshow Test yielded a *P* value of .21, χ^2_8=10.9, Aikake Information Criterion of 542, and coefficient of determination (R^2) of 48%.

[b]Italicized *P* values are significant.

Discussion

Principal Findings

Use of health and fitness mobile apps was positively associated with a single status, >10 years of education, and higher digital literacy scores, and negatively associated with depressive symptoms. No significant associations were found with other variables, including gender, age, socioeconomic level, general health status, smoking status, anxiety, and asthma control.

Comparison With Previous Studies

According to our results, single participants are more likely to use health and fitness apps. In a recent mixed methods study, Zhou et al [18] explored the barriers to and facilitators of the use of mobile health apps and found that single users had less strong concerns about information security and privacy and less desire to have stringent security protection, which could contribute to higher usage levels in this group.

Consistent with our findings, previous studies have described educational attainment as an important predictor of use of mobile devices and apps [14-16]. However, the relationship may be more complex than initially predicted, as in another study by Carroll et al [14] where both patients with a degree and those with less than high school education were significantly associated with a reduced likelihood of using health apps. The reasons for the educational differences are not fully understood but may reflect the effect of digital skills and confidence, and social norms related to the perceived value of using health and fitness apps [16].

A significant association between digital health literacy and use of health and fitness apps was also found. Digital health literacy has been previously shown to affect the use of health apps [17]. However, comparisons between different studies are limited by the heterogeneity of tools used to evaluate patients' digital health literacy; therefore, standardization of the assessment methods used is recommended in the future. It is also important to note that although individuals with low general health literacy tend to use less health information technology [31], previous evidence has shown that tailored approaches including apps programmed with computer-animated characters, text, and graphics to provide health communication and education could be a widely accepted option for these patients [32].

Interestingly, our study found a negative association between the use of health and fitness apps and the presence of depressive symptoms. Despite the breadth of research exploring the determinants of use of mental health apps [33-35], there was a lack of evidence specifically exploring the impact of the presence of mental health symptoms on usage rates. However, several studies show that patients with depressive symptoms often exhibit poor engagement with health services, and health care avoidance [36,37], and such behaviors may also contribute to a lower interest and usage of health and fitness apps by these groups of patients. This finding opens a new research avenue, and future work should explore the drivers for this effect and should involve patients with mental health problems in the co-design of digital solutions and interventions that promote both early and sustained use. The other needed variables, such as the ones related to asthma control, were not significant in explaining mHealth behaviors, although they showed a similar trend to the one observed with the variable depression. This is in line with previous asthma mHealth studies showing that patients with worse asthma control engage less with the tested apps [15,38]. Regarding age, odds ratios were higher for patients

aged 18-30 years (1.93, 95%CI 0.97-3.89), but no significant differences were detected, despite the *P* value being close to the significance threshold (*P*=.06). Increasing the sample size in future studies could help further explore this effect and to confirm or exclude a potential type II error [39]. Previous studies found that younger adults were more likely to engage with health apps [14-16] and suggested that the effect of age likely reflects both social norms and cohort effects, such as the increased exposure to these devices and apps at younger ages [14]. Previous evidence also suggests that younger adults seem to have higher digital health literacy levels, which can contribute to increased use of digital solutions [40]. While it is recognized that the use of digital health solutions by older persons could improve patient engagement and reduce both financial burden and pressure on health systems, usage rates among this group remain low. According to a mixed methods study conducted by Fox et al, this digital health divide is deepening owing to older adults' perceived inability and unwillingness to use digital technologies, stemming from mistrust, high-risk perceptions, and a strong desire for privacy [41].

Finally, no significant associations were found with gender. Previous literature shows mixed evidence on this subject: while some studies had found a higher use among male subjects [15], others reported the opposite [14,16]. The reasons for gender differences in some samples are unclear but may reflect sample-specific differences in health-seeking behavior, and interest and participation in healthy lifestyle interventions in general.

Strengths and Limitations

This study has several strengths. It is the first national-level study performed in Portugal, evaluating the use of health and fitness apps among patients with asthma, covering the majority of the geographic regions of the country. A comprehensive set of individual-level characteristics was collected and analyzed, which allowed us to explore the impact of a range of sociodemographic factors, health literacy, and cofactors such as general health status and asthma status. Although no power calculation was performed (which was associated with the secondary analysis nature of this study), the overall large sample size contributes to the robustness of these findings.

Other limitations also need to be acknowledged. Intrinsic to the study design using convenience sampling, a potential selection bias cannot be excluded. This can possibly explain the low number of patients above 65 years of age. Nevertheless, this risk was mitigated by sampling patients from different health care settings, centers, and geographic regions. Health and fitness app use was patient-reported; therefore, a potential information bias cannot be excluded either. As an alternative, future studies could use patient log-in as a measure of app use. Furthermore, digital literacy was assessed using selected items of MTUAS scale, not the complete instrument. This choice emerged as a mitigation measure to reduce the data collection burden and allow us to efficiently collect data on an aspect seldom reported in the literature. Future studies should explore the possibility of including complete validated tools, such as MTUAS [23] or the eHealth Literacy Scale [42]. We also need to consider that although the questionnaires were independently answered by the study participants, they were provided to them by their family physicians provided during an in-person visit, which may have influenced them to give desirable responses, namely those related to their general health status and asthma control.

Finally, most subjects included in this study were relatively young, had at least 7 years of education, were in the 3 higher quintiles of the socioeconomic deprivation index, had an overall good health status, and had good asthma control. Consequently, attempts to generalize these findings to other populations that do not share the same characteristics need to be cautious. Future studies should consider involving patients with specific needs (ie, lower education attainment, lower sociodemographic levels, poorer overall health status, or poorer asthma status control) and evaluate the replicability of these findings.

Conclusions

Our results show a negative association with low literacy and the presence of depressive symptoms, highlighting the need for future research to explore the effect of disparities on app use, particularly among those with lower digital health literacy or mental health issues. Importantly, such studies should also assess participants' general health-seeking behavior (ie, interest and participation in healthy lifestyle interventions in general) and the key constructs of their acceptance the unified theory of acceptance and use of technology) [43,44]. This theory was developed through a consolidation of the constructs of 8 previous models (theory of reasoned action, technology acceptance model, motivational model, theory of planned behavior, a combined theory of planned behavior/technology acceptance model, model of personal computer use, diffusion of innovations theory, and social cognitive theory), and includes four constructs: performance expectancy, effort expectancy, social influence, and facilitating conditions. Gender, age, experience, and voluntariness of use have been suggested to moderate the impact of the 4 key constructs on usage intention and behavior [43,44].

A better understanding of which patients with asthma are (and are not) using general health and fitness apps is key to design tailored mHealth interventions to improve sustained use by this specific group of patients. Additionally, this knowledge can also inform high-level delivery strategies to ensure that these solutions reach out comprehensive groups of patients with asthma, thus improving rather than entrenching health inequities within this population.

Acknowledgments

This study was funded by the European Regional Development Fund through the following programs: POCI-01-0145-36 FEDER-029130 ("mINSPIRE—mHealth to measure and improve adherence to medication in chronic obstructive respiratory diseases - generalisation and evaluation of gamification, peer support and advanced image processing technologies") cofunded

by the COMPETE2020 (Programa Operacional Competitividade e Internacionalização), Portugal 2020 and by Portuguese Funds through Fundação para a Ciência e a Tecnologia. ALN is funded by the National Institute for Health Research's Imperial Patient Safety Translational Research Centre and supported by the National Institute of Health Research's Imperial Biomedical Research Centre. This study was supported by National Funds through Fundação para a Ciência e a Tecnologia, IP, within the research and development unit of CINTESIS (reference UIDB/4255/2020).

Conflicts of Interest

JAF is the co-founder of the University of Porto Spin-off Medida that is co-responsible for the Inspirermundi app. The other authors have no conflicts to declare.

References

1. Bert F, Giacometti M, Gualano MR, Siliquini R. Smartphones and health promotion: a review of the evidence. J Med Syst 2014 Jan;38(1):9995. [doi: 10.1007/s10916-013-9995-7] [Medline: 24346929]
2. Number of smartphone users from 2016 to 2021. Statista. 2021 Jun. URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2021-06-29]
3. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
4. Bhuyan SS, Lu N, Chandak A, Kim H, Wyant D, Bhatt J, et al. Use of Mobile Health Applications for Health-Seeking Behavior Among US Adults. J Med Syst 2016 Jun;40(6):153. [doi: 10.1007/s10916-016-0492-7] [Medline: 27147516]
5. Total global mHealth market forecast from 2016 to 2025. Statista. 2018 Nov. URL: https://www.statista.com/statistics/938544/mhealth-market-size-forecast-globally/ [accessed 2021-06-29]
6. Alvarez-Perea A, Sánchez-García S, Muñoz Cano R, Antolín-Amérigo D, Tsilochristou O, Stukus DR. Impact Of "eHealth" in Allergic Diseases and Allergic Patients. J Investig Allergol Clin Immunol 2019;29(2):94-102 [FREE Full text] [doi: 10.18176/jiaci.0354] [Medline: 30457105]
7. Partridge MR, Dal Negro RW, Olivieri D. Understanding patients with asthma and COPD: insights from a European study. Prim Care Respir J 2011 Sep;20(3):315-323, 17 p following 323 [FREE Full text] [doi: 10.4104/pcrj.2011.00056] [Medline: 21660394]
8. Masoli M, Fabian D, Holt S, Beasley R, Global Initiative for Asthma (GINA) Program. The global burden of asthma: executive summary of the GINA Dissemination Committee report. Allergy 2004 May;59(5):469-478. [doi: 10.1111/j.1398-9995.2004.00526.x] [Medline: 15080825]
9. British Guideline on the Management of Asthma: Quick Reference Guide. British Thoracic Society, Scottish Intercollegiate Guidelines Network. 2008. URL: http://globalasthmanetwork.org/management/guides/bosnia_and_herzegovina/British%20guideline%20for%20asthma.pdf [accessed 2021-08-24]
10. Nikolova S. Why do asthma apps only capture less than 1% of the addressable market? (Top 10 asthma apps). Research2Guidance. 2018. URL: https://research2guidance.com/why-do-asthma-apps-only-capture-less-than-1-of-the-addressable-market-top-10-asthma-apps/ [accessed 2021-06-29]
11. Unni E, Gabriel S, Ariely R. A review of the use and effectiveness of digital health technologies in patients with asthma. Ann Allergy Asthma Immunol 2018 Dec;121(6):680-691.e1. [doi: 10.1016/j.anai.2018.10.016] [Medline: 30352288]
12. Matricardi PM, Dramburg S, Alvarez-Perea A, Antolín-Amérigo D, Apfelbacher C, Atanaskovic-Markovic M, et al. The role of mobile health technologies in allergy care: An EAACI position paper. Allergy 2020 Feb;75(2):259-272. [doi: 10.1111/all.13953] [Medline: 31230373]
13. Jácome C, Almeida R, Pereira AM, Araújo L, Correia MA, Pereira M, INSPIRERS group. Asthma App Use and Interest Among Patients With Asthma: A Multicenter Study. J Investig Allergol Clin Immunol 2020 Apr;30(2):137-140 [FREE Full text] [doi: 10.18176/jiaci.0456] [Medline: 32327403]
14. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who Uses Mobile Phone Health Apps and Does Use Matter? A Secondary Data Analytics Approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
15. Chan YY, Wang P, Rogers L, Tignor N, Zweig M, Hershman SG, et al. The Asthma Mobile Health Study, a large-scale clinical observational study using ResearchKit. Nat Biotechnol 2017 Apr;35(4):354-362 [FREE Full text] [doi: 10.1038/nbt.3826] [Medline: 28288104]
16. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
17. Lin TTC, Bautista JR. Understanding the Relationships between mHealth Apps' Characteristics, Trialability, and mHealth Literacy. J Health Commun 2017 Apr;22(4):346-354. [doi: 10.1080/10810730.2017.1296508] [Medline: 28323546]

18. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and Facilitators of the Use of Mobile Health Apps From a Security Perspective: Mixed-Methods Study. JMIR Mhealth Uhealth 2019 Apr 16;7(4):e11223 [FREE Full text] [doi: 10.2196/11223] [Medline: 30990458]
19. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? J Health Soc Behav 1995 Mar;36(1):1-10. [doi: 10.2307/2137284] [Medline: 7738325]
20. Jácome C, Guedes R, Almeida R, Teixeira JF, Pinho B, Vieira-Marques P, pelo grupo INSPIRERS. mINSPIRERS – Estudo da exequibilidade de uma aplicação móvel para medição e melhoria da adesão à medicação inalada de controlo em adolescentes e adultos com asma persistente: Protocolo de um estudo observacional multicêntrico. Rev Port Imunoalergologia 2018;26(1):47-61 [FREE Full text]
21. Jácome C, Pereira AM, Almeida R, Ferreira-Magalhaes M, Couto M, Araujo L, Inspirers group. Patient-physician discordance in assessment of adherence to inhaled controller medication: a cross-sectional analysis of two cohorts. BMJ Open 2019 Nov 07;9(11):e031732 [FREE Full text] [doi: 10.1136/bmjopen-2019-031732] [Medline: 31699737]
22. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Int J Surg 2014 Dec;12(12):1500-1524 [FREE Full text] [doi: 10.1016/j.ijsu.2014.07.014] [Medline: 25046751]
23. Rosen LD, Whaling K, Carrier LM, Cheever NA, Rokkum J. The Media and Technology Usage and Attitudes Scale: An empirical investigation. Comput Human Behav 2013 Nov 01;29(6):2501-2511 [FREE Full text] [doi: 10.1016/j.chb.2013.06.006] [Medline: 25722534]
24. Ribeiro AI, Mayer A, Miranda A, Pina MF. The Portuguese Version of the European Deprivation Index: An Instrument to Study Health Inequalities. Acta Med Port 2017 Jan 31;30(1):17-25 [FREE Full text] [doi: 10.20344/amp.7387] [Medline: 28501033]
25. Pickard AS, Wilke C, Jung E, Patel S, Stavem K, Lee TA. Use of a preference-based measure of health (EQ-5D) in COPD and asthma. Respir Med 2008 Apr;102(4):519-536 [FREE Full text] [doi: 10.1016/j.rmed.2007.11.016] [Medline: 18180151]
26. Pais-Ribeiro J, Silva I, Ferreira T, Martins A, Meneses R, Baltar M. Validation study of a Portuguese version of the Hospital Anxiety and Depression Scale. Psychol Health Med 2007 Mar;12(2):225-235; quiz 235. [doi: 10.1080/13548500500524088] [Medline: 17365902]
27. 2020 Gold Reports. Global Initiative for Chronic Obstructive Lung Disease. URL: https://goldcopd.org/gold-reports/ [accessed 2021-06-29]
28. Loymans RJB, Ter Riet G, Sterk PJ. Definitions of asthma exacerbations. Curr Opin Allergy Clin Immunol 2011 Jun;11(3):181-186. [doi: 10.1097/ACI.0b013e3283466f45] [Medline: 21467926]
29. Paintmaps. URL: https://paintmaps.com/ [accessed 2021-08-24]
30. População residente: total e por grandes grupos etários. Pordata.: INE URL: https://www.pordata.pt/Municipios/Popula%C3%A7%C3%A3o+residente+total+e+por+grandes+grupos+et%C3%A1rios-390 [accessed 2020-11-10]
31. Herndon JB, Chaney M, Carden D. Health literacy and emergency department outcomes: a systematic review. Ann Emerg Med 2011 Apr;57(4):334-345. [doi: 10.1016/j.annemergmed.2010.08.035] [Medline: 21035902]
32. Kim H, Xie B. Health literacy and internet‐ and mobile app‐based health services: A systematic review of the literature. Proc Assoc Info Sci Tech 2016 Feb 24;52(1):1-4. [doi: 10.1002/pra2.2015.145052010075]
33. Lipschitz J, Miller CJ, Hogan TP, Burdick KE, Lippin-Foster R, Simon SR, et al. Adoption of Mobile Apps for Depression and Anxiety: Cross-Sectional Survey Study on Patient Interest and Barriers to Engagement. JMIR Ment Health 2019 Jan 25;6(1):e11334 [FREE Full text] [doi: 10.2196/11334] [Medline: 30681968]
34. Huang H, Bashir M. Users' Adoption of Mental Health Apps: Examining the Impact of Information Cues. JMIR Mhealth Uhealth 2017 Jun 28;5(6):e83 [FREE Full text] [doi: 10.2196/mhealth.6827] [Medline: 28659256]
35. Williams MG, Stott R, Bromwich N, Oblak SK, Espie CA, Rose JB. Determinants of and barriers to adoption of digital therapeutics for mental health at scale in the NHS. BMJ Innov 2020 May 07;6(3):92-98. [doi: 10.1136/bmjinnov-2019-000384]
36. Lo C, Calzavara A, Kurdyak P, Barbera L, Shepherd F, Zimmermann C, et al. Depression and use of health care services in patients with advanced cancer. Can Fam Physician 2013 Mar;59(3):e168-a174 [FREE Full text] [Medline: 23486819]
37. Ganson KT, Weiser SD, Tsai AC, Nagata JM. Associations between Anxiety and Depression Symptoms and Medical Care Avoidance during COVID-19. J Gen Intern Med 2020 Nov;35(11):3406-3408 [FREE Full text] [doi: 10.1007/s11606-020-06156-8] [Medline: 32875507]
38. Jácome C, Almeida R, Pereira AM, Amaral R, Mendes S, Alves-Correia M, et al. Feasibility and Acceptability of an Asthma App to Monitor Medication Adherence: Mixed Methods Study. JMIR Mhealth Uhealth 2021 May 25;9(5):e26442 [FREE Full text] [doi: 10.2196/26442] [Medline: 34032576]
39. Rothman KJ. Curbing type I and type II errors. Eur J Epidemiol 2010 Apr;25(4):223-224 [FREE Full text] [doi: 10.1007/s10654-010-9437-5] [Medline: 20232112]
40. Knitza J, Simon D, Lambrecht A, Raab C, Tascilar K, Hagen M, et al. Mobile Health Usage, Preferences, Barriers, and eHealth Literacy in Rheumatology: Patient Survey Study. JMIR Mhealth Uhealth 2020 Aug 12;8(8):e19661 [FREE Full text] [doi: 10.2196/19661] [Medline: 32678796]
41. Fox G, Connolly R. Mobile health technology adoption across generations: Narrowing the digital divide. Info Systems J 2018 Jan 29;28(6):995-1019. [doi: 10.1111/isj.12179]

42. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
43. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
44. Venkatesh V, Thong JYL, Xu X. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly 2012;36(1):157. [doi: 10.2307/41410412]

Abbreviations

aOR: adjusted odds ratio
MTUAS: Media and Technology Usage and Attitudes Scale
VAS: Visual Analog Scale

Edited by R Kukafka; submitted 13.11.20; peer-reviewed by C Mueller, O El-Gayar, J Alqahtani, S Rostam Niakan Kalhori; comments to author 29.01.21; revised version received 01.04.21; accepted 14.06.21; published 22.09.21.

Please cite as:
Neves AL, Jácome C, Taveira-Gomes T, Pereira AM, Almeida R, Amaral R, Alves-Correia M, Mendes S, Chaves-Loureiro C, Valério M, Lopes C, Carvalho J, Mendes A, Ribeiro C, Prates S, Ferreira JA, Teixeira MF, Branco J, Santalha M, Vasconcelos MJ, Lozoya C, Santos N, Cardia F, Moreira AS, Taborda-Barata L, Pinto CS, Ferreira R, Morais Silva P, Monteiro Ferreira T, Câmara R, Lobo R, Bordalo D, Guimarães C, Espírito Santo M, Ferraz de Oliveira J, Cálix Augusto MJ, Gomes R, Vieira I, da Silva S, Marques M, Cardoso J, Morete A, Aroso M, Cruz AM, Nunes C, Câmara R, Rodrigues N, Abreu C, Albuquerque AL, Vieira C, Santos C, Páscoa R, Chaves-Loureiro C, Alves A, Neves Â, Varanda Marques J, Reis B, Ferreira-Magalhães M, Almeida Fonseca J
Determinants of the Use of Health and Fitness Mobile Apps by Patients With Asthma: Secondary Analysis of Observational Studies
J Med Internet Res 2021;23(9):e25472
URL: https://www.jmir.org/2021/9/e25472
doi:10.2196/25472
PMID:34550077

Original Paper

Attitudes Toward the Environment and Use of Information and Communication Technologies to Address Environmental Health Risks in Marginalized Communities: Prospective Cohort Study

Jose G Perez-Ramos[1], PhD; Scott McIntosh[2], PhD; Emily S Barrett[3], PhD; Carmen M Velez Vega[4], PhD; Timothy D Dye[1], PhD

[1]Department of Obstetrics and Gynecology, School of Medicine and Dentistry, University of Rochester, Rochester, NY, United States

[2]Department of Public Health Sciences, School of Medicine and Dentistry, University of Rochester, Rochester, NY, United States

[3]Environmental and Occupational Health Sciences Institute, Rutgers School of Public Health, Rutgers University, Piscataway Township, NJ, United States

[4]Escuela Graduada de Salud Pública, Recinto de Ciencias Médicas, Universidad de Puerto Rico, San Juan, Puerto Rico

Corresponding Author:
Jose G Perez-Ramos, PhD
Department of Obstetrics and Gynecology
School of Medicine and Dentistry
University of Rochester
601 Elmwood Ave
Box 668
Rochester, NY, 14622
United States
Phone: 1 5852768755
Email: j.perezramos@rochester.edu

Abstract

Background: Information and communication technologies, including mobile health (mHealth), can help isolated communities address environmental health challenges. The Puerto Rican island of Culebra has faced multiple sociopolitical and economic factors that have distressed the island's environment and health. *Culebrenses* are technologically engaged and have demonstrated a use of technology that transcends socioeconomic barriers. As a result, technological interventions could potentially help manage environmental risks on the island.

Objective: This study aims to test and evaluate the potential benefits of an mHealth tool, termed ¡mZAP! (*Zonas, Acción y Protección*), for engaging communities with environmental risks through technology.

Methods: Participants using ¡mZAP! (N=111) were surveyed. Bivariate analyses were used to examine associations of mHealth use with sociodemographics, technology use, an adapted environmental attitudes inventory, and the multidimensional health locus of control. Logistic regression was used to examine associations between attitudes toward environmental health risks and mHealth use.

Results: Higher positive attitudes toward the environment were significantly associated with the use of ¡mZAP! (odds ratio 5.3, 95% CI 1.6-17.0). Environmental attitudes were also associated with the multidimensional health locus of control *powerful others* subscale (*P*=.02), indicating that attitudes toward the environment become more negative as feelings controlled by others increase. Participants felt that the authorities would resolve the challenges (63/111, 56.7%).

Conclusions: Perceived lack of control could present barriers to collective actions to address salient environmental health challenges in communities. The ongoing dependency on government-based solutions to community problems is worrisome, especially after the hurricane experiences of 2017 (which may potentially continue to be an issue subsequent to the more recent 2020 earthquakes).

(J Med Internet Res 2021;23(9):e24671) doi:10.2196/24671

KEYWORDS

community engagement; environmental health risk; epidemiology; ICT; mHealth; mobile phone; Puerto Rico

Introduction

Background

Information and communication technologies (ICTs), including mobile health (mHealth), can help isolated communities address environmental health challenges [1,2]. The World Health Organization considers environmental risks a public health priority [3]. Social and ecological determinants of health, along with environmental health challenges, often exacerbate pre-existing health disparities [4]. Traditional health paradigms have been changed by mHealth tools to address such health disparities. For example, a study used a mobile app termed *FAITH!* (Fostering African American Improvement in Total Health) to improve cardiovascular health among an African American church community (n=86) in the United States following a multiphase community-based participatory research design [5]. Other examples include using a web-based intervention for smoking cessation among low socioeconomic status populations in the United States (n=1440) [6] and an mHealth tool to reduce health care access gaps among marginalized rural populations in Laos (n=983) and Thailand (n=1158) [7].

The benefits of using ICTs to address environmental risks following crowdsourcing approaches have been successfully implemented in the United States using the mobile phones of users [8,9]. Mobile phones are also being used to engage populations in new and innovative learning modalities [10,11]. Previous studies have shown that mobile tools such as cell phones can help users develop positive attitudes toward the environment [12,13]. Mobile tools facilitate engaging with users by using their devices (eg, smartphones and smartwatches) as part of the solution to existing environmental challenges [8,13].

Although the use of technology has transcended socioeconomics barriers, providing digital access to different levels of society, a significant gap perpetuates digital access disparities among marginalized communities worldwide [14-16].

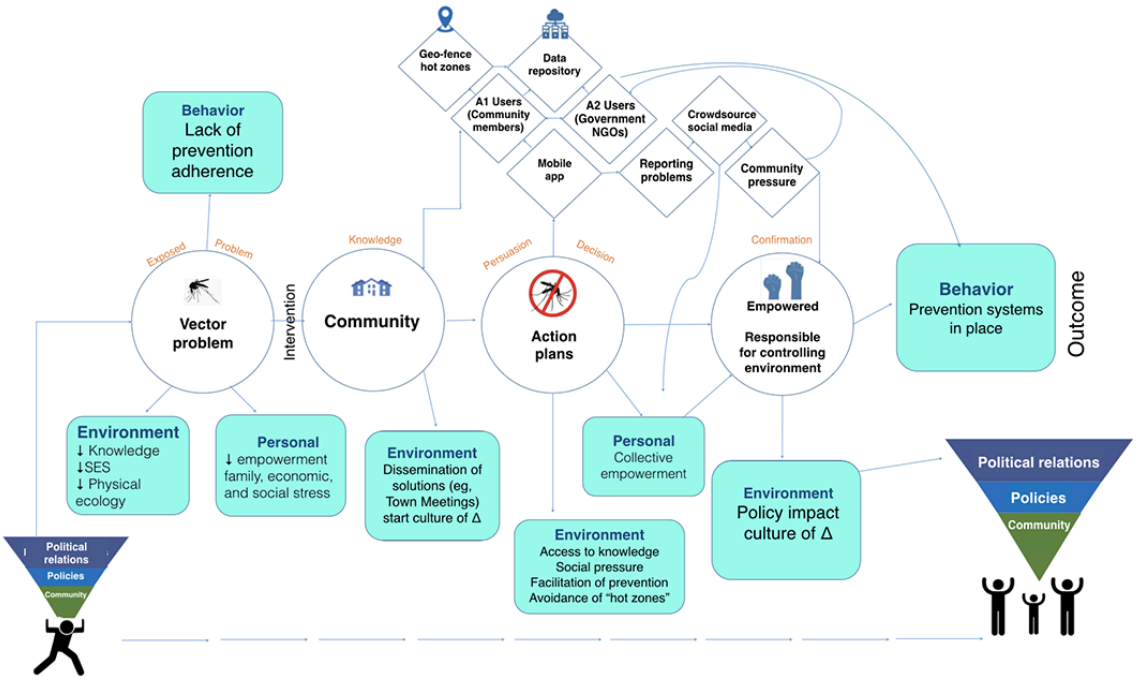
The power of being interconnected via modern technology (eg, social media, cell phones, and the internet) has been shown to (1) catalyze sociopolitical changes [17,18], (2) close the gap in physical distances [19], (3) foster the development of health advances [20], and (4) provide access to information [21]. In contrast, the use of technology has also been linked to the spread of misinformation, resulting in the exacerbation of social injustices among marginalized groups. For example, WhatsApp was used in Maceió, Brazil, to diffuse conspiracy theories about vaccines as the cause of microcephaly in children testing positive for the Zika virus [22,23]. Thus, communities often have untapped and unrecognized resources that may be mobilized to work with the government sector, health programs, and nonprofit organizations to address risks in their environments [24].

Many communities, particularly islands, are dependent on ICT to reduce the impact of their geographical and societal isolation [25]. Island communities face many social and health challenges that are not necessarily faced by other communities, which can intensify risk [26-28]. The remote island of Culebra, off the eastern coast of Puerto Rico's *Big Island* (*Isla Grande*), is geographically and systemically isolated from the rest of Puerto Rico. Culebra's isolation also results in limited political power because of the small population, leading to a system of dependency that negatively impacts health access, utilities, waste management, transportation, water safety, and air and soil contamination, which is common in rural and isolated areas [29-33].

Hurricanes Irma and María in 2017 demonstrated this dependency, with Culebra losing connection to the national electrical grid for more than a year [34] and having difficulties in resuming ferry and other island services. Although impacted by these challenges, *Culebrenses* have a strong history of collective engagement against social injustices, such as facilitating the US Navy's exit from the island in 1975 and protesting against maritime transportation policies [35]. In addition to this community engagement, *Culebrenses* are technologically engaged and have demonstrated a use of technology that transcends socioeconomic barriers [36]. As a result, technological interventions could potentially help manage environmental risks on the island.

This study was guided by a multi-theoretical framework (social cognitive theory [SCT] and the diffusion of innovations theory) which, along with community-centered design methodology, can (1) provide a model for behavioral changes and (2) guide the process to understand the adoption of this mHealth tool better [37-39]. This study also followed a human-centered design framework that provides an inclusion process whereby users are central and active agents necessary for the conception, testing, implementation, and evaluation of solutions to Culebra's environmental health risk challenges [40]. The SCT addresses behavior change and the effect of the community members' environment [41]. The SCT also helps identify community problems and priorities, allowing for an in-depth understanding of personal factors, current community behaviors, and environmental influences associated with environmental health risks that can affect community-wide behavioral change. The diffusion of innovations theory facilitates the process of development and sustainability [42]. Owing to the complexity of these two theories, a theoretical framework model titled *Community-Centered Environmental Health Risk Control Model* was developed for this study (Figure 1).

Figure 1. Community-centered environmental health risk control model. NGO: nongovernmental organization; SES: socioeconomic status.



Objective

Following this theoretically informed model facilitated the work as part of a community-driven research project to collaborate with community partners to implement a tailored, community-centered, and crowdsourced ICT termed ¡mZAP! (*Zonas, Acción y Protección*) as part of a community-driven approach to assess the ability of ICTs to engage islanders in collective environmental actions. Specifically, in this study, we evaluate the ability of psychosocial, environmental attitudes, and technological variables to predict the use of the ¡mZAP! app.

Methods

Overview

We used a prospective design to assess the ¡mZAP! app use. The study surveyed community participants at baseline to capture their perceptions and attitudes toward the environment, access to and use of technology, community actions toward environmental challenges, knowledge about vector-borne diseases, and health control in Culebra, Puerto Rico. App registrants were followed prospectively for 3 months (November 2018-February 2019) to assess the use of ¡mZAP!

The Archipelago of Culebra

Culebra is an archipelago municipality located on the east coast of Puerto Rico, 17 miles from *Isla Grande* (*Big Island*—as it is commonly referred to in Puerto Rico by the residents of outer islands; Figure 2). On the basis of the 2018 US Census annual estimates of the resident population, Culebra has an estimated population of 1500 [43]. Culebra's geographical location provides a unique, biodiverse ecosystem of drylands, mangroves, lagoons, white sand beaches, and coral reefs, all in almost 12 square miles [35,44]. Largely owing to this biodiversity and geographical location, tourism is Culebra's primary economy.

Culebra also presents another face; for many decades, this set of islands has experienced marginalization by the United States and from the local government [33,35], including military bombing drills that left unknown contamination and pollution effects [29-31], unreliable maritime transportation [45], environmental hazards [46], high presence of vector agents (mosquitoes) [35], and the lack of a proper health care facility [32,47].

Figure 2. Map of Culebra in relation to the rest of Puerto Rico.

¡mZAP! App

The ¡mZAP! app was designed and developed as a community-oriented ICT to stimulate collective action to mitigate environmental health risks (eg, trash and mosquitoes). This study design involved direct input from community organizations and community members in Culebra, including women's advocacy organizations, sports organizations, private businesses, health organizations, and schools, primarily through focus group and qualitative engagement processes, as previously described [48].

In this app, users may take actions to identify, report, and address community-based priorities such as abandoned structures, trash, stray animals, stagnant water, mosquito breeding grounds, and pests. Although not directly relevant to environmental issues, a feature to track ferry and ship locations was included to further incentivize app use as the community had identified maritime transportation as a major challenge.

Survey

A survey was deployed at baseline using the REDCap (Research Electronic Data Capture, Vanderbilt University) platform version 9.0, which also hosted the survey data. As defined in previous studies, the REDCap database is a secured, encrypted, and Health Insurance Portability and Accountability Act–compliant data capture application developed for large-scale research projects [49].

The survey contained adaptations of two previously validated instruments and original items self-administered by participants via the mobile app tool. The survey aimed to better understand users' attitudes toward the environment, community technology use, and community challenges. Questions related to community health, demographic characteristics, the environment, attitudes toward community actions, and technology use were included. Consistent with recommendations for pretesting and pilot-testing of surveys evaluating mHealth interventions [49,50], the survey was pretested with approximately 10 iterations with feedback from 4 bilingual Latinos from the author's institution and volunteers from the target community.

¡mZAP! Databases

The ¡mZAP! app was built as a cross-platform mobile and web-based app using the *Ionic* framework by an app developer that worked as part of the ¡mZAP! core team at the University of Rochester. *Ionic* is an open-source toolkit that facilitates the building of mobile apps and web-based applications using a single coding structure (eg, HTML, Cascading Style Sheets, and JavaScript), eliminating the need to create a new coding process for each mobile and web-based platform. An integration of *Angular*, a Google-maintained platform, was used to build the app across mobile and web-based systems (eg, iOS and Android). *Firebase* software was used as a database repository. *Firebase* is a Google-based software with the capacity to secure file uploads and downloads, including images, videos, and other user-generated data under Google encrypted (AES256 and AES128) and cloud storage. The ¡mZAP! mHealth tool back-end process was built to assign each user a unique ID linking all the data generated by users while using the app. This process facilitated a better understanding of ¡mZAP! users' behaviors. An application programming interface was used to access the REDCap survey directly from the app. Unique user identifications and email addresses were used to merge the two databases (¡mZAP! use and the REDCap survey). A weekly data transfer stored participants' app use to secure servers at the lead author's institution.

Sample Size Calculation

This study was powered to detect an effect size of 3.0 (α=.05 and β=.20) of app use by the environmental attitudes' category. We estimated that 50% of participants with positive attitudes toward the environment would use the app and 25% of participants with negative attitudes would use ¡mZAP! On the basis of these calculations, the sample size required for this study was 116 participants.

Participants and Study Implementation

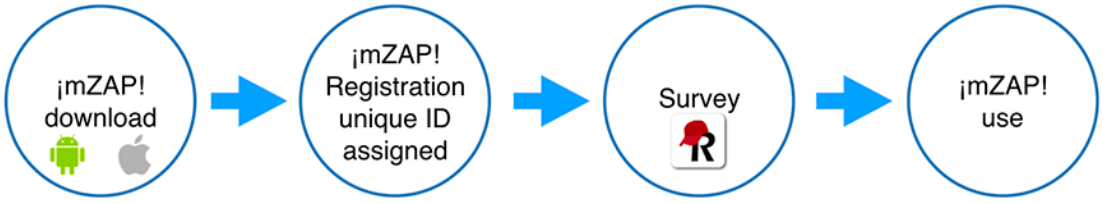
Eligible participants were ≥18 years and self-reported living in Culebra for at least 6 months before the start of the study. Survey questions were administered in Spanish, which is the primary language of the island.

Participants were recruited at baseline following a snowball sampling methodology [51], community meetings, flyer distributions, and social media posts. During in-person visits, eligible participants were invited to download the ¡mZAP! app and complete the survey during the registration process before starting the use of app functionalities (Figure 3). Participants completed the survey using their devices, and via REDCap, the surveys were uploaded to the University of Rochester's firewall-protected, encrypted cloud-based servers. Participants did not receive any compensation for completing the survey or using ¡mZAP! For recruitment purposes, a local area code telephone number and a unique email address were acquired, and a Facebook fan page was created specifically for the project, enabling participants to reach the research team for questions and concerns about the project. This study adhered to the Reporting of Studies Conducted using Observational Routinely Collected Data guidelines and checklist to ensure the inclusion of all components deemed necessary for a scientific research report [52].

¡mZAP! was launched in a series of in-person community meetings in fall of 2018 and was disseminated among community members during a 3-month trial period (November 2018-February 2019). A total of 124 people downloaded the app (Apple iOS 53/124, 42.7% and Android 71/124, 57.3%). Out of the 124 people, 116 (93.5%) participants started registration; 5 (4%) respondents did not meet the inclusion criteria of being a resident of Culebra and were excluded. Out of the 116 participants, 111 (95.7%) participants successfully registered and completed the survey.

This study was reviewed and approved by the institutional review board of the first author's (RSRB00073777) university and followed the 1964 Declaration of Helsinki and its subsequent amendments. HealthproMed, a federally qualified health center in Culebra, reviewed the study protocol and approved the study. An information letter was electronically provided to the participants during their registration for the app.

Figure 3. ¡mZAP! user implementation process. ¡mZAP!: Zonas, Acción y Protección.



Measures

Overview

The study data included two distinct databases: ongoing baseline survey data and ¡mZAP! use data. The app documented user actions and collected data within a 3-month period. *¡mZAP! use* was defined as any action by users within the mobile app (eg, reports, comments, messages, likes, and resolves) during this follow-up period. Multiple user data points were synchronized by random user unique identification, which traced each user's behavior while using the ¡mZAP! tool.

The Environmental Attitudes Inventory

The environmental attitudes inventory [53] scale was adapted (adapted environmental attitudes inventory [a-EAI]) to assess participants' attitudes toward the environment. This environmental scale was developed to evaluate the multidimensional nature of environmental attitudes and the perceptions or beliefs of participants regarding the natural environment and associated factors. For this study, 12 items from the environmental attitudes inventory scale were selected and combined for the a-EAI summary scale. We converted the EAI summary scale to a standardized *t* score with mean 50 (SD 10) [54]. In addition to analyzing the a-EAI as a continuous variable, we reduced the a-EAI to discrete categories with tertiles as the cutoff.

The Wallston Multidimensional Health Locus of Control

The multidimensional health locus of control (MHLC) scale assesses perceptions of control over one's health. The MHLC scale has been used in previous studies to measure the influences and perceptions of community members regarding external factors that affect health outcomes [55]. A Spanish-validated version of the MHLC was used in this study [56]. The three subscales were MHLC internal (internal health locus of control), MHLC external (powerful others health locus of control), and MHLC chance (chance health locus of control) [57]. The rationale for using the MHLC was that perceptions of health control could be associated with overall control and a positive inclination to act against environmental health risks. The subscale scores were standardized using the Streiner and Norman [54] method.

Community Environmental Priorities, Technology Use, Mosquito Knowledge, and Demographics

For addressing community environmental priorities, two additional sections were incorporated from previous study findings [48,49,58] to assess the use of technology, environmental challenges, community resolutions, and access to technology at home (eg, cell phones and the internet).

Data Analysis

The survey data were exported from REDCap, cleaned, and analyzed using SPSS Statistics (IBM Corporation) version 25 [59]. Chi-square tests and analysis of variance were used to test bivariate associations among all key variables. Odds ratios (ORs) with 95% CIs were used to examine the magnitude of the association of predictor variables, including the main predictor variable (attitudes toward the environment, a-EAI) and the primary outcome (use of ¡mZAP!). Forced and forward stepwise logistic regression models [60] were generated to control for potential confounders that were marginally associated (ie, $P<.25$) with a-EAI and mZAP! use. Where necessary, $P<.05$ was considered significant.

Results

Study Sociodemographic Characteristics

In total, 110 participants completed the download of ¡mZAP! and the intake survey. As Table 1 shows, most (71/110, 64.5%) of the survey participants were aged >35 years. In addition, 65.4% (72/110) had completed a higher education degree and 68.5% (76/111) had lived in Culebra for >11 years. An equal number of females and males participated in the study (n=54 each). A comparison of the study sample with the US Census 2013-2017 American Community Survey in Puerto Rico [43] revealed no statistically significant differences between the study population and the population of the municipality of Culebra regarding gender and ethnicity (data not shown), suggesting that the study sample was similar to the overall population of Culebra.

Table 1. Sociodemographics and technology attitudes among ¡mZAP! (Zonas, Acción y Protección) study survey participants in Culebra, Puerto Rico, 2018-2019 (N=110).

Variables	Participants, n (%)
Age (years)	
18-25	10 (9.1)
26-36	29 (26.4)
37-47	34 (30.9)
48-54	22 (20)
≥55	15 (13.6)
Age category (years)	
≤36	39 (35.5)
≥37	71 (64.5)
Education level	
Less than or equal to high school	34 (32.1)
Higher education	72 (67.9)
Gender	
Female	54 (49.1)
Male	54 (49.1)
Prefer not to answer	2 (1.8)
Time living in Culebra (years; N=111)	
1-10	35 (31.5)
More than 11	76 (68.5)
Do you have a cell phone data plan?	
Yes	105 (96.3)
No	4 (3.7)
How is your cell phone reception on the island?	
Good	11 (10)
Regular or bad	99 (90)
How many cell phones do you have at home?	
1	28 (25.5)
>1	82 (74.6)
Does your home have internet access?	
Yes	96 (87.3)
No	14 (12.7)
What type of internet do you have in your home?	
Wi-Fi and satellite	31 (32.3)
Data plan	63 (65.6)
How many hours per day do you spend using electronic devices?	
<1-5	89 (80.9)
≥6	21 (19.1)
Do you use social media?	
Yes	105 (95.5)
No	5 (4.5)
Daily use of social media	

Variables	Participants, n (%)
WhatsApp	99 (94.3)
Facebook	89 (84.8)
Instagram	43 (41.3)
Agreement with the following statements	
Technology is a tool to help us daily	100 (90.9)
The use of technologies helps to fix community problems	66 (60)
The use of technologies brings us closer to the community	37 (33.9)
The constant use of technology limits the community work	43 (39.1)
The technology cause problems in the community	36 (32.7)

Technology and Community Characteristics Among ¡mZAP! Survey Participants

As Table 1 shows, every participant reported owning a cell phone, and 96.3% (105/110) of participants also had a data plan as part of their cell phone plan. With respect to technology at home, 87.3% (96/110) of participants reported having internet access at home, with 66% (63/95) using cell phone data plans for connecting to the internet. Most of the participants owned more than one cell phone (62/83, 74.6%), used electronic devices for up to 5 hours daily (89/110, 80.9%), and had negative perceptions about cell phone reception (99/110, 90%). In addition, 95.5% (105/110) of participants reported being social media users, with WhatsApp being the most used daily social media network (99/105, 94.3%), followed by Facebook (89/105, 84.8%) and Instagram (43/105, 40.1%).

In a series of statements about the relationship between technology and the community, 90.9% (100/110) of participants believed that technology is a tool that can be useful daily and 60% (66/110) agreed that technology could facilitate the resolution of problems in the community. Participants disagreed or were undecided about the role of technology in bringing communities closer (72/110, 65.4%), whether using technologies was an impediment for individuals to do community work (67/110, 60.9%), and whether technology caused community problems (74/110, 67.3%).

Attitudes Toward the Environment and Predictors

On the basis of the responses on the a-EAI scale, participants often disagreed with items reflecting negative statements about protecting the environment. For example, 44.2% (49/116) disagreed with “One of the most important reasons to keep lakes, beaches and rivers clean is so that people have a place to enjoy water sports,” and 46.8% (51/109) disagreed with “Modern science will not be able to solve our environmental problems.” Most participants agreed with or were ambivalent about items that reflected positive inclinations toward the environment; for example, “Protecting the environment is more important than protecting peoples’ jobs” (65/109, 59.6%) and “I would like to join and actively participate in an environmentalist group” (73/110, 66.4%). When comparing a-EAI with sociodemographic variables, as shown in Table 2, having a higher educational level (OR 4.5, 95% CI 1.2-16.3) and having lived fewer total years in Culebra were significantly associated with a stronger positive attitude toward the environment (OR 3.0, 95% CI 1.2-7.6).

Table 2. Associations of the a-EAI[a] with sociodemographics, technology attitudes, and knowledge about mosquito-borne diseases among ¡mZAP! (Zonas, Acción y Protección) study survey participants in Culebra, Puerto Rico, 2018-2019.

Variables	Total participants, n (%)	a-EAI, n (%)		Odds ratio (95% CI)	*P* value[b]
		High (more positive)	Low (less positive)		
Age (years; n=108)					
≤36	37 (34.3)	11 (29.7)	26 (70.3)	1.5 (0.6-3.6)	.41
≥37	71 (65.7)	16 (22.5)	55 (77.5)	Referent	N/A[c]
Education level (n=104)					
Higher education	71 (68.3)	22 (31)	49 (69)	4.5 (1.2-16.3)	.01
Less than or equal to high school	33 (31.7)	3 (9.1)	30 (90.9)	Referent	N/A
Time living in Culebra (years; n=108)					
>11	73 (67.6)	13 (17.8)	60 (82.2)	Referent	N/A
1-10	35 (32.4)	14 (40)	21 (60)	3.0 (1.2-7.6)	.01
How is your cell phone reception on the island? (n=108)					
Regular or bad	97 (89.8)	25 (25.8)	72 (74.2)	1.5 (0.3-7.7)	.58
Good	11 (10.2)	2 (18.2)	9 (81.8)	Referent	N/A
How many cell phones do you have at home? (n=108)					
>1	80 (74.1)	22 (27.5)	58 (72.5)	1.7 (0.6-5.2)	.31
1	28 (25.9)	5 (17.9)	23 (82.1)	Referent	N/A
Does your home have internet access? (n=108)					
Yes	95 (87.9)	24 (25.3)	71 (74.7)	1.1 (0.3-4.4)	.86
No	13 (12.1)	3 (23.1)	10 (76.9)	Referent	N/A
What type of internet do you have in your home? (n=95)					
Wi-Fi and satellite	33 (34.7)	6 (18.2)	27 (81.8)	Referent	N/A
Data plan	62 (65.3)	18 (29)	44 (71)	1.8 (0.7-5.2)	.24

[a]a-EAI: adapted environmental attitudes inventory.

[b]*P* value of chi-square test.

[c]N/A: not applicable for reference groups.

Collective Action Strategies

A series of questions on the importance of community actions to address environmental risks were posed. As shown in Table 3, most participants (79/108, 73.1%) believed that environmental problems are not typically resolved among community members. Similarly, 58.3% (63/108) of participants expected the government to resolve environmental problems. Other participants expressed that these problems would be resolved by the community (29/108, 26.9%) or would be resolved without making any effort to fix them (28/108, 25.9%). No statistically significant associations were found among the series of items on community actions and a-EAI scale scores.

Table 3. Associations of the a-EAI[a] with community environmental actions resolutions among ¡mZAP! (Zonas, Acción y Protección) study survey participants in Culebra, Puerto Rico, 2018-2019 (N=108).

How does generally environmental differences or problems get resolved in your community?	Total participants, n (%)	a-EAI, n (%)		Odds ratio (95% CI)	*P* value[b]
		Hi (more positive)	No (more negative)		
Resolved between each other					
No	79 (73.1)	20 (25.3)	59 (74.7)	1.0 (0.4-2.9)	.90
Yes	29 (26.9)	7 (24.1)	22 (75.9)	Referent	N/A[c]
Go to government authorities (police and mayor's office)					
Yes	63 (58.3)	13 (20.6)	50 (79.4)	Referent	N/A
No	45 (41.6)	14 (31.1)	31 (68.9)	1.7 (0.7-4.2)	.22
Ignore and *pass the page*					
No	83 (76.8)	21 (25.3)	62 (74.7)	1.0 (0.4-3.0)	.89
Yes	25 (23.1)	6 (24)	19 (76)	Referent	N/A
Think it will eventually get resolved					
No	82 (75.9)	22 (26.8)	60 (73.2)	1.5 (0.5-4.6)	.44
Yes	26 (24)	5 (19.2)	21 (80.8)	Referent	N/A
It will get resolved by itself					
Yes	28 (25.9)	8 (28.6)	20 (71.4)	1.2 (0.4-3.3)	.61
No	80 (74)	19 (23.8)	61 (76.3)	Referent	N/A

[a]a-EAI: adapted environmental attitudes inventory.

[b]*P* value of chi-square test.

[c]N/A: not applicable for reference groups.

Multidimensional Health Locus of Control

Most participants reported moderate beliefs about health control, MHLC internal (76/118, 64.4%), MHLC *powerful others* (78/108, 72.2%), and MHLC chance (71/108, 65.7%; data not shown). No statistically significant associations were found between the a-EAI and MHLC subscales. In contrast, when comparing MHLC subscales with the a-EAI as a scale score, as MHLC *powerful others* scores increased, overall a-EAI scores decreased (*P*=.02; data not shown). This association indicates that attitudes toward the environment become more negative as the sense that one's health is controlled by *powerful others* (eg, parents, doctors, and authorities) increases.

¡mZAP! User's Characteristics and Predictors

Of the 14 participants who used the ¡mZAP! app, 5 (36%) participants used the app more than once; 11 (78%) participants used the app to report environmental health risks, including trash, stray animals, mosquito breeding grounds, abandoned structures, and deposits of stagnant water; and 2 (14%) participants commented on reports. Of the 14 users, 3 (21%) used the tool to report the maritime transportation location (location of the ferries). In addition, of the 14 ¡mZAP! users, 9 (64%) expressed the expectation that the government would resolve the environmental challenges in the community.

Equal proportions of women and men used ¡mZAP! As shown in Table 4, most users (11/14, 79%) were aged >35 years, lived in Culebra for >11 years (10/14, 71%), and had completed higher education (10/14, 71%). No statistically significant differences were found between nonusers and users in terms of sociodemographic variables. Although not significant, several variables had substantial ORs (ORs>1.5): age (OR 2.2, 95% CI 0.5-8.4), educational level (OR 1.6, 95% CI 0.4-6.5), cell phone number per household (OR 2.2, 95% CI 0.5-10.6), household type of internet access (OR 1.8, 95% CI 0.7-5.2), daily hours spent using electronic devices (OR 2.8, 95% CI 0.8-9.3), and *the use of technologies helps to fix community problems* (OR 3.1, 95% CI 0.9-10.0).

Table 4. Association of ¡mZAP![a] use with sociodemographics, technology attitudes, and knowledge about mosquito-borne diseases among ¡mZAP! study survey participants in Culebra, Puerto Rico, 2018-2019.

Variables	Total participants, n (%)	¡mZAP! use		Odds ratio (95% CI)	*P* value[b]
		Yes	No		
Age category (years; n=110)					
≤36	39 (35.5)	3 (7.7)	36 (92.3)	Referent	N/A[c]
≥37	71 (64.5)	11 (15.5)	60 (84.5)	2.2 (0.5-8.4)	.24
Education level (n=106)					
Higher education	72 (67.9)	10 (13.9)	62 (86.1)	1.6 (0.4-6.5)	.46
Less than equal to high school	34 (32.1)	3 (8.8)	31 (91.2)	Referent	N/A
Time living in Culebra (years; n=111)					
>11	76 (68.5)	10 (13.2)	66 (86.8)	1.1 (0.3-4.0)	.79
1-10	35 (31.5)	4 (11.4)	31 (88.6)	Referent	N/A
How is your cell phone reception on the island? (n=110)					
Regular or bad	99 (90)	14 (14.1)	85 (85.9)	N/A	N/A
Good	11 (10)	0 (0)	11 (100)	N/A	N/A
How many cell phones do you have at home? (n=110)					
>1	82 (74.5)	12 (14.6)	70 (85.4)	2.2 (0.5-10.6)	.30
1	28 (25.5)	2 (7.1)	26 (92.9)	Referent	N/A
Does your home have internet access? (n=110)					
Yes	96 (87.3)	12 (12.5)	84 (87.5)	Referent	N/A
No	14 (12.7)	2 (14.3)	12 (85.7)	1.1 (0.2-5.8)	.85
What type of internet do you have in your home? (n=95)					
Wi-fi and satellite	33 (34.7)	6 (18.2)	27 (81.8)	Referent	N/A
Data plan	62 (65.3)	18 (29)	44 (71)	1.8 (0.7-5.2)	.25
How many hours per day do you spend using electronic devices? (n=110)					
≥6 hours	21 (19.1)	5 (23.8)	16 (76.2)	2.8 (0.8-9.3)	.90
<1 hour or up to 5 hours	89 (80.9)	9 (10.1)	80 (89.9)	Referent	N/A
a-EAI[d] scores (n=108)					
Higher (top tertile) score and lower (lowest two tertiles) score	N/A	N/A	N/A	N/A	N/A
Two-third low score	81 (75)	6 (7.4)	75 (92.6)	Referent	N/A
One-third high score	27 (25)	8 (29.6)	19 (70.4)	5.3 (1.6-17.0)	.003
Uses social media (n=110)	105 (95.5)	14 (13.3)	91 (86.7)	N/A	N/A
How often do you use WhatsApp? (n=105)					
Less than daily	6 (5.7)	1 (16.7)	5 (83.3)	1.3 (0.1-12.2)	.81
Daily	99 (94.3)	13 (13.1)	86 (86.9)	Referent	N/A
How often do you use Facebook? (n=105)					
Daily	89 (84.8)	12 (13.5)	77 (86.5)	1.1 (0.2-5.4)	.91
Less than daily	16 (15.2)	2 (12.5)	14 (87.5)	Referent	N/A
How often do you use Instagram? (n=104)					
Daily	43 (41.3)	8 (18.6)	35 (81.4)	2.1 (0.7-6.6)	.19
Less than daily	61 (58.7)	6 (9.8)	55 (90.2)	Referent	N/A
Technology is a tool to help us daily (n=110)					

Variables	Total participants, n (%)	¡mZAP! use		Odds ratio (95% CI)	*P* value[b]
		Yes	No		
Agree	100 (90.9)	14 (14)	86 (86)	N/A	N/A
The use of technologies helps to fix community problems (n=110)					
Undecided or disagree	44 (40)	9 (20.5)	35 (79.5)	3.1 (0.9-10.0)	.04
Agree	66 (60)	5 (7.6)	61 (92.4)	Referent	N/A
The use of technologies brings us closer to the community (n=109)					
Undecided or disagree	72 (66.1)	10 (13.9)	62 (86.1)	1.3 (0.3-4.6)	.65
Agree	37 (33.9)	4 (10.8)	33 (89.2)	Referent	N/A
The constant use of technology limits the community work (n=110)					
Undecided or disagree	67 (60.9)	9 (13.4)	58 (86.6)	1.2 (0.3-3.8)	.78
Agree	43 (39.1)	5 (11.6)	38 (88.4)	Referent	N/A
The technology cause problems in the community (n=110)					
Undecided or disagree	74 (67.3)	10 (13.5)	64 (86.5)	1.3 (0.4-4.3)	.72
Agree	36 (32.7)	4 (11.1)	32 (88.9)	Referent	N/A

[a]¡mZAP!: Zonas, Acción y Protección.

[b]*P* value of chi-square test.

[c]N/A: not applicable.

[d]a-EAI: adapted environmental attitudes inventory.

No item from the MHLC was statistically significantly associated with the use of the ¡mZAP! app, including when compared with each subscale's total score (MHLC subscale 1 [internal health locus of control], MHLC subscale 2 [powerful others health locus of control], and MHLC subscale 3 [chance health locus of control]; data not shown).

Higher positive attitudes toward the environment were significantly associated with using the ¡mZAP! app (OR 5.3, 95% CI 1.6-17.0). To assess potential confounding factors, we selected variables that were marginally associated with a-EAI and the use of the ¡mZAP! app, although none of the variables met the statistical criteria for remaining in the model. Clinically relevant sociodemographic variables (education level and age) were forced into the logistic regression model and showed no confounding between environmental attitudes and the use of ¡mZAP! app.

A statistically significant ($P<.01$) association persisted when comparing the highest tertile of a-EAI versus the other two tertiles and ¡mZAP! app use (adjusted OR 5.4, 95% CI 1.4-20.4). The Hosmer and Lemeshow goodness-of-fit test [60] was not statistically significant ($P=.89$), validating the tested model.

Discussion

Principal Findings

This study was followed by the *Community-Centered Environmental Health Risk Control Model*. This theoretically informed model facilitated the process of learning about *Culebrenses'* environmental health priorities and, more importantly, the implementation of their community perspectives in the development of ¡mZAP! app as an mHealth tool tailored to the community and by the community. This new model contributes to our understanding of behaviors that can lead to better ICT use among community members, and this model can be adopted in future studies worldwide.

The main findings of this study are as follows: (1) ¡mZAP! users were five times more likely to have stronger environmental attitudes than ¡mZAP! nonusers; (2) a negative relationship between environmental attitudes and the MHLC *powerful others* was observed; and (3) an expectation that the government will meet the needs of *Culebrenses* was found.

The study results also suggest that *Culebrenses* have a strong use of ICT, where most participants spend up to 5 hours daily using electronic devices. In contrast, this study also found important aspects of *Culebrenses'* potential interpretations of and implications for their attitudes toward the environment. For example, participants with a low perception of health control were also more likely to have a less positive attitude toward the environment. This association indicates that attitudes toward the environment become more negative as the sense that one's health is controlled by *powerful others* increases. Higher scores on the MHLC *powerful others* subscale have been previously demonstrated to affect health outcomes negatively [61,62]. This relationship could also lead people to become disengaged in efforts to protect their environment. People may hold conflicting beliefs because most participants agreed with or were ambivalent about items that reflected positive inclinations toward the environment.

As described in previous studies, *Culebrenses* expressed a feeling of *being forgotten* by the Puerto Rican government, which could partly explain why *Culebrenses* have a sense of pride and ownership with respect to Culebra [48]. As a result, *Culebrenses* could demonstrate a greater inherent consciousness of protecting natural resources, the environment, and fragile

ecosystems, which are also the island's main tourist attractions. The results of this study help to empirically reaffirm Culebra's positive community attitudes and beliefs toward the environment. Although the intentions to protect the island were in the minds of community members, these intentions did not necessarily translate into the use of ¡mZAP! to protect the environment and address environmental health risks that affect islanders.

This study suggests that less time spent living in Culebra was associated with stronger positive attitudes toward the environment. This finding is important, particularly considering that most ¡mZAP! users have lived in Culebra for >11 years. Perhaps these results suggest that the longer one lives in a location, the more likely one is to be disincentivized to engage in behaviors to respond to community environmental challenges. For example, *Culebrenses* have been exposed to many environmental health risks in the past, including solid waste polluting beaches, improper land development, air pollution, and most notably, the US military bombing practices [35,63-67]. Although the community advocated against these environmental injustices in the past [66,67], the ongoing repercussions of these challenges [31,68] have potentially resulted in reduced enthusiasm. ¡mZAP! users had higher positive attitudes toward the environment, which could also translate into an increase in environmental conscientiousness.

Participants' beliefs that government agencies are responsible for responding to environmental risks and community discrepancies related to environmental challenges help us better understand *Culebrenses*' perceptions of lack of control over the community's health. Societal issues that affect the community's poor health outcomes extend into the social and ecological determinants of health [69,70]. The ongoing dependency on government-based solutions to community problems is worrisome, especially after the hurricane experiences of 2017 (which may potentially continue to be an issue subsequent to the more recent 2020 earthquakes).

The catastrophic impact of hurricanes Irma and María and the inappropriate responses of the Puerto Rican and US governments may have cemented this perception of lack of control. Community members experienced the isolation of being disconnected from the main island of Puerto Rico and the exacerbation of an unreliable maritime transportation system, resulting in a societal crisis where health access, common goods, and food became scarce [71,72].

Culebrenses' dependency systems and the way social injustices from the past and the present may have resulted in the perception of lack of control, which was associated in this study with negative environmental attitudes, can undermine the good intentions of community members to protect their land. Therefore, implementing community-driven approaches to address environmental health risks or health disparities may not be sufficient to ensure a successful mHealth tool intervention, especially when other social conditions work as oppressors. Socioecological conditions force community priorities to be in constant transformative change. This transformative change was especially true after the 2017 hurricane season in Culebra, where the emergence of new challenges such as rebuilding destroyed properties became a new top priority for *Culebrenses*.

Previous studies suggest that new mobile apps are used between 4% and 20% of the time, with an average session app use of <1 minute. Furthermore, 51% of apps are deleted after the first week of use [73-78]. In this study, only 12.6% (14/111) of participants adopted the use of ¡mZAP!. Although the number of initial users was small, it may be noteworthy that 36% (5/14) of participants used the mHealth tool more than once. These adoption and reuse rates are typical when compared with the adoption rates of other mobile apps. In addition, identifying proper champions on the island to promote ¡mZAP! presented a challenge that could have also affected the use of the mHealth tool. Early identification of community champions has been previously studied as an approach to develop and increase trust between community members and investigators, facilitating an engaging process and a successful community-based research intervention [79-81].

This research study lacked sufficient power to detect some potentially meaningful and theoretically driven associations. The observed associations that were not statistically significant but which had ORs >1.5 can, therefore, be useful for hypothesis generation and to inform future studies. Therefore, in a larger sample, certain relevant predictor variables would have statistically significant associations with the primary outcome, such as age, education level, cell phone number and type of internet per household, daily use of electronic devices, and the perception of how technology contributes to help fixing community problems.

The results of this study provide some support for suggesting that people used ¡mZAP! and were taking self-directed action by using ICTs to address environmental health challenges in Culebra. These results, although limited, confirm the objective of this study to assess the ability of ICTs as tools to engage islanders in collective actions that address environmental health risks. However, for other *Culebrenses*, although they may have similar intentions to protect the environment, existing higher community priorities and potential oppressive challenges, including health care and maritime transportation access, prevent them from translating their intentions into actions.

Limitations and Strengths

Although this pilot study may offer some insight into directional relationships via statistically significant associations in this sample, interpretation of results is limited because of the small sample size, which restricted power. There may have been a selection bias, including volunteer bias. The data collected for this study reflect a one-time *snapshot* where questions were self-administered by participants, potentially generating respondent bias. The hurricane disasters of 2017 may have substantially affected community perceptions and priorities to the extent that the development of ¡mZAP! was affected in unexpected ways.

The empirical findings concerning perceived lack of control and government dependency systems support conclusions about how these situations could have a stronger influence on people's behaviors. These behaviors could lead to a perpetuation of the

challenges associated with this population's unique social determinants of health. The results from this study should be further qualified by the fact that people who are less familiar with technology may not be as represented in this study as those who are more familiar (eg, those who are younger and have higher education). These contextual factors are important in future research to ensure the potential adoption and success of any ICT.

The study results might only apply to the users of the ¡mZAP! app, a tool that proactively focuses on environmental health risks. Anecdotally we know that some community members were positively affected by the tool (social influence from other users). The perceptions and priorities of nonusers are likely to differ. Future studies could explore the relationship between offline and internet-based generations to address communities' environmental priorities.

This study had several strengths. Existing partnerships with collaborators in Puerto Rico facilitated access to the target population and necessary local ethical review and approval. The study addressed key goals and objectives of Healthy People 2030, including "Use health communication strategies and health information technology to improve population health outcomes and health care quality, and to achieve health equity" [82]. This is the first study to provide an overview of *Culebrenses'* intentions to protect the island's environment in the context of mHealth resource use. In addition, this study provides a unique perspective on how people in remote and underresourced communities perceive environmental health risk and how those perceptions affect the use of rapidly advancing mobile technologies, which can help decrease barriers to access to health in rural areas [2,83,84]. In conclusion, the study findings demonstrated the capacity to stimulate collective action by using ICTs as a novel and engaging approach in underresourced rural locations.

Implications and Future Research

Future research studies should seek to better understand the factors preventing the use of ¡mZAP! and other ICTs, including changes in the community's perceptions and priorities after natural disasters or other major community-wide challenges. It is important to explore how ICT use can support, facilitate, or even drive collective community actions. For example, there are multiple community-based centers, groups, and institutions, including a federally qualified health center, a women's health community organization, and other grassroot-level environmental organizations that can be supplied with ¡mZAP! and other mHealth ICT–related mobile technologies. These tools could be coupled with educational materials and community-based initiatives aimed at increasing positive perceptions of the environment by community members in this rural setting.

Future research should work toward a better understanding of community members' priorities and addressing pre-existing social determinants, such as those found in this study. Although other social determinants of health in this study, including colonialism, were not assessed, this research establishes a contributing baseline to further investigate the relationship between health or environmental disparities and the sociopolitical power imbalances that affect community islands such as Culebra [85,86]. As community-driven interventions have been successful in the past, with the understanding that community work takes time to develop, the findings of this study can serve as a foundation for future community and ICT research in Culebra and other locations with similar environmental health conditions.

Acknowledgments

The authors would like to thank the community of Culebra, in particular the collaborators in this study from the Amigos de la Reserva Natural Caño Luis Peña organization (Nicolás Gómez Andujar, Abimarie Otaño, and Alfredo Montañez), Dulce María del Río Pineda and MariCarmen Carbonell from the Mujeres de Islas organization, HealthproMed Inc, Escuela Ecológica de Culebra, Abbie White-Thomas (Abbie's School), Foundation for a Better Puerto Rico, José D Pérez from the Municipality of Culebra, Office of Emergencies Management, and Yahaira Rivera, former Culebra's social worker from the Administración de Servicios de Salud y Contra la Adicción. The authors would also like to acknowledge Ahona Shirin, Matthew Quick, Derek Guo, and Astghik Baghinyan for their contributions to this manuscript. The authors are also thankful to The Dye Lab, including the ¡mZAP! (Zonas, Acción y Protección) developing team, Michael Kelly, and Noah Wilson. This study was funded by the University of Rochester, School of Medicine and Dentistry, Department of Obstetrics and Gynecology Pilot Research Award and the University of Rochester, Office of Equity and Inclusion supported journal publication fees. At the time of this work, the corresponding author (JGPR) was enrolled in a doctoral program and accessed REDCap (Research Electronic Data Capture) resources supported by the University of Rochester, Clinical and Translational Science Awards, award numbers TL1TR002000 and UL1TR002001 from the National Center for Advancing Translational Sciences of the National Institutes of Health. The content of this work is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Adibi S. Mobile Health: A Technology Road Map. Berlin, Germany: Springer; 2015.

2. Mbuthia F, Reid M, Fichardt A. mHealth communication to strengthen postnatal care in rural areas: a systematic review. BMC Pregnancy Childbirth 2019 Nov 06;19(1):406 [FREE Full text] [doi: 10.1186/s12884-019-2531-0] [Medline: 31694578]
3. Environment and health in developing countries. Health and Environment Linkages Initiative, World Health Organization. URL: https://www.who.int/heli/risks/ehindevcoun/en/ [accessed 2021-08-11]
4. Neira M, Prüss-Ustün A. Preventing disease through healthy environments: A global assessment of the environmental burden of disease. Toxicol Lett 2016 Oct 10;259:S1. [doi: 10.1016/j.toxlet.2016.07.028]
5. Brewer LC, Fortuna KL, Jones C, Walker R, Hayes SN, Patten CA, et al. Back to the future: achieving health equity through health informatics and digital health. JMIR Mhealth Uhealth 2020 Jan 14;8(1):e14512 [FREE Full text] [doi: 10.2196/14512] [Medline: 31934874]
6. Norman CD, McIntosh S, Selby P, Eysenbach G. Web-assisted tobacco interventions: empowering change in the global fight for the public's (e)Health. J Med Internet Res 2008 Nov 25;10(5):e48 [FREE Full text] [doi: 10.2196/jmir.1171] [Medline: 19033147]
7. Haenssgen MJ, Charoenboon N, Zanello G. You've got a friend in me: how social networks and mobile phones facilitate healthcare access among marginalised groups in rural Thailand and Lao PDR. World Dev 2021 Jan;137:105156. [doi: 10.1016/j.worlddev.2020.105156]
8. Johnston FH, Wheeler AJ, Williamson GJ, Campbell SL, Jones PJ, Koolhof IS, et al. Using smartphone technology to reduce health impacts from atmospheric environmental hazards. Environ Res Lett 2018 Mar 29;13(4):044019. [doi: 10.1088/1748-9326/aab1e6]
9. Jafari H, Li X, Qian L, Chen Y. Community based sensing: a test bed for environment air quality monitoring using smartphone paired sensors. In: Proceedings of the 36th IEEE Sarnoff Symposium. 2015 Presented at: 36th IEEE Sarnoff Symposium; Sept 20-22, 2015; Newark, NJ, USA. [doi: 10.1109/sarnof.2015.7324635]
10. Van 'T Hooft M, DÍaz S, Swan K. Examining the potential of handheld computers: findings from the Ohio PEP project. J Educ Comput Res 2016 Jun 1;30(4):295-311. [doi: 10.2190/m1w6-a94d-3nkm-kbuu]
11. Swan K, Hooft M, Kratcoski A, Unger D. Uses and effects of mobile computing devices in K–8 classrooms. J Res Technol Educ 2005;38(1):99-112. [doi: 10.1080/15391523.2005.10782451]
12. English P, Richardson M, Garzón-Galvis C. From crowdsourcing to extreme citizen science: participatory research for environmental health. Annu Rev Public Health 2018 Apr 01;39(1):335-350. [doi: 10.1146/annurev-publhealth-040617-013702] [Medline: 29608871]
13. Uzunboylu H, Cavus N, Ercag E. Using mobile learning to increase environmental awareness. Comp Educ 2009 Feb;52(2):381-389. [doi: 10.1016/j.compedu.2008.09.008]
14. Benelli G, Jeffries C, Walker T. Biological control of mosquito vectors: past, present, and future. Insects 2016 Oct 03;7(4):52 [FREE Full text] [doi: 10.3390/insects7040052] [Medline: 27706105]
15. Cañizares JC. The information society: technological, socioeconomic and cultural aspects - prolegomena for a sustainability-oriented ethics of ICT, with glossary and PhD prop. University of Twente Student Theses. 2017. URL: http://essay.utwente.nl/73135/ [accessed 2021-08-11]
16. Hoffman D, Novak T, Schlosser A. The evolution of the digital divide: how gaps in Internet access may impact electronic commerce. J Comput Mediat Commun 2000 Mar 1;5(3):JCMC534. [doi: 10.1111/j.1083-6101.2000.tb00341.x]
17. Nemer D. Online favela: the use of social media by the marginalized in Brazil. Inf Technol Dev 2015 Apr 18;22(3):364-379. [doi: 10.1080/02681102.2015.1011598]
18. Nemer D, Tsikerdekis M. Political engagement and ICTs: internet use in marginalized communities. J Assoc Inf Sci Technol 2017 Apr 28;68(6):1539-1550. [doi: 10.1002/asi.23779]
19. Xu Y, Yen DC, Lin B, Chou DC. Adopting customer relationship management technology. Ind Manag Data Syst 2002 Nov;102(8):442-452. [doi: 10.1108/02635570210445871]
20. Levine AR, Buchner JA, Verceles AC, Zubrow MT, Mallemat HA, Papali A, et al. Ultrasound images transmitted via FaceTime are non-inferior to images on the ultrasound machine. J Crit Care 2016 Jun;33:51-55. [doi: 10.1016/j.jcrc.2016.02.019] [Medline: 27006267]
21. Madon S, Krishna S. The Digital Challenge: Information Technology in the Development Context. Oxfordshire, England: Routledge; Feb 6, 2018:1-386.
22. Taub A, Fisher M. How YouTube misinformation resolved a WhatsApp mystery in Brazil. The New York Times. 2019 Aug 15. URL: https://www.nytimes.com/2019/08/15/the-weekly/how-youtube-misinformation-resolved-a-whatsapp-mystery-in-brazil.html?action=click&module=Briefings&pgtype=Homepage [accessed 2021-08-11]
23. Boyd A, Gatewood J, Thorson S, Dye T. Data diplomacy. Science & Diplomacy. URL: https://www.sciencediplomacy.org/article/2019/data-diplomacy [accessed 2021-08-11]
24. Minkler M. Community Organizing and Community Building for Health and Welfare. New Brunswick, New Jersey: Rutgers University Press; 2012.
25. Qureshi S, Najjar L. Information and communications technology use and income growth: evidence of the multiplier effect in very small island states. Inf Technol Dev 2017 Apr 26;23(2):212-234. [doi: 10.1080/02681102.2016.1173634]
26. Baldacchino G. A World of Islands: An Island Studies Reader. Charlottetown: University of Prince Edward Island; 2007.

27. Dye TD, Sy A, Albert P, Cash H, Hadley J, Tomeing T, et al. Critical medical ecological perspectives on diabetes in the Pacific Islands: colonialism, power, and balance in human-environment interaction over time. Lancet Glob Health 2018 Mar 1;6:S36. [doi: 10.1016/s2214-109x(18)30165-7]
28. Douglas CH. Small island states and territories: sustainable development issues and strategies – challenges for changing islands in a changing world. Sust Dev 2006 Mar 3;14(2):75-80. [doi: 10.1002/sd.297]
29. Davis JS, Hayes-Conroy JS, Jones VM. Military pollution and natural purity: seeing nature and knowing contamination in Vieques, Puerto Rico. GeoJ 2007 Aug 3;69:165-179. [doi: 10.1007/s10708-007-9095-7]
30. Cintrón C. Culebra y la Marina de Estados Unidos. San Juan, Puerto Rico: Editorial Edil; Jan 1, 1989.
31. Rogers C, Cintrón G, Goenaga C. The impact of military operations on the coral reefs of Vieques and Culebra. Report to Department of Natural Resources, San Juan, Puerto Rico. 1978. URL: http://www.reefbase.org/resource_center/publication/main.aspx?refid=4823 [accessed 2021-08-11]
32. Vistazo a la Salud Pública en Culebra. Dialogo UPR. 2015. URL: http://dialogoupr.com/vistazo-la-salud-publica-en-culebra/ [accessed 2021-08-11]
33. Martinez RB. La Independencia de Puerto Rico (Razon y Lucha). San Juan, Puerto Rico: Editorial Linea; Jan 1, 1983.
34. Culebra recibirá desde hoy electricidad por cable submarino desde Vieques. Agencia EFE. URL: https://www.efe.com/efe/usa/puerto-rico/culebra-recibira-desde-hoy-electricidad-por-cable-submarino-vieques/50000110-3930344 [accessed 2021-08-11]
35. Feliciano C. Apuntes y Comentarios de la Colonización y Liberación de la Isla de Culebra. : Self-Published; Jan 1, 2001.
36. Asociación de Ejecutivos de Ventas y Mercadeo de Puerto Rico (SME). Digital & Mobile Behavioral Study. 2016. URL: https://www.smepr.org/index.php?src=news&srctype=detail&category=Highlighted&refno=175 [accessed 2019-05-30]
37. Haider M, Kreps GL. Forty years of diffusion of innovations: utility and value in public health. J Health Commun 2004;9 Suppl 1:3-11. [doi: 10.1080/10810730490271430] [Medline: 14960400]
38. Rawstorn JC, Gant N, Meads A, Warren I, Maddison R. Remotely delivered exercise-based cardiac rehabilitation: design and content development of a novel mHealth platform. JMIR Mhealth Uhealth 2016 Jun 24;4(2):e57 [FREE Full text] [doi: 10.2196/mhealth.5501] [Medline: 27342791]
39. Henrich J. Cultural transmission and the diffusion of innovations: adoption dynamics indicate that biased cultural transmission is the predominate force in behavioral change. Am Anthropol 2001 Dec;103(4):992-1013. [doi: 10.1525/aa.2001.103.4.992]
40. IDEO (Firm). The Field Guide to Human-centered Design: Design Kit. San Francisco, California: IDEO; 2015.
41. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. London, United Kingdom: Pearson; Nov 1, 1985.
42. Rogers E. Diffusion of Innovations. Mumbai: Free Press; Jul 6, 2010.
43. American Community Survey (ACS). The United States Census Bureau. 2017. URL: https://www.census.gov/programs-surveys/acs [accessed 2021-08-11]
44. Culebra National Wildlife Refuge. United States Fish and Wildlife Service. URL: https://www.fws.gov/refuge/Culebra/wildlife_and_habitat.html [accessed 2021-08-11]
45. Manifestantes llevan reclamos sobre lanchas para Vieques y Culebra a La Fortaleza y al Capitolio. Elnuevodia.com. URL: https://www.elnuevodia.com/noticias/locales/notas/manifestantes-llevan-reclamos-sobre-lanchas-para-vieques-y-culebra-a-la-fortaleza-y-al-capitolio/ [accessed 2021-08-11]
46. Hernández-Delgado EA, Acuña AM, Otaño-Cruz A, Suleiman-Ramos SE. Los arrecifes de coral con craters-bomba en Puerto Rico, la historia no contada sobre un hábitat inusual: desde la destrucción de arrecifes hasta la rehabilitación ecológica basada en la comunidad. Rev Biol Trop 2014 Sep 01;62(Suppl 3):183-200. [doi: 10.15517/rbt.v62i0.15913]
47. Rivera-Arguinzoni A. Seis municipios donde no nacen bebés. Elnuevodía.com. URL: https://www.elnuevodia.com/noticias/locales/notas/seis-municipios-donde-no-nacen-bebes/ [accessed 2021-08-11]
48. Perez-Ramos JG, McIntosh S, Barrett ES, Vélez-Vega CM, Dye TD. Qualitative assessment of environmental health risk perceptions and community challenges in a Puerto Rican community: change and continuity in response to hurricanes Irma and María. Behav Med 2020;46(3-4):231-244. [doi: 10.1080/08964289.2019.1700894] [Medline: 31860413]
49. McIntosh S, Perez-Ramos J, Demment MM, Vélez Vega C, Avendaño E, Ossip DJ, et al. Development and implementation of culturally tailored offline mobile health surveys. JMIR Public Health Surveill 2016 Jun 02;2(1):e28 [FREE Full text] [doi: 10.2196/publichealth.5408] [Medline: 27256208]
50. Whittaker R, Merry S, Dorey E, Maddison R. A development and evaluation process for mHealth interventions: examples from New Zealand. J Health Commun 2012;17 Suppl 1:11-21. [doi: 10.1080/10810730.2011.649103] [Medline: 22548594]
51. Biernacki P, Waldorf D. Snowball sampling: problems and techniques of chain referral sampling. Sociol Methods Res 1981 Nov 1;10(2):141-163. [doi: 10.1177/004912418101000205]
52. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med 2015 Oct 6;12(10):e1001885 [FREE Full text] [doi: 10.1371/journal.pmed.1001885] [Medline: 26440803]
53. Milfont TL, Duckitt J. The environmental attitudes inventory: a valid and reliable measure to assess the structure of environmental attitudes. J Environ Psychol 2010 Mar;30(1):80-94. [doi: 10.1016/j.jenvp.2009.09.001]

54. Streiner D, Norman G. Health Measurement Scales: A Practical Guide to Their Development and Use. Oxford, United Kingdom: Oxford University Press; 2008.
55. De Jesus M, Xiao C. Predicting health care utilization among Latinos: health locus of control beliefs or access factors? Health Educ Behav 2014 Aug 04;41(4):423-430. [doi: 10.1177/1090198114529130] [Medline: 24707031]
56. Tomás-Sábado J, Montes-Hidalgo J. [Spanish version of the Multidimensional health locus of control scale innursing students]. Enferm Clin 2016;26(3):181-187. [doi: 10.1016/j.enfcli.2015.12.005] [Medline: 26922046]
57. Wallston KA, Wallston BS, DeVellis R. Development of the Multidimensional Health Locus of Control (MHLC) scales. Health Educ Monogr 1978;6(2):160-170. [doi: 10.1177/109019817800600107] [Medline: 689890]
58. An innovative training program in Latin America for maternal health improvement. MundoComm. URL: http://mundocomm.org/ [accessed 2021-08-11]
59. IBM SPSS Statistics, Version 25. IBM Corp. 2018. URL: https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-25 [accessed 2021-08-11]
60. Hosmer JD, Lemeshow S, Sturdivant R. Applied Logistic Regression, 3rd Edition. Hoboken, New Jersey: Wiley; Mar 2013.
61. Christensen AJ, Turner CW, Smith TW, Holman JM, Gregory MC. Health locus of control and depression in end-stage renal disease. J Consult Clin Psychol 1991 Jun;59(3):419-424. [doi: 10.1037//0022-006x.59.3.419] [Medline: 2071727]
62. Poortinga W, Dunstan FD, Fone DL. Health locus of control beliefs and socio-economic differences in self-rated health. Prev Med 2008 Apr;46(4):374-380. [doi: 10.1016/j.ypmed.2007.11.015] [Medline: 18177930]
63. López-Alicea K. Montaña de basura al lado de playa Tamarindo en Culebra. Montaña De Basura Al Lado De Playa Tamarindo en Culebra. URL: https://www.cesampr.com/blog/montaa-de-basura-al-lado-de-playa-tamarindo-en-culebra [accessed 2021-08-11]
64. Varsavky M. Vivir entre la basura en Culebra, Puerto Rico. MartinVarsavsky.net. URL: https://spanish.martinvarsavsky.net/general/vivir-entre-la-basura-en-culebra.html [accessed 2021-08-11]
65. Siguen los daños ambientales en Playa Flamenco, Culebra. PrimeraHora.Com. URL: https://www.primerahora.com/noticias/puerto-rico/notas/siguen-los-danos-ambientales-en-playa-flamenco-culebra/ [accessed 2021-08-11]
66. Landauer J. Culebrans fire back: islanders seek to end role as a Navy target - shells fall close to governor of Puerto Rico. Wall St J 1970 Jun 10:A [FREE Full text]
67. Pérez-Rodríguez T. Movilización Social y Política Contra la Presencia de la Marina en Culebra 1960-1975. Publicaciones Puertorriqueñas 2013:A [FREE Full text]
68. Carrasco C. Munitions cleanup on Culebra advances, increasing safety for public recreational use. Defense Visual Information Distribution Service. URL: https://www.dvidshub.net/news/333084/munitions-cleanup-culebra-advances-increasing-safety-public-recreational-use [accessed 2021-08-11]
69. Skinner EA. A guide to constructs of control. J Pers Soc Psychol 1996 Sep;71(3):549-570. [doi: 10.1037//0022-3514.71.3.549] [Medline: 8831161]
70. Bosma H, van de Mheen HD, Borsboom G, Mackenbach J. Neighborhood socioeconomic status and all-cause mortality. Am J Epidemiol 2001 Feb 15;153(4):363-371. [doi: 10.1093/aje/153.4.363] [Medline: 11207154]
71. Protestan culebrenses por servicio de lanchas. Elnuevodia.com. URL: https://www.elnuevodia.com/noticias/locales/nota/protestanculebrensesporserviciodelanchas-779136/ [accessed 2021-08-11]
72. Ruiz-Kuilan G. Residentes de las islas municipio denuncian que la transportación ha empeorado. Elnuevodia.com. URL: https://www.elnuevodia.com/noticias/locales/nota/residentesdelasislasmunicipiodenuncianquelatransportacionhaempeorado-2472156/ [accessed 2021-08-11]
73. Gordon N. 55+ jaw dropping app usage statistics in 2021 [infographic]. Techjury. URL: https://techjury.net/stats-about/app-usage/ [accessed 2021-08-11]
74. Böhmer M, Hecht B, Schöning J, Krüger A, Bauer G. Falling asleep with Angry Birds, Facebook and Kindle: a large scale study on mobile application usage. In: Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services. 2011 Aug Presented at: MobileHCI '11: 13th International Conference on Human Computer Interaction with Mobile Devices and Services; Aug 30-Sep 2, 2011; Stockholm Sweden p. 47-56. [doi: 10.1145/2037373.2037383]
75. Lim S, Bentley P. Investigating app store ranking algorithms using a simulation of mobile app ecosystems. In: Proceedings of the IEEE Congress on Evolutionary Computation. 2013 Presented at: IEEE Congress on Evolutionary Computation; June 20-23, 2013; Cancun, Mexico. [doi: 10.1109/cec.2013.6557892]
76. Shepard C, Rahmati A, Tossell C, Zhong L, Kortum P. LiveLab: measuring wireless networks and smartphone users in the field. SIGMETRICS Perform Eval Rev 2010 Dec;38(3):15-20. [doi: 10.1145/1925019.1925023]
77. O'Connell C. 23% of users abandon an app after one use. DZone. URL: https://dzone.com/articles/23-of-users-abandon-an-app-after-one-use [accessed 2021-05-11]
78. The State of Mobile 2021 Report. AppAnnie, San Francisco, California, USA. URL: https://www.appannie.com/en/go/state-of-mobile-2021/ [accessed 2021-08-11]
79. Worthy SL, Downey L, Dyk PH, Monroe PA, Tyler-Mackey C, Welborn R. Turning the tide on poverty: community champions as critical elements of success in economically distressed communities. Community Dev 2016 Mar 31;47(3):341-357. [doi: 10.1080/15575330.2016.1167096]

80. Shields-Zeeman L, Pathare S, Walters BH, Kapadia-Kundu N, Joag K. Promoting wellbeing and improving access to mental health care through community champions in rural India: the intervention approach. Int J Ment Health Syst 2017 Jan 4;11:6 [FREE Full text] [doi: 10.1186/s13033-016-0113-3] [Medline: 28066505]
81. Leadbeater BJ. Engaging Community Champions in the Prevention of Bullying: Understanding and Addressing Bullying: An International Perspective. Bloomington: AuthorHouse; 2008.
82. United States Department of Health and Human Services. Healthy people 2030 framework. Office of Disease Prevention and Health Promotion. URL: https://www.healthypeople.gov/2020/About-Healthy-People/Development-Healthy-People-2030/Framework [accessed 2020-04-04]
83. Uddin J, Biswas T, Adhikary G, Ali W, Alam N, Palit R, et al. Impact of mobile phone-based technology to improve health, population and nutrition services in rural Bangladesh: a study protocol. BMC Med Inform Decis Mak 2017 Jul 06;17(1):101 [FREE Full text] [doi: 10.1186/s12911-017-0502-9] [Medline: 28683742]
84. Peprah P, Abalo EM, Agyemang-Duah W, Budu HI, Appiah-Brempong E, Morgan AK, et al. Lessening barriers to healthcare in rural Ghana: providers and users' perspectives on the role of mHealth technology. A qualitative exploration. BMC Med Inform Decis Mak 2020 Feb 10;20(1):27 [FREE Full text] [doi: 10.1186/s12911-020-1040-4] [Medline: 32041608]
85. Czyzewski K. Colonialism as a broader social determinant of health. Int Indig Policy J 2011 May;2(1):5. [doi: 10.18584/iipj.2011.2.1.5]
86. Rodríguez-Díaz CE. Community resistance as public health activism in Puerto Rico. Am J Public Health 2020 Oct;110(10):1454-1455. [doi: 10.2105/AJPH.2020.305884] [Medline: 32903086]

Abbreviations

¡mZAP!: Zonas, Acción y Protección
a-EAI: adapted environmental attitudes inventory
FAITH!: Fostering African American Improvement in Total Health
ICT: information and communication technology
mHealth: mobile health
MHLC: multidimensional health locus of control
OR: odds ratio
REDCap: Research Electronic Data Capture
SCT: social cognitive theory

Edited by R Kukafka; submitted 29.09.20; peer-reviewed by Z Quiñones, YC Chen; comments to author 13.02.21; revised version received 27.02.21; accepted 19.07.21; published 23.09.21.

Please cite as:
Perez-Ramos JG, McIntosh S, Barrett ES, Velez Vega CM, Dye TD
Attitudes Toward the Environment and Use of Information and Communication Technologies to Address Environmental Health Risks in Marginalized Communities: Prospective Cohort Study
J Med Internet Res 2021;23(9):e24671
URL: https://www.jmir.org/2021/9/e24671
doi:10.2196/24671
PMID:34554103

Original Paper

Patterns for Patient Engagement with the Hypertension Management and Effects of Electronic Health Care Provider Follow-up on These Patterns: Cluster Analysis

Dan Wu[1*], BSc; Jiye An[1*], PhD; Ping Yu[2], PhD; Hui Lin[1], BSc; Li Ma[3], MSc; Huilong Duan[1], PhD; Ning Deng[1], PhD

[1]College of Biomedical Engineering and Instrument Science, Ministry of Education Key Laboratory of Biomedical Engineering, Zhejiang University, Hangzhou, China

[2]School of Computing and Information Technology, Faculty of Engineering and Information Sciences, University of Wollongong, Wollongong, Australia

[3]General Hospital of Ningxia Medical University, Yinchuan, China

[*]these authors contributed equally

Corresponding Author:
Ning Deng, PhD
College of Biomedical Engineering and Instrument Science
Ministry of Education Key Laboratory of Biomedical Engineering
Zhejiang University
Zhouyiqing Bldg 512, Yuquan Campus
38 Zheda Rd
Hangzhou, 310027
China
Phone: 86 571 2295 269
Fax: 86 571 2295 269
Email: zju.dengning@gmail.com

Abstract

Background: Hypertension is a long-term medical condition. Electronic and mobile health care services can help patients to self-manage this condition. However, not all management is effective, possibly due to different levels of patient engagement (PE) with health care services. Health care provider follow-up is an intervention to promote PE and blood pressure (BP) control.

Objective: This study aimed to discover and characterize patterns of PE with a hypertension self-management app, investigate the effects of health care provider follow-up on PE, and identify the follow-up effects on BP in each PE pattern.

Methods: PE was represented as the number of days that a patient recorded self-measured BP per week. The study period was the first 4 weeks for a patient to engage in the hypertension management service. K-means algorithm was used to group patients by PE. There was compliance follow-up, regular follow-up, and abnormal follow-up in management. The follow-up effect was calculated by the change in PE (CPE) and the change in systolic blood pressure (CSBP, SBP) before and after each follow-up. Chi-square tests and *z* scores were used to ascertain the distribution of gender, age, education level, SBP, and the number of follow-ups in each cluster. The follow-up effect was identified by analysis of variances. Once a significant effect was detected, Bonferroni multiple comparisons were further conducted to identify the difference between 2 clusters.

Results: Patients were grouped into 4 clusters according to PE: (1) PE started low and dropped even lower (PELL), (2) PE started high and remained high (PEHH), (3) PE started high and dropped to low (PEHL), and (4) PE started low and rose to high (PELH). Significantly more patients over 60 years old were found in the PEHH cluster ($P \leq .05$). Abnormal follow-up was significantly less frequent ($P \leq .05$) in the PELL cluster. Compliance follow-up and regular follow-up can improve PE. In the clusters of PEHH and PELH, the improvement in PE in the first 3 weeks and the decrease in SBP in all 4 weeks were significant after follow-up. The SBP of the clusters of PELL and PELH decreased more (–6.1 mmHg and –8.4 mmHg) after follow-up in the first week.

Conclusions: Four distinct PE patterns were identified for patients engaging in the hypertension self-management app. Patients aged over 60 years had higher PE in terms of recording self-measured BP using the app. Once SBP reduced, patients with low PE tended to stop using the app, and a continued decline in PE occurred simultaneously with the increase in SBP. The duration

and depth of the effect of health care provider follow-up were more significant in patients with high or increased engagement after follow-up.

(J Med Internet Res 2021;23(9):e25630) doi:10.2196/25630

KEYWORDS

hypertension; health care services; mHealth; patient engagement; electronic follow-up; cluster analysis

Introduction

Background

Hypertension is a lifestyle-induced chronic disease that affects health-related quality of life. Once the disease deteriorates, it may cause complications such as stroke, myocardial infarction, cardiac failure, and renal failure [1]. Controlling hypertension requires patients to follow long-term self-management plans, including measuring and recording blood pressure (BP), taking medicine, conducting physical activities, and healthy eating. These plans should preferably be established with guidance from the health care providers and be followed continuously for hypertension control [2].

The rapid pace of electronic and mobile technology development has promoted the implementation of out-of-hospital health care services [3-5]. Mobile health (mHealth) service is defined as the use of mobile phone and wireless technologies to support medical and public health care services [6]. Abundant evidence supports the introduction of mHealth services into chronic disease control and promoting positive patient behavior [4,7-10]. Although recent studies have demonstrated efficacy for some mHealth services, some have performed poorly [11-14]. Thus, there are mixed results for using mHealth services to support patient self-management of hypertension in the community [15,16].

Patient engagement (PE) is essential for bringing an improvement of health outcomes using the mHealth app [11,17,18]. PE refers to the activation of a patient to engage with the interventions that are designed to promote positive health behavior [19]. In a study of a hypertension self-management app, Toto-Ramos et al [20] found that the patients with sustained PE, as measured by the number of weeks that the patients engaged with the app, experienced significant reduction in systolic blood pressure (SBP). Goyal et al [21] evaluated the behavior of daily blood glucose reading of patients and found a significant relationship between increased number of readings and improved glycated hemoglobin. In comparison with the traditional methods, mHealth service is advantageous in improving PE for self-management of hypertension [22,23]. For example, Kaplan et al [24] designed an mHealth app to facilitate PE in managing hypertension. The app provides a mobile platform for recording and tracking self-measured BP, periodic reminders to measure BP, and wireless BP measurement devices. Their research resulted in significant improvement in BP for patients and a high level of PE.

In the mHealth field, complex PE patterns emerge when mobile apps are used for self-management. Some patients register but never use the app, some use it intermittently, and others frequently use it for a long period of time. PE can be objectively measured as amount, duration, breadth, and depth of using the mHealth app [11], and it can also be subjectively measured as interest, intrigue, focus, inattention, enjoyment, pleasure, etc [25]. In a study of pain management with a mobile app, Rahman et al [26] used a clustering technique to identify PE patterns with the app. They measured PE by 3 key usage features: duration of app use, frequency of app use, and the number of usage records. They then used a K-means clustering algorithm to find groups of patients as indicated by PE because the PE pattern reflects the behavioral characteristics of patients using the app. Sanatkar et al [27] also used a clustering technique to analyze the mHealth app usage data so as to distinguish PE with the app in an e-mental health community. They measured 5 usage features of PE: number of user logins, number of daily trackers used, number of learning activities started and completed, and number of reminders received. Distinct usage patterns were observed in the frequency of using the app.

To date, the studies on patterns of PE and the associations between these PE patterns and health outcomes have usually used various types of static count data (eg, number of logins and number of records). However, PE is a dynamic, ongoing process [28]. The change of this dynamic process cannot be captured simply by analyzing count data captured at a single date point in the cross-sectional data analysis. Longitudinal change is useful for identifying trends through analyzing the time series data. Furthermore, as hypertension management requires long-term efforts, understanding the trends of PE is important for the long-term successful management of hypertension.

Health care provider follow-up can offer continuous and personalized attention to guide patient's self-management behavior in response to the patient's current BP level [29]. Without health care provider support, it is difficult for patients to maintain BP control and high levels of PE with an mHealth app [30], which has often resulted in a high level of dropout [31,32]. Therefore, follow-up by health care providers is recommended by the hypertension management guideline in order to know, track, and intervene in patient's hypertension self-management in the long-term [33-35]. In the traditional outpatient management model, follow-up often takes the form of home visits by health care providers to promote PE and then achieve BP control. Despite its proven benefits for hypertension management, follow-up is not fully implemented by health care providers due to the high human resources required for patient follow-up.

In comparison with the traditional format of follow-up, mHealth health care provider follow-up is a low-cost, convenient means of follow-up, and patient's BP, medication, exercise, diet, etc., can be checked in the office. In a series of our previous studies into hypertension management service [36,37], after receiving

the BP data that a patient entered through the mobile app, the server could automatically conduct BP data analysis, with the web-based platform reporting the patient's BP condition to the health care provider and reminding the health care provider which patients needed to follow-up. In response to different patient health statuses, health care providers can conduct follow-up through calling and sending short SMS text messages. Our previous study designed 3 types of follow-ups: compliance follow-up, regular follow-up, and abnormal follow-up [36,37]. These were conducted to improve compliance, track BP status, redesign a new plan, and respond to the abnormal BP of patients.

To date, few studies have explored the effect of health care provider follow-up on PE and BP control in mHealth services for hypertension management [23]. It is not clear if follow-up is effective, which type of follow-up is effective, when the follow-up effect emerges, if there is a differential effect of follow-up in different patient groups, and how long the follow-up effect lasts [28,38]. Answers to these questions are essential for informing the implementation of value-based health care and providing optimal outcomes in improving the quality and reducing cost of hypertension management.

Objective

The aim of this study was to explore the patterns of PE with a hypertension management service and the effect of health care provider follow-up on PE and BP within the first 4 weeks of mHealth app usage. This included 3 objectives: (1) to discover the patterns of PE with the mHealth app and the association between a series of related variables (age, gender, education level, the mean SBP, and the number of follow-ups) and the PE pattern, (2) to explore the health care provider follow-up effect on PE, and (3) to examine the effect of follow-up on SBP in each PE pattern.

Methods

Description of the mHealth App

Blood Pressure Assistant (BPA) is a mHealth hypertension self-management app available for patients in the General Hospital of Ningxia Medical University [36,37,39]. BPA was launched in 2015, and since then, 2129 patients have registered to use the app. The mHealth app was designed in accordance with a customized care pathway in compliance with the Chinese guideline for hypertension management. The care pathway defines tasks for hypertension management for patients and health care providers.

In this care pathway, each patient is required to register and provide basic demographic information (ie, name, gender, date of birth, and education level). The patient is then assigned to a health care provider, who is responsible for formulating a tailored management plan, conducting follow-up, and supervising patients' uploaded data. The management plan includes the frequency of self-measured and recorded BP, and recommendations for medication, physical activities, and diet. The patient's uploaded data includes self-measured BP, medications taken, physical activities, and diet records. The most important task for health care providers is to track patient's current BP level through the web-based platform and use mobile phones for patient follow-ups to assist in BP control. There are 3 types of follow-ups. (1) Compliance follow-up is performed as a response to the identified low patient compliance with the hypertension management plan [36]. In this case, health care providers need to remind the patient to measure and record BP on a regular basis. (2) Regular follow-up is conducted on regular basis in order to check the BP level and decide whether or not to maintain or update the management plan [37]. (3) Abnormal follow-up is required when patient's self-measured BP data are abnormal [36]. Health care providers need to check and understand the causes of abnormal conditions and intervene in a timely manner. With consent, the health care provider enrolls the patient they manage into the mHealth hypertension management program in an online community. Patients can use the app to check their self-management plans. Their main task is to measure and record BP data on a daily basis.

Data Collection

Statement of Ethics

Ethical approval was granted by the Ethic Committee for the Conduct of Human Research at General Hospital of Ningxia Medical University (#NXMU-GH-2017-273). Patients in this study signed the informed consent forms.

Sample

All data were stored and extracted from the BPA server, which contains the demographic information, self-management plans, patient-uploaded data, and follow-up records of health care providers. The primary data set included 2129 patients. We selected the patient records based on the following criteria: patients were at least 18 years of age; patients registered to the app between March 27, 2016, and July 10, 2019 (as the main functions of the app were consistent during this period, this could ensure that the patient's behavior was not affected by the changes in app functions); and patients continuously measured and recorded BP more than 4 weeks after registration (as the control of BP requires at least a 4-week observation period according to the relevant guidelines) [33-35].

Data Extraction

We extracted 3 types of data from the database: demographics, BP records, and follow-up records. The demographics information included the patient identification, data of birth, gender (male and female), and education level (primary or secondary school, high school, or university and above). The BP record data included SBP, diastolic blood pressure, and the uploaded date (containing year, month, day, minute, and second). Due to the high correlation between SBP and diastolic blood pressure, only SBP was used for analysis. The information specific to the follow-up records included patient ID, follow-up type (compliance follow-up, regular follow-up, and abnormal follow-up), and follow-up date.

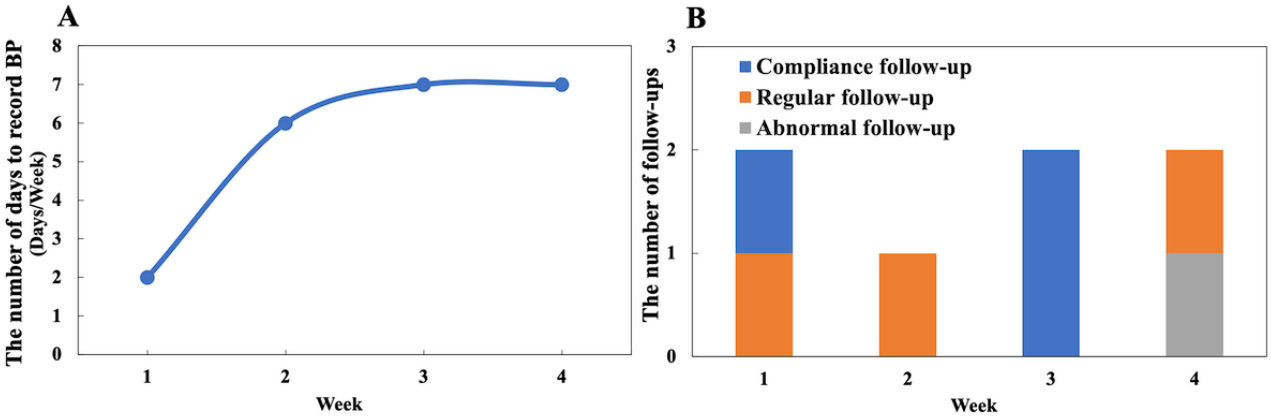
Data Analysis

Measuring PE

In accordance with the hypertension management guidelines [33-35], we defined the time unit of observation as 1 week. The behavior of measuring BP is a basic behavior in hypertension

management because other behaviors (eg, taking medicine, doing physical activities, and eating healthy) need to be based on patient's current BP level. Therefore, in this study, patient engagement was indicated by the number of days that a patient recorded self-measured BP per week (see Figure 1). The analysis period was the first 4 consecutive weeks after the initial patient registration.

Figure 1. Example of patient engagement and follow-up for a patient within 4 weeks. (A) Example of patient engagement. (B) Example of the number of different types of follow-ups. BP: blood pressure.



Cluster Analysis

Cluster analysis was conducted to group the patients into different PE patterns. K-means algorithm (Python 3; Python Software Foundation) was used to cluster patients by PE [40]. Dynamic time warping was used to calculate the similarity of PE between patients because this feature was represented as time series data [41]. We then used silhouette score to determine the optimal number of clusters [42]. The silhouette score measured the distance between clusters based on the distance between the PE of patients as determined by dynamic time warping. A higher silhouette score indicates tighter clusters, where each cluster is completely separate from the others.

Characterizing the Clusters

An optimal clustering result was reached based on the silhouette score of a different number of clusters. For each cluster, we analyzed the demographic features (the distribution of age, gender, and education level); BP features, including the distribution of mean SBP in the first week (representing the initial BP conditions in hypertension management) and the trend of weekly mean SBP; and follow-up features (the distribution of the number of the 3 types of follow-ups in each of the 4 weeks; see Figure 1).

The Effect of 3 Types of Follow-ups on PE and SBP

The follow-up effect was calculated by the change in PE (CPE) and the change in SBP (CSBP) before and after each follow-up, which were defined as follows:

where PE_{i+1} is PE in the week after a follow-up event, PE_i is PE in the week of the follow-up even, SPB_{i+1} is the mean SBP in the week after a follow-up event, SPB_i is the mean SBP in the week of the follow-up event, and N is the number of follow-ups performed by health care providers.

Statistical analysis was conducted in SPSS version 24 (IBM Corp). Chi-square test was performed to evaluate the statistical significance of associations between the clusters and the discrete variables (demographic, BP, and the number of follow-ups). Pairwise comparisons were conducted using the z scores to compare difference in the proportion of discrete variables between the clusters. Analysis of variance (ANOVA) was employed to analyze the changes in outcome parameters (CPE and CSBP) after follow-up in each of the 4 weeks. Once a significant change was detected, Bonferroni multiple comparisons were further conducted to examine the differences between the clusters. A P value <.05 was used to determine whether the difference was statistically significant.

Results

PE Patterns

A total of 562 patients met the selection criteria and were included in the study. We found that the silhouette score was the highest with 4 clusters of patients (see Figure 2). Hence, we accepted the 4-cluster output of K-means (see Table 1) for further analysis. The PE of each cluster within 4 weeks was significantly different (P<.001).

Figure 2. Comparison of the silhouette score for the different numberd of clusters (range from 3 to 7).

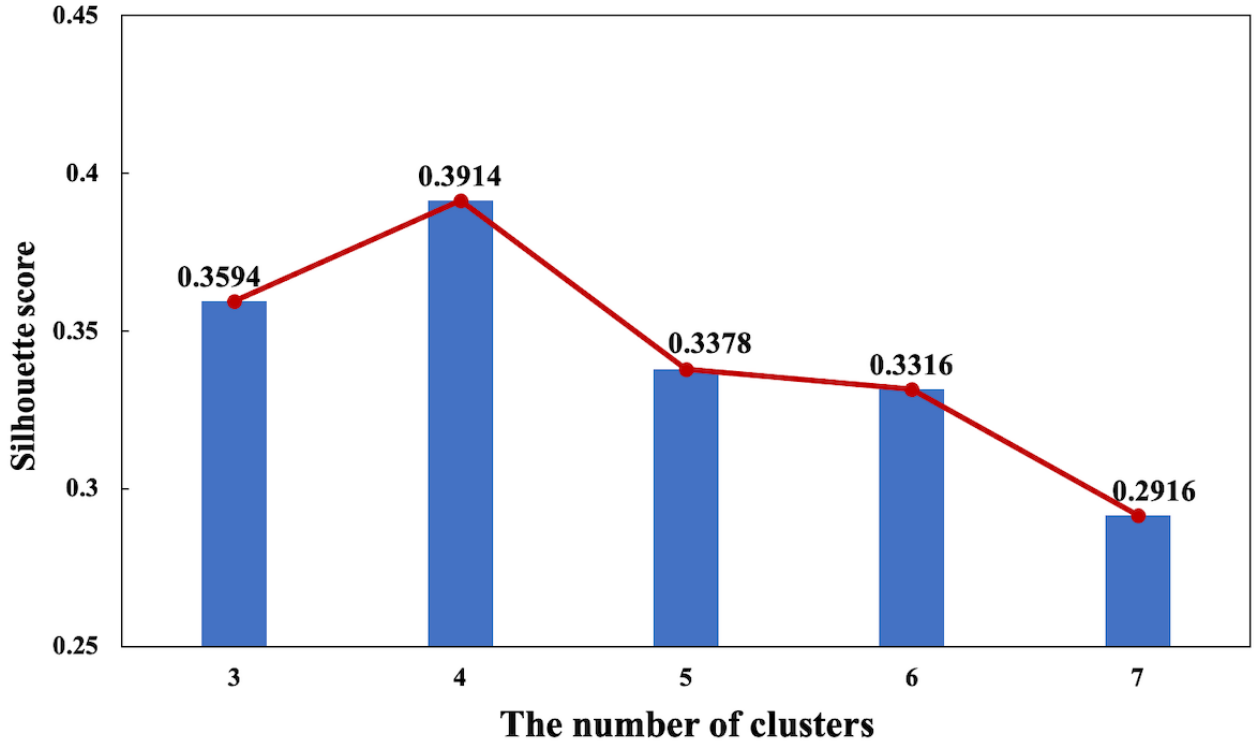


Table 1. Comparison of PE in the 4 clusters.

PE[a] (Days/week)	PELL[b], mean (SD) (n=183)	PEHH[c], mean (SD) (n=142)	PEHL[d], mean (SD) (n=148)	PELH[e], mean (SD) (n=89)	*P* value
Week 1	1.9 (1.1)	6.4 (1.1)	5.8 (1.2)	2.4 (1.3)	<.001
Week 2	0.8 (1.3)	6.6 (0.8)	4.8 (1.9)	3.5 (2.0)	<.001
Week 3	0.8 (1.3)	6.4 (0.9)	3.2 (2.1)	4.3 (2.0)	<.001
Week 4	0.6 (0.8)	6.3 (0.9)	1.8 (1.3)	4.6 (1.5)	<.001

[a]PE: patient engagement.

[b]PELL: patient engagement started low and dropped even lower.

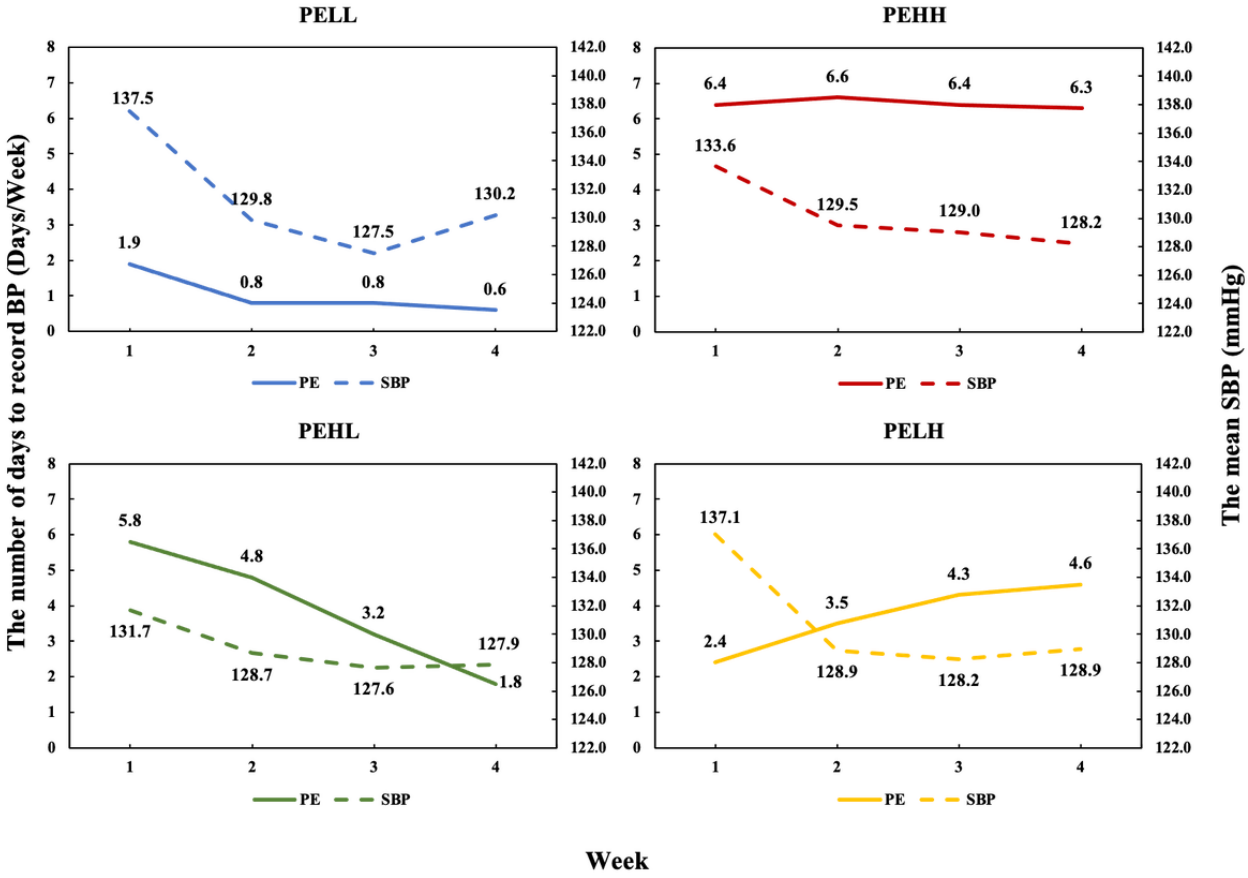
[c]PEHH: patient engagement started high and remained high.

[d]PEHL: patient engagement started high and dropped to low.

[e]PELH: patient engagement started low and rose to high.

There were 4 distinctive change patterns of PE (see Figure 3). The first cluster contained 183 patients. They started recording their BP 1.9 days/week (SD 1.1 days/week) and then decreased every week to around 0.6 days/week (SD 0.8 d days/week) in the fourth week. We referred this cluster as “PE started low and dropped even lower” (PELL) cluster. The second cluster contained 142 patients who were consistently active engaging in BP recording throughout the whole period. On average they recorded the BP for more than 6 days per week. Therefore, we referred them as the “started high and remained high” (PEHH) cluster. The third cluster contained 148 patients who started with a high level of recording (5.8 days/week, SD 1.2 days/week) that then decreased every week to around 1.8 days/week (SD 1.3 days/week) in the fourth week. We referred them as the “PE started high and dropped to low” (PEHL) cluster. The fourth cluster contained 89 patients who began recording self-measured BP 2.4 days/week (SD 1.3 days/week) and gradually increased every week to around 4.6 days/week (SD 1.5 days/week) in the fourth week. We referred them as the “PE started low and then rose to high” (PELH) cluster.

Figure 3. The patient engagement patterns of the 4 clusters and the 4-week mean SBP curve of the 4 clusters. BP: blood pressure; SBP: systolic blood pressure. PELL: patient engagement started low and dropped even lower; PEHH: patient engagement started high and remained high; PEHL: patient engagement started high and dropped to low; PELH: patient engagement started low and rose to high.



The Demographics, SBP, and Follow-up Characteristics of the PE Patterns

The Association Between Demographic Factors and the PE Pattern

The distribution of age, gender, and education level in each cluster is presented in Table 2. The majority of patients (387/562, 68.9%) were between 40 and 60 years old. Chi-square test found a statistically significant association between age and the PE pattern (P=.001). There was no significant difference in the distribution of patients under 60 years old in each PE pattern, but there were many more patients aged over 60 years old in the PEHH cluster than in the other clusters (P<.05).

Chi-square test did not find a statistically significant association between the PE pattern and gender or education level. There were many more male patients (338/562, 60.1%) using the BPA than female patients. Moreover, 50.5% (284/562) of patients had a university degree.

Table 2. The distribution of demographic characteristics in each cluster.

Characteristic	PELL[a] (n=183)	PEHH[b] (n=142)	PEHL[c] (n=148)	PELH[d] (n=89)	*P* value
Age, n (%)					.001
18<age≤40	24 (13.1)	14 (9.8)	27 (18.2)	11 (12.4)	
40<age≤60	131 (71.6)	86 (60.6)	104 (70.3)	66 (74.1)	
Age>60	28 (15.3)	42 (29.6)	17 (11.5)	12 (13.5)	
Gender, n (%)					.37
Male	102 (55.7)	84 (59.2)	95 (64.2)	57 (64.0)	
Female	81 (44.3)	58 (40.8)	53 (35.8)	32 (36.0)	
Education, n (%)					.45
Primary or secondary school	14 (7.6)	6 (4.2)	5 (3.4)	2 (2.2)	
High school	79 (43.2)	61 (43.0)	68 (45.9)	43 (48.3)	
University and above	90 (49.2)	75 (52.8)	75 (50.7)	44 (49.4)	

[a]PELL: patient engagement started low and dropped even lower.

[b]PEHH: patient engagement started high and remained high.

[c]PEHL: patient engagement started high and dropped to low.

[d]PELH: patient engagement started low and rose to high.

The Association Between the Initial BP Conditions and the PE Pattern

The mean SBP in the first week represented the initial BP conditions in hypertension management. The chi-square test showed a statistically significant association between the mean SBP in the first week and the PE pattern (*P*<.001; see Table 3). *z* scores identified a significantly higher proportion of patients with high SBP (between 160 mmHg and 179 mmHg) in the PELL cluster in the first week than in the PEHL cluster (*P*<.05). The majority of patients (316/562, 56.2%) had a mean SBP between 120 mmHg and 139 mmHg in the first week.

Table 3. The distribution of the first week mean SBP in each cluster (*P*<.001).

SBP[a]	PELL[b], n (%) (n=183)	PEHH[c], n (%) (n=142)	PEHL[d], n (%) (n=148)	PELH[e], n (%) (n=89)
SBP<120	19 (10.4)	13 (9.2)	23 (15.5)	7 (7.9)
120≤SBP≤139	90 (49.2)	91 (64.1)	91 (61.5)	44 (49.4)
140≤SBP≤159	53 (29.0)	35 (24.6)	32 (21.6)	34 (38.2)
160≤SBP≤179	16 (8.7)	3 (2.1)	2 (1.4)	4 (4.5)
SBP≥180	5 (2.7)	0 (0)	0 (0)	0 (0)

[a]SBP: systolic blood pressure.

[b]PELL: patient engagement started low and dropped even lower.

[c]PEHH: patient engagement started high and remained high.

[d]PEHL: patient engagement started high and dropped to low.

[e]PELH: patient engagement started low and rose to high.

The Trend of Weekly Mean SBP in Each PE Pattern

Across 4 weeks of management, the mean SBP of all 4 clusters declined at different rates (see Figure 3). Pairwise comparisons of the mean SBP between 4 weeks in each cluster revealed that in the clusters of PELL (*P*=.002), PEHH (*P*=.006), and PELH (*P*<.001), the mean SBP had significantly reduced from the second week. In the PEHL cluster, the significant reduction of SBP appeared from the third week (*P*=.02). The PELL cluster experienced a greater reduction in the mean SBP in the third week compared to the other clusters, but the SBP of users in this cluster increased more than did the others in the fourth week.

The Number of Follow-ups in Each PE Pattern

The amount of the 3 types of follow-ups were significantly different (*P*=.03) in the 4 clusters (see Table 4). Abnormal follow-up was performed significantly less frequently (*P*<.05) in the PELL cluster than in the clusters of PEHH and PEHL. The distribution of follow-up in each of the 4 weeks was not significantly different in the 4 clusters. Most (338/680, 49.7%) follow-ups were performed in the first week.

Table 4. The distribution of the number of 3 types of follow-ups and the number of follow-ups in each of the 4 weeks in each cluster.

Value	PELL[a] (n=183)	PEHH[b] (n=142)	PEHL[c] (n=148)	PELH[d] (n=89)	*P* value
Type, n (%)					.02
Compliance	77 (46.1)	53 (31.4)	74 (39.2)	57 (41.9)	
Regular	80 (47.9)	87 (51.5)	85 (45.0)	61 (44.9)	
Abnormal	10 (6.0)	29 (17.2)	30 (15.9)	18 (13.2)	
Time, n (%)					.11
Week 1	90 (51.7)	86 (50.3)	96 (49.2)	66 (47.1)	
Week 2	50 (28.7)	44 (25.7)	40 (20.5)	36 (25.7)	
Week 3	13 (7.5)	14 (8.2)	23 (11.8)	23 (16.4)	
Week 4	21 (12.1)	27 (15.8)	36 (18.5)	15 (10.7)	

[a]PELL: patient engagement started low and dropped even lower.

[b]PEHH: patient engagement started high and remained high.

[c]PEHL: patient engagement started high and dropped to low.

[d]PELH: patient engagement started low and rose to high.

The Effect of Follow-up on PE and SBP in Each PE Pattern

The Effect of Each Type of Follow-up on PE

The results of ANOVA revealed the significant main effects of cluster (F=15.09; P<.001) and type (F=5.442, P=.005), and the cluster–type interaction effect (F=2.60; P=0.02) on PE (see Table 5). Bonferroni multiple comparisons found that after the 3 types of follow-up, the CPE between the 4 clusters had the following statistically significant differences: (1) for the compliance follow-up, PE in the PELL cluster had a significantly larger drop than did the PEHH (P=.02) and PELH (P<.001) clusters, and improvement in PE in the PELH cluster was significantly higher than that of the PELL (P<.001) and PEHL (P<.001) clusters; (2) for the regular follow-up, PE in the PEHL cluster had a significantly larger drop than did the clusters of PELL (P=.01), PEHH (P=.01), and PELH (P<.001); and (3) for the abnormal follow-up, the decrease in PE in the PEHL cluster was significantly lower than that of the PEHH cluster (P=.01). Compliance follow-up improved PE in the PEHH and PELH clusters. Regular follow-up improved PE in the PELH cluster.

Table 5. The effect of 3 types of follow-ups on patient engagement and the effect of follow-up in each of the 4 weeks on PE. Multiple comparisons of the 4 clusters (at the .05 level).

CPE[a] (Days/week)	PELL[b], mean (SD) (n=183)	PEHH[c], mean (SD) (n=142)	PEHL[d], mean (SD) (n=148)	PELH[e], mean (SD) (n=89)	*P* value
Type					
Compliance	–0.6 (1.7)	0.5 (1.5)[f]	–0.3 (2.5)	1.31 (2.3)[f,h]	<.001
Regular	–0.1 (1.2)	–0.2 (1.4)	–1.1 (2.3)[f,g]	0.67 (2.7)[h]	<.001
Abnormal	–0.1 (1.0)	–0.1 (1.4)	–1.8 (2.3)[g]	–0.39 (2.3)	.01
Time					
Week 1	–0.7 (1.6)	0.3 (1.7)[f]	–0.8 (2.3)[g]	1.2 (2.5)[f,g,h]	<.001
Week 2	0.1 (1.2)	0 (0.9)	–1.5 (2.7)[f,g]	1.4 (2.1)[f,g,h]	<.001
Week 3	–0.2 (1.3)	0.4 (1.0)	–1.6 (2.1)[g]	0.3 (2.5)[h]	.01
Week 4	–0.2 (1.4)	–0.7 (1.7)	–0.1 (2.7)	–1.5 (2.5)	.10

[a]CPE: change in patient engagement.

[b]PELL: patient engagement started low level and dropped even lower.

[c]PEHH: patient engagement started high and remained high.

[d]PEHL: patient engagement started high and dropped to low.

[e]PELH: patient engagement started low and rose to high.

[f]The given cluster is significantly different from the PELL cluster.

[g]The given cluster is significantly different from the PEHH cluster.

[h]The given cluster is significantly different from the PEHL cluster.

The Effect of Follow-up on PE in Each of the 4 Weeks

The results of ANOVA revealed the significant main effects of the cluster (F=10.335; P<.001) and time (F=2.870; P=.04), and the cluster–time interaction effect (F=5.168; P<.001) on PE (see Table 5). Bonferroni multiple comparisons found that after the follow-up in each of the 4 weeks, the CPE between the 4 clusters had the following statistically significant differences: in week 1, the improvement in PE in the PELH cluster was significantly higher than that of the clusters of PELL (P<.001), PEHH (P=.03), and PEHL (P<.001); in week 2, the PE in the PEHL cluster had a significantly larger drop than that of the clusters of PELL (P=.001), PEHH (P=.01), and PELH (P<.001), and the improvement in PE in the PELH was significantly higher than that of the clusters of PELL (P=.01), PEHH (P=.01), and PEHL (P<.001); and in week 3, the PE in the PEHL cluster had a significantly larger drop than that of the PEHH (P=.02) and PELH (P=.01) clusters. Follow-up improved PE in the PEHH and PELH clusters in first 3 weeks and only in the second week in the PELL cluster, but had no effect on the PEHL cluster in any of the 4 weeks.

The Effect of Each Type of Follow-up on SBP

The results of ANOVA revealed significant main effects of cluster (F=2.789; P=.04) and type (F=1.137; P=.32), but no cluster–type interaction (F=0.956; P=.46) effect on SBP (see Table 6). The mean SBP of all 4 clusters of patients reduced after 3 types of follow-ups except for the abnormal follow-up in the PELL cluster. Bonferroni multiple comparisons showed that after the compliance follow-up, the PELH cluster had a significantly higher level of SBP decline than did the PEHL cluster (P=.003).

Table 6. The effect of 3 types of follow-ups on systolic blood pressure and the effect of follow-up in each of the 4 weeks on systolic blood pressure. Multiple comparisons of the 4 clusters (at the .05 level).

CSBP[a] (mmHg)	PELL[b], mean (SD) (n=183)	PEHH[c], mean (SD) (n=142)	PEHL[d], mean (SD) (n=148)	PELH[e], mean (SD) (n=89)	*P* value
Type					
Compliance	–3.4 (11.4)	–2.5 (5.8)	–0.9 (6.2)	–6.7 (10.5)[f]	.01
Regular	–3.7 (11.3)	–2.1 (6.4)	–2.4 (9.1)	–4.5 (8.7)	.33
Abnormal	1.8 (8.5)	–2.6 (6.3)	–1.7 (7.2)	–3.1 (11.0)	.66
Time					
Week 1	–6.1 (12.7)	–3.0 (6.5)	–2.5 (7.1)	–8.4 (9.8)[f,g]	<.001
Week 2	–2.1 (7.7)	–1.3 (5.2)	–0.1 (5.5)	–4.2 (9.0)	.09
Week 3	1.4 (13.5)	–3.2 (6.3)	2.2 (8.9)	–0.3 (10.3)	.04
Week 4	2.6 (6.2)	–1.7 (6.3)	–3.7 (10.7)	–3.4 (4.8)	.28

[a]CSBP: change in systolic blood pressure

[b]PELL: patient engagement started low and dropped even lower.

[c]PEHH: patient engagement started high and remained high.

[d]PEHL: patient engagement started high and dropped to low.

[e]PELH: patient engagement started low and rose to high.

[f]The given cluster is significantly different from the PEHL cluster.

[g]The given cluster is significantly different from the PEHH cluster.

The Effect of Follow-up on SBP in Each of the 4 Weeks

The results of ANOVA revealed significant main effects of cluster (F=2.697; P=.045) and time (F=7.561; P<.001), but no cluster–time interaction effect (F=1.600; P=.11) on SBP (see Table 6). Bonferroni multiple comparisons showed that the PELH cluster in the first week had a significantly higher level of SBP decline than did the PEHH (P=.001) and PEHL (P<.001) clusters. The SBP decreased within the first 2 weeks in all 4 clusters and continued to fall in the PEHH and PELH clusters over 4 weeks. Overall, the SBP was more reduced in the PELL (–6.1 mmHg) and PELH (–8.4 mmHg) clusters than in the others in the first week.

Discussion

Principal Findings

PE Patterns

This study explored patterns of PE with the hypertension self-management app and identified the effect of health care provider follow-up on PE and SBP in each PE pattern in the first 4 weeks after registration. For the first time, we found 4 dynamic trends of PE in a sample of 562 patients who used the mHealth app to record self-measured BP. Two clusters started with a high level of engagement, with one remaining at a high level throughout and the other dropping. Two clusters started with low engagement, with one increasing the level of engagement and the other dropping.

The majority of patients (387/562, 68.9%) were mainly between 40 and 60 years old. There was no difference in the distribution of patients under 60 years of age in each PE cluster; however, there were more patients aged over 60 years in the continuously high engagement cluster than in the other clusters. This may suggest that the patients over 60 years of age were more likely to engage in recording self-measured BP using the mHealth app for hypertension self-management. This may be attributed to the high level of awareness of the risk of hypertension because age is an important contribution factor to the development of hypertension [16,43]. Our finding is consistent with that of Goyal et al [15] in which older participants completed more planning challenges for chronic disease management than did younger participants. However, the finding of Kruse et al [10] was not consistent with our observations. They found that patients over the age of 65 years were less likely to use the mHealth service due to problems understanding the information, difficulty using technology, and inability to access the internet [10].

In this study, there were more male patients than female patients, which may be attributed to males being more prone to using mHealth services for hypertension management [44,45]. However, we did not find a significant difference in the proportion of males and females in each PE pattern. It appears that, although more males used the mHealth services than did females [44], they had the same level of PE with the mHealth service. Abd-alrazaq et al [30] found that gender did not affect intention to use the mHealth service, which was in line with our findings. However, Chung et al [46] found that females with heart failure were more adherent to the sodium-restricted diet than males, and Goyal et al [15] found that PE levels among female users with chronic disease were slightly higher than those among male users. Therefore, the gender differences in PE with mHealth apps does not appear to be conclusive.

The relationship between PE and SBP in the PELL cluster is worth noting. Patients in the PELL cluster reported a higher

SBP in the first week, and after 3 weeks of management, the SBP decreased more compared with the other clusters. Patients then began to stop using the app, and this decline in PE continued simultaneously with the increase in SBP. It may be the case that patients felt they reached the BP goal and ceased use of the app [27,47]. This may give evidence to suggest that controlling BP requires continuous engagement in hypertension self-management [35].

The Effect of 3 Types of Follow-ups on PE and SBP

Health care provider follow-up is essential for the prevention and treatment of hypertension. It is necessary to understand which type of follow-up is effective for which patient group [2]. Compliance follow-up, which was provided when the patients showed signs of reducing compliance in recording self-measured BP, had a positive effect on PE in the PEHH and PELH clusters. Regular follow-up, which was provided at a fixed interval, only had a positive effect on PE in the PELH cluster. These follow-ups might have been taken as the cue for engagement behavior for these patients, who may have awareness of hypertension self-management, but need cues for action. Interestingly, abnormal follow-up had no effect on any cluster of patients despite it being performed when the patient's BP was abnormal. These findings suggest that different types of follow-ups had different effects on patients' behavior of recording self-measured BP.

The content and method of follow-up are important to achieving a follow-up effect [48-50]. One study showed that health care providers have individual inherent preferences for the type of follow-up content to be provided to the patient [18]. Abd-alrazaq et al [30] found that the health care provider factor was related to patients' intension to engage in mHealth services. In our study, follow-up was performed through the mobile phone after health care providers received a reminder from the mHealth app. Other ways to improve PE, such as routine reminders combined with games, have yielded higher user engagement [16], and users seem to prefer simple, short voice messages over text messages because of communication trust and increased accessibility [51]. Cechetti et al [52] developed and implemented an mHealth app with a gamification method for hypertension management, which proved to be effective in promoting PE [52].

The mean SBP of all 4 clusters of patients reduced after 3 types of follow-ups except for abnormal follow-up in the PELL cluster. We also found abnormal follow-up was performed less frequently in the PELL cluster than in the others. This may be attributed to patients in the PELL cluster having a low level of engagement in recording self-measured BP, which resulted in the mHealth app not detecting the patient's abnormal BP condition and reminding the health care provider to follow-up. The PELH cluster developed positive PE and achieved SBP control after 3 types of follow-ups. This is in agreement with the observation that follow-up can motivate patients to engage in self-management. Positive feedback from PE after follow-up has been shown to be beneficial for BP control [53-55].

The Effect of Follow-up in Each of the 4 Weeks on PE and SBP

One study reported that 74.84% of app-only users stopped using an mHealth physical activity management app by day 43 [56]. This suggests that the duration of patient "stickiness" with the fully automatic mHealth services is limited and needs to be complemented by human support to keep momentum. Little is still known concerning the extent and duration of the effect of health care provider follow-up on self-management of hypertensive patients. This study found that the effect of follow-up on the PEHH and PELH clusters lasted until the third week, only had a small impact on the PELL cluster in the second week, and had no effect on the PEHL cluster at any time point. The effect duration of different types of follow-ups varied across each PE pattern. This may suggest that a fixed care pathway based only on patient's BP level would not work for all types of patients. Chronic disease management also needs to consider the patient's behavior and personal preferences. Our results support the design of patient-centered follow-up plans that incorporate social behavioral characteristics and preferences of patients into chronic disease management [14].

After 4 weeks of mHealth services being used, the mean SBP of all 4 clusters decreased. Interestingly, in the first week, the PELL and PELH clusters had a higher level of decline in SBP than did the other clusters. This may be the reason why the PELL cluster was complacent and did not actively engage in recording self-measured BP. Taking the advice of the health care providers may motivate those in the PELH cluster to increase their level of engagement. In the first 2 weeks, all 4 clusters of patients experienced a decline in SBP. Only the PEHH and PELH clusters maintained the SBP reduction through the fourth week, supporting the notion that hypertension management requires ongoing effort in monitoring BP to help patients improve their awareness of their own condition.

Strengths and Limitations

The strength of this study was that, first, the study used the longitudinal data collected in the first 4 weeks of patient registration with the mHealth service, which is informative for characterizing the changing trend of PE with hypertension management under an mHealth service. The findings are useful for continuous improvement of mHealth services for hypertension management. To the best of our knowledge, no other study has used longitudinal data to describe patterns of PE with a hypertension self-management app. Second, we analyzed the effect of health care provider follow-up on PE and SBP in each PE pattern. Our findings revealed the relationship between PE, BP, and health care provider follow-up. This provides evidence to support the further design of appropriate types of follow-ups for patients. Finally, we observed PE from real-world patients, which can reveal patient behavior in a natural setting, rather than recruited patients who would be more likely to overcome the burden associated with research work [57]. This helped to generate implementable, practical insights for the engagement of actual patients in daily hypertension management.

The study had 4 limitations. First, we excluded patients who used the app for fewer than 4 weeks after registration. This

might have produced a bias toward more positive findings of PE with the mHealth app and thus may limit the generalizability of our findings to those patients who drop out of the mHealth service early. Second, we only investigated the PE trend with the mHealth service in the first 4 weeks of hypertension management. Future study can further investigate the PE pattern after this period of time. Third, we defined PE as the behavior of recording self-measured BP. There are many other usage behaviors, such as taking medicine, engaging in physical activities, and keeping a healthy diet, which should be analyzed in future studies. Fourth, the different engagement behaviors could have arisen from various demographic and social psychological characteristics of patients (such as marital status, profession, anxiety, depression), so these factors need to be considered in future studies.

Conclusions

By analyzing the 4-week log data from a hypertension self-management app, BPA, this study identified the 4 distinct PE patterns in using an mHealth app for hypertension self-management. We also characterized the different effects of 3 types of health care provider follow-up on PE and SBP. Results showed how patients engaged with the mobile app and how health care provider follow-up affects or does not affect their engagement and BP. Our findings may inform the design and help strengthen health care provider follow-up strategies to improve outpatient engagement with mHealth apps for hypertension management. Future work needs to clarify the long-term engagement of patients with hypertension health care services. The indicators of PE should be broadened to include multiple types of usage behavior, and the effect of patient provider follow-up needs to be associated with patients' various demographic and socio-psychological characteristics.

Acknowledgments

This study was supported by the National Key Research and Development Program of China (#2020YFC2006400 and #2020YFC2003403), and the Major Science and Technology Project in Hainan Province of China (# ZDKJ2019012).

Conflicts of Interest

None declared.

References

1. Global status report on noncommunicable diseases 2014. World Health Organization. 2014. URL: https://apps.who.int/iris/bitstream/handle/10665/148114/9789241564854_eng.pdf;jsessionid=3C0C9A267DA218A8836BDCA9EB2C5C0D?sequence=1 [accessed 2021-09-05]
2. Reynolds R, Dennis S, Hasan I, Slewa J, Chen W, Tian D, et al. A systematic review of chronic disease management interventions in primary care. BMC Fam Pract 2018 Jan 09;19(1):11 [FREE Full text] [doi: 10.1186/s12875-017-0692-3] [Medline: 29316889]
3. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
4. Osborn CY, Mayberry LS, Wallston KA, Johnson KB, Elasy TA. Understanding patient portal use: implications for medication management. J Med Internet Res 2013;15(7):e133 [FREE Full text] [doi: 10.2196/jmir.2589] [Medline: 23823974]
5. Wong A, Bhyat R, Srivastava S, Boissé Lomax L, Appireddy R. Patient Care During the COVID-19 Pandemic: Use of Virtual Care. J Med Internet Res 2021 Jan 21;23(1):e20621 [FREE Full text] [doi: 10.2196/20621] [Medline: 33326410]
6. Chen X, Koskela M, Hyväkkä J. mHealth new horizons for health through mobile technologies. World Health Organization. 2011. URL: https://apps.who.int/iris/bitstream/handle/10665/44607/9789241564250_eng.pdf?sequence=1&isAllowed=y [accessed 2021-09-05]
7. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018;1:12 [FREE Full text] [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
8. Li R, Liang N, Bu F, Hesketh T. The effectiveness of self-management of hypertension in adults using mobile health: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e17776 [FREE Full text] [doi: 10.2196/17776] [Medline: 32217503]
9. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019 Apr 11;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
10. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
11. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]

12. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
13. Sakakibara BM, Ross E, Arthur G, Brown-Ganzert L, Petrin S, Sedlak T, et al. Using mobile-health to connect women with cardiovascular disease and improve self-management. Telemed J E Health 2017 Mar;23(3):233-239. [doi: 10.1089/tmj.2016.0133] [Medline: 27623231]
14. Rothman AA, Wagner EH. Chronic illness management: what is the role of primary care? Ann Intern Med 2003 Feb 04;138(3):256-261. [Medline: 12558376]
15. Goyal S, Morita PP, Picton P, Seto E, Zbib A, Cafazzo JA. JMIR Mhealth Uhealth 2016 Mar 24;4(1):e32 [FREE Full text] [doi: 10.2196/mhealth.4730] [Medline: 27012937]
16. Edney S, Ryan JC, Olds T, Monroe C, Fraysse F, Vandelanotte C, et al. User engagement and attrition in an app-based physical activity intervention: secondary analysis of a randomized controlled trial. J Med Internet Res 2019 Nov 27;21(11):e14645 [FREE Full text] [doi: 10.2196/14645] [Medline: 31774402]
17. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
18. Tyagi H, Sabharwal M, Dixit N, Pal A, Deo S. Leveraging providers' preferences to customize instructional content in information and communications technology-based training interventions: retrospective analysis of a mobile phone-based intervention in India. JMIR Mhealth Uhealth 2020 Mar 03;8(3):e15998 [FREE Full text] [doi: 10.2196/15998] [Medline: 32130191]
19. James J. Patient Engagement. Health Affairs Health Policy Brief 2013;14. [doi: 10.1377/hpb20130214.898775]
20. Toro-Ramos T, Kim Y, Wood M, Rajda J, Niejadlik K, Honcz J, et al. Efficacy of a mobile hypertension prevention delivery platform with human coaching. J Hum Hypertens 2017 Dec;31(12):795-800 [FREE Full text] [doi: 10.1038/jhh.2017.69] [Medline: 28972573]
21. Goyal S, Nunn CA, Rotondi M, Couperthwaite AB, Reiser S, Simone A, et al. A mobile app for the self-management of type 1 diabetes among adolescents: a randomized controlled trial. JMIR Mhealth Uhealth 2017 Jun 19;5(6):e82 [FREE Full text] [doi: 10.2196/mhealth.7336] [Medline: 28630037]
22. Rehman H, Kamal AK, Morris PB, Sayani S, Merchant AT, Virani SS. Mobile health (mHealth) technology for the management of hypertension and hyperlipidemia: slow start but loads of potential. Curr Atheroscler Rep 2017 Mar;19(3):12. [doi: 10.1007/s11883-017-0649-y] [Medline: 28210974]
23. Cheikh-Moussa K, Mira JJ, Orozco-Beltran D. Improving engagement among patients with chronic cardiometabolic conditions using mHealth: critical review of reviews. JMIR Mhealth Uhealth 2020 Apr 08;8(4):e15446 [FREE Full text] [doi: 10.2196/15446] [Medline: 32267239]
24. Kaplan AL, Cohen ER, Zimlichman E. Improving patient engagement in self-measured blood pressure monitoring using a mobile health technology. Health Inf Sci Syst 2017 Dec;5(1):4 [FREE Full text] [doi: 10.1007/s13755-017-0026-9] [Medline: 29081974]
25. Perski O, Blandford A, Garnett C, Crane D, West R, Michie S. A self-report measure of engagement with digital behavior change interventions (DBCIs): development and psychometric evaluation of the "DBCI Engagement Scale". Transl Behav Med 2020 Feb 03;10(1):267-277. [doi: 10.1093/tbm/ibz039] [Medline: 30927357]
26. Rahman QA, Janmohamed T, Pirbaglou M, Ritvo P, Heffernan JM, Clarke H, et al. Patterns of user engagement with the mobile app, manage my pain: results of a data mining investigation. JMIR Mhealth Uhealth 2017 Jul 12;5(7):e96 [FREE Full text] [doi: 10.2196/mhealth.7871] [Medline: 28701291]
27. Sanatkar S, Baldwin PA, Huckvale K, Clarke J, Christensen H, Harvey S, et al. Using cluster analysis to explore engagement and e-attainment as emergent behavior in electronic mental health. J Med Internet Res 2019 Nov 28;21(11):e14728 [FREE Full text] [doi: 10.2196/14728] [Medline: 31778115]
28. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
29. Mosby. Follow-up. (n.d.). Medical Dictionary for the Health Professions and Nursing Internet. URL: https://medical-dictionary.thefreedictionary.com/follow-up [accessed 2021-09-05]
30. Abd-Alrazaq AA, Bewick BM, Farragher T, Gardner P. Factors that affect the use of electronic personal health records among patients: A systematic review. Int J Med Inform 2019 Jun;126:164-175. [doi: 10.1016/j.ijmedinf.2019.03.014] [Medline: 31029258]
31. Kohl LFM, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
32. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: Recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
33. Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Dennison HC, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation,

and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension 2017 Nov 13:e13-e115. [doi: 10.1161/HYP.0000000000000065] [Medline: 29133356]
34. National Center for Cardiovascular Diseases of China. China Hypertension Community Management Guidelines 2014. Chinese Journal of Health Management 2015 Sep 1;9(1):10-30. [doi: 10.3760/cma.j.issn.1674-0815.2015.01.004]
35. Joint Committee for Guideline Revision. 2018 Chinese Guidelines for Prevention and Treatment of Hypertension-A report of the Revision Committee of Chinese Guidelines for Prevention and Treatment of Hypertension. J Geriatr Cardiol 2019 Mar;16(3):182-241 [FREE Full text] [doi: 10.11909/j.issn.1671-5411.2019.03.014] [Medline: 31080465]
36. Duan H, Wang Z, Ji Y, Ma L, Liu F, Chi M, et al. Using goal-directed design to create a mobile health app to improve patient compliance with hypertension self-management: development and deployment. JMIR Mhealth Uhealth 2020 Feb 25;8(2):e14466 [FREE Full text] [doi: 10.2196/14466] [Medline: 32130161]
37. Wang Z, Li C, Huang W, Chen Y, Li Y, Huang L, et al. Effectiveness of a pathway-driven eHealth-based integrated care model (PEICM) for community-based hypertension management in China: study protocol for a randomized controlled trial. Trials 2021 Jan 22;22(1):81 [FREE Full text] [doi: 10.1186/s13063-021-05020-2] [Medline: 33482896]
38. Serrano KJ, Coa KI, Yu M, Wolff-Hughes DL, Atienza AA. Characterizing user engagement with health app data: a data mining approach. Transl Behav Med 2017 Jun;7(2):277-285 [FREE Full text] [doi: 10.1007/s13142-017-0508-y] [Medline: 28616846]
39. Dou K, Yu P, Deng N, Liu F, Guan Y, Li Z, et al. Patients' acceptance of smartphone health technology for chronic disease management: a theoretical model and empirical test. JMIR Mhealth Uhealth 2017 Dec 06;5(12):e177 [FREE Full text] [doi: 10.2196/mhealth.7886] [Medline: 29212629]
40. Jain AK. Data clustering: 50 years beyond K-means. Pattern Recognition Letters 2010 Jun;31(8):651-666. [doi: 10.1016/j.patrec.2009.09.011]
41. Huy VT, Anh DT. An efficient implementation of anytime K-medoids clustering for time series under dynamic time warping. New York, NY, USA: Association for Computing Machinery; 2016 Presented at: Proceedings of the Seventh Symposium on Information and Communication Technology (SoICT '16); 2016; Ho Chi Minh, Vietnam p. 22-29. [doi: 10.1145/3011077.3011128]
42. Liu Y, Li Z, Xiong H, Gao X, Wu J. Understanding of internal clustering validation measures. 2010 Presented at: 2010 International Conference on Data Mining; Dec 13, 2010; Sydney, NSW, Australia p. 911-916.
43. Rao J, Chen X, Zhang J, Yang J, Pang M, Pan B, et al. The risk of hypertension doubles every 10 years in China: Age, period, and birth cohort effects on the prevalence of hypertension from 2004 to 2013. Am J Hypertens 2019 Apr 22;32(5):492-502. [doi: 10.1093/ajh/hpz003] [Medline: 30668634]
44. Zhang X, Guo X, Lai K, Guo F, Li C. Understanding gender differences in m-health adoption: a modified theory of reasoned action model. Telemed J E Health 2014 Jan;20(1):39-46. [doi: 10.1089/tmj.2013.0092] [Medline: 24161004]
45. Khatun F, Heywood AE, Hanifi SMA, Rahman MS, Ray PK, Liaw S, et al. Gender differentials in readiness and use of mHealth services in a rural area of Bangladesh. BMC Health Serv Res 2017 Aug 18;17(1):573 [FREE Full text] [doi: 10.1186/s12913-017-2523-6] [Medline: 28821243]
46. Chung ML, Moser DK, Lennie TA, Worrall-Carter L, Bentley B, Trupp R, et al. Gender differences in adherence to the sodium-restricted diet in patients with heart failure. J Card Fail 2006 Oct;12(8):628-634 [FREE Full text] [doi: 10.1016/j.cardfail.2006.07.007] [Medline: 17045182]
47. Christensen H, Mackinnon A. The law of attrition revisited. J Med Internet Res 2006 Sep 29;8(3):1-4 [FREE Full text] [doi: 10.2196/jmir.8.3.e20] [Medline: 17032636]
48. Billue KL, Safford MM, Salanitro AH, Houston TK, Curry W, Kim Y, et al. Medication intensification in diabetes in rural primary care: A cluster-randomised effectiveness trial. BMJ Open 2012 Sep 17;2(5):e000959. [doi: 10.1136/bmjopen-2012-000959]
49. Estrada CA, Safford MM, Salanitro AH, Houston TK, Curry W, Williams JH, et al. A web-based diabetes intervention for physician: a cluster-randomized effectiveness trial. Int J Qual Health Care 2011 Dec;23(6):682-689 [FREE Full text] [doi: 10.1093/intqhc/mzr053] [Medline: 21831967]
50. Crenshaw K, Curry W, Salanitro AH, Safford MM, Houston TK, Allison JJ, et al. Is physician engagement with web-based CME associated with patients' baseline hemoglobin A1c levels? The rural diabetes online care stud. Academic Medicine 2010;85(9):1511-1517. [doi: 10.1097/acm.0b013e3181eac036]
51. Steinman L, Heang H, van Pelt M, Ide N, Cui H, Rao M, et al. Facilitators and barriers to chronic disease self-management and mobile health interventions for people living with diabetes and hypertension in Cambodia: qualitative study. JMIR Mhealth Uhealth 2020 Apr 24;8(4):e13536 [FREE Full text] [doi: 10.2196/13536] [Medline: 32329737]
52. Cechetti NP, Bellei EA, Biduski D, Rodriguez JPM, Roman MK, De Marchi ACB. Developing and implementing a gamification method to improve user engagement: A case study with an m-Health application for hypertension monitoring. Telematics and Informatics 2019 Aug;41:126-138. [doi: 10.1016/j.tele.2019.04.007]
53. Miller AS, Cafazzo JA, Seto E. A game plan: Gamification design principles in mHealth applications for chronic disease management. Health Informatics J 2014 Jul 1;22(2):184-193. [doi: 10.1177/1460458214537511] [Medline: 24986104]

54. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
55. Giles EL, Robalino S, McColl E, Sniehotta FF, Adams J. The effectiveness of financial incentives for health behaviour change: systematic review and meta-analysis. PLoS One 2014;9(3):e90347 [FREE Full text] [doi: 10.1371/journal.pone.0090347] [Medline: 24618584]
56. Guertler D, Vandelanotte C, Kirwan M, Duncan MJ. Engagement and nonusage attrition with a free physical activity promotion program: the case of 10,000 steps Australia. J Med Internet Res 2015;17(7):e176 [FREE Full text] [doi: 10.2196/jmir.4339] [Medline: 26180040]
57. Sanders I, Short CE, Bogomolova S, Stanford T, Plotnikoff R, Vandelanotte C, et al. Characteristics of adopters of an online social networking physical activity mobile phone app: cluster analysis. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e12484 [FREE Full text] [doi: 10.2196/12484] [Medline: 31162130]

Abbreviations

ANOVA: analysis of variance
BP: blood pressure
BPA: Blood Pressure Assistant
CPE: change in patient engagement
CSBP: the change in systolic blood pressure
mHealth: mobile health
PE: patient engagement
PELL: patient engagement started low and dropped even lower
PEHH: patient engagement started high and remained high
PEHL: patient engagement started high and dropped to low
PELH: patient engagement started low and rose to high
SBP: systolic blood pressure

Edited by R Kukafka; submitted 09.11.20; peer-reviewed by L Liu, J Lacroix; comments to author 05.02.21; revised version received 10.03.21; accepted 10.08.21; published 28.09.21.

Please cite as:
Wu D, An J, Yu P, Lin H, Ma L, Duan H, Deng N
Patterns for Patient Engagement with the Hypertension Management and Effects of Electronic Health Care Provider Follow-up on These Patterns: Cluster Analysis
J Med Internet Res 2021;23(9):e25630
URL: https://www.jmir.org/2021/9/e25630
doi:10.2196/25630
PMID:34581680

Original Paper

Smartphone Technology to Empower People Experiencing Homelessness: Secondary Analysis

Whitney Thurman[1], PhD, RN; Monika Semwal[2], MPharm, PhD; Leticia R Moczygemba[2], PharmD, PhD; Mark Hilbelink[3], MDiv

[1]School of Nursing, University of Texas at Austin, Austin, TX, United States

[2]Health Outcomes Division, College of Pharmacy, University of Texas at Austin, Austin, TX, United States

[3]Sunrise Homeless Navigation Center, Austin, TX, United States

Corresponding Author:
Whitney Thurman, PhD, RN
School of Nursing
University of Texas at Austin
1710 Red River St
Austin, TX, 78701
United States
Phone: 1 5126573855
Email: wthurman@utexas.edu

Abstract

Background: In the United States, the number of people experiencing homelessness has continually increased over the last 3 years. Homelessness is associated with poor health, and people experiencing homelessness are often burdened with high rates of chronic and mental health conditions, functional limitations, and cognitive impairment. Despite the high burden of chronic illness and functional limitations, there is limited literature exploring self-management among homeless populations.

Objective: This study aims to investigate how access to smartphone technology facilitates self-management, including the attainment of social needs within the context of homelessness.

Methods: A secondary analysis of 33 exit interviews from 2 feasibility studies related to mobile health interventions among people experiencing homelessness was conducted. Iterative thematic analysis was used to identify themes representative of participants' experiences using smartphone technology.

Results: Collectively, participants revealed not only how the context of homelessness constrained their ability to engage in activities necessary to self-manage health and meet social needs but also how consistent and predictable access to the tools available through a smartphone changed their behaviors and outlook. The global theme of *empowered by technology* was identified and defined as how having a smartphone with a plan for unlimited text, calling, data, and transportation allowed participants to navigate homelessness and facilitated self-management.

Conclusions: People experiencing homelessness used the tools on a smartphone to make decisions, take action, solve problems, and use the resources—skills necessary for fulfilling tasks required for effective self-management. Further, consistent access to smartphone technology and transportation empowered participants to meet the requirements for the attainment of social needs.

(J Med Internet Res 2021;23(9):e27787) doi:10.2196/27787

KEYWORDS

homelessness; self-management; smartphone technology; social needs; mobile phone

Introduction

Background

In the United States, the number of people experiencing homelessness has continually increased over the last 3 years, with a 3% increase from 552,830 in 2018 to 567,715 in 2019 [1]. This number is likely to increase during the ongoing COVID-19 pandemic [2]. Homelessness is associated with poor health, and people experiencing homelessness are often burdened with high rates of chronic and mental health conditions, functional limitations, and cognitive impairment [3-5]. Furthermore, the rates of disability are high among people experiencing homelessness. In 2016, almost 50% of people

experiencing homelessness had some type of disability compared with 20% of the general US population [6].

The reasons for poor health are multifactorial and include risk factors that increase the risk of homelessness in the first place, such as early life poverty, mental illness, and substance use disorders. However, many risk factors are secondary to the conditions of homelessness, such as poor nutrition, exposure to communicable diseases, harsh living conditions, and violence [7,8]. Many people experiencing homelessness face difficulties in connecting to and accessing appropriate health and social services because of social and psychological barriers such as stigma, lack of trust in health care providers, and transient lifestyles [9-11]. Furthermore, people experiencing homelessness struggle to meet the basic needs such as food, shelter, and transportation, resulting in self-management of chronic health conditions being overlooked or neglected [12-14]. These barriers and competing priorities contribute to adverse outcomes, including reliance on emergency departments (EDs) or hospitals for nonemergent health care services and increased risk of 30-day hospital readmissions [15-17].

Self-management has been defined as "the ability of the individual, in conjunction with family, community, and health care professionals, to manage symptoms; treatments; lifestyle changes; and psychosocial, cultural, and spiritual consequences of health conditions" [18] and requires patients to perform a variety of day-to-day tasks to manage symptoms, treatments, and consequences of chronic conditions [19]. A component of self-management also includes developing the skills needed to devise, implement, evaluate, and revise plans to take charge of one's own health [20]. Given that self-management requires planning ahead, dedicated resources, and access to information, it is not surprising that homelessness poses unique barriers to successful self-management. For example, people experiencing homelessness have described poor sleeping conditions, limited food choices, limited income, and a lack of access to storage as barriers to self-management [21].

Despite the high burden of chronic illness and functional limitations, there is limited literature exploring self-management among homeless populations. To the extent that self-management has been investigated among people experiencing homelessness, outcomes are encouraging [22-24]. In addition, Merdsoy et al [21] explored the perceptions, needs, and preferences for self-management among men experiencing homelessness. They found that the majority of participants perceived self-management as important, but complex challenges related to homelessness, including lack of social support and limited access to communication technology, dictated unique self-management support needs [21].

Given the distinct context of homelessness and the importance of social support and access to information and communication for effective self-management, smartphone technology may be a useful tool to facilitate self-management activities among people experiencing homelessness. In the United States, smartphones are ubiquitous. For people experiencing homelessness who often encounter many barriers to communication and information seeking, smartphones allow for multiple resources in one tool, helping overcome access barriers. Prior evidence suggests that the connectivity provided by smartphones can decrease social isolation and facilitate communication between peers, family, and service providers [25]. In addition, pilot studies have investigated using phone calls for improving medication adherence [26], apps for improving mental health [27], and text messaging for appointment reminders [28] and improving physical activity [29] among people experiencing homelessness. This burgeoning body of evidence suggests that people experiencing homelessness are welcoming and supportive of using smartphone technology to address their health care issues, and smartphone technology has also been identified as a feasible tool for self-management support among people experiencing homelessness [25].

Objective

The purpose of this study is to investigate how access to smartphone technology facilitates self-management, including the attainment of social needs within the context of homelessness. The methods and results are reported in accordance with the Consolidated Criteria for Reporting Qualitative studies checklist [30].

Methods

Overview

This study is a secondary analysis of exit interviews from 2 pilot studies investigating the preliminary outcomes of two mobile health interventions among people experiencing homelessness [25]. Both studies were conducted in Austin, Texas, United States. The first pilot study was a 4-month study designed to improve care coordination and reduce acute health care utilization among people experiencing homelessness [25]. The second was a 1-month study designed to improve medication adherence among people experiencing homelessness. In each study, participants were recruited from 3 churches that provided services to people experiencing homelessness. Convenience sampling was used, and potential participants were referred by community partner site staff or volunteers to the research team for study eligibility and screening (for eligibility criteria for each study, refer to Textbox 1). At baseline, the first author (WT) administered a series of baseline assessments to collect demographic data and self-reported health history, medication adherence, social support, and recent hospitalizations or ED visits. Upon enrollment into the study, participants received a study-assigned smartphone activated with a plan for unlimited text, calling, data, and public transportation.

Exit interviews were conducted between February and June 2019 with 16 people experiencing homelessness who participated in the first study and between August and September 2019 with 17 people experiencing homelessness who participated in the second study. The semistructured interviews included questions about participating in the study, having and using a smartphone, ED or hospital visits during the study period, and medication use. The interviews allowed participants to share their experiences with smartphone technology and how access to a smartphone with a data plan and access to unlimited public transportation affected their daily lives and routines. For this secondary analysis, investigators focused on the component

of the interview in which participants described how they used their smartphones to accomplish various daily tasks. Interviews were conducted by the first author (WT) in quiet, private locations in church-based entities from which the participants had been recruited. Interviews were digitally recorded, and a team of graduate students transcribed and verified each interview. Both studies were approved by the university institutional review board, and informed consent was obtained before participant enrollment.

Textbox 1. Inclusion criteria for research studies.

Study 1

- Aged ≥18 years
- Currently experiencing homelessness (defined as where the person had slept most nights in the previous 30 days)
- Currently prescribed at least two medications for chronic health conditions (self-report)
- Diagnosis of at least one chronic health condition (self-report)
- At least two hospitalizations or emergency department visits in the past 6 months (self-report)
- Score of at least 10 on the Patient Health Questionnaire-9

Study 2

- Aged ≥18 years
- Currently experiencing homelessness (defined as where the person had slept most nights in the previous 30 days)
- Currently prescribed at least two medications for chronic health conditions (self-report)
- Diagnosis of at least one chronic health condition (self-report)
- At least two hospitalizations or emergency department visits in the past 6 months (self-report)
- Score of at least 4 on the Rapid Estimate of Adult Literacy in Medicine-Short Form

Data Analysis

Overview

Iterative thematic analysis situated within a critical realism framework [31] was used to explore how participants incorporated smartphone technology into their lives and how the broader social context of homelessness influenced their experiences with technology and their interpretations of those experiences [32]. Three authors (WT, MS, and LRM) of this paper independently read 3 transcripts to familiarize themselves with the data before independently conducting line-by-line coding and then meeting to discuss the initial themes. After 2 subsequent rounds of independent coding and discussion, the team reached consensus on the themes and finalized the codebook. Once consensus regarding the codebook was reached, the remaining transcripts were divided between the 3-member coding team to independently code, and iterative analyses continued until saturation of major themes was achieved. Dedoose software (SocioCultural Research Consultants) was used to facilitate coding and data management.

Qualitative Rigor

Transferability was addressed by providing detailed descriptions of data and context of homelessness and reporting data using participants' direct quotes. Credibility of data analysis was enhanced by having each of the first 3 authors (WT, MS, and LRM) critically analyze the data before reflecting on the analytic process. This process enabled the authors to discuss negative cases, which facilitated a thick description of the data [33]. Reflexivity was maintained through ongoing documentation in a research log, comparing these notes against the data and engaging in whole-group discussions about the reflections.

Sample

A total of 31 unique individuals are represented in the 33 interviews. On average, participants were 42.7 (SD 9.67) years and had been experiencing homelessness for a mean of 7.4 years (range 1 month to 30 years). Of the 31 participants, 25 (81%) were unsheltered, sleeping on the street or in tents, whereas 4 (13%) reported staying in a shelter, 1 (3%) had just been released from jail, and 1 (3%) reported staying with a friend. Participants reported a substantial burden of chronic illness as 77% (24/31) experienced multiple chronic conditions (≥2) and 55% (17/31) reported being prescribed at least four different medications. Participant demographics and health-related characteristics are given in Table 1.

Table 1. Participant demographics (N=31).

Demographic variable	Values
Age (years), mean (SD)	42.7 (1.73)
Length of homelessness (years), mean (SD)	8.1 (7.7)
Gender, n (%)	
Male	21 (68)
Female	7 (23)
Transgender female	1 (3)
Transgender male	1 (3)
Other	1 (3)
Race and ethnicity, n (%)	
White and non-Hispanic	21 (68)
Black and non-Hispanic	5 (16)
Hispanic	2 (6)
Asian	1 (3)
Other	2 (6)
Functional limitations, n (%)	
0	11 (36)
1	9 (29)
2	7 (23)
3+	4 (13)
Chronic conditions, n (%)	
Hypertension	20 (64)
High cholesterol	10 (32)
Asthma	9 (29)
COPD[a]	9 (29)
Diabetes mellitus	8 (26)
Anxiety or depression	7 (23)
Bipolar disorder	5 (16)
Seizure disorder	5 (16)
Brain injury	4 (13)
Myocardial infarction or stroke	2 (6)
Multiple chronic conditions (≥2)	24 (77)
Prescribed medications, n (%)	
2-3	14 (45)
4-5	7 (23)
6+	10 (32)

[a]COPD: chronic obstructive pulmonary disease.

Just over half of the participants (16/31, 52%) reported having a personal cell phone with service upon enrollment in the study. Of those with phones at baseline, the majority (14/16, 87%) reported that their phones were capable of text messaging, picture messaging, and mobile apps use. Of the 48% (15/31) of the participants who did not have a personal cell phone at baseline, 93% (14/15) reported that they did not have a cell phone because it was expensive, 13% (2/15) reported having no use of a phone, 7% (1/15) reported not knowing how to use a phone, and 27% (4/15) reported concerns about privacy.

Results

Overview

The qualitative analysis identified two major themes and eight subthemes that captured the ways in which smartphone technology facilitated self-management within the social context of homelessness. Collectively, participants revealed not only how the *context of homelessness* constrained their ability to engage in activities necessary to self-manage health and meet social needs but also how consistent and predictable access to the tools available through the smartphone changed their behaviors, thoughts, and outlook toward daily life. Thus, the global theme of *empowered by technology* was identified and defined as how having the smartphone activated with a plan for unlimited text, calling, data, and public transportation allowed participants to navigate the context of homelessness and facilitated self-management. Textbox 2 includes representative quotes from each of the themes and subthemes.

Textbox 2. Themes and representative quotes.

Context of homelessness

- Being stigmatized
  - "The first time I had a stroke, it wasn't diagnosed. I was put in a cab and sent to my room."
  - "...a lot of businesses don't allow people—homeless people—to come in and charge their phones. They cut the outlets off."
- Encountering barriers
  - "I mean it's hard to try and get bus passes, when-especially when you're broke. I mean, you either have to make deals with people who trade them for cigarettes or dollar here, dollar there. Or you got to get up and go to churches like super early in the morning where they hand them out."
  - "I been sitting three weeks for this guy just to bring me two letters. Two letters so I could return it to the people from [apartment complex] trying to get me an apartment. I am still sitting there, so I guess I get up early tomorrow and go knock on his office door."
- Precarious environments
  - "Why would I want someone to find me? Well, anything could happen. Get sick. I could get you know beat up, passed out somewhere. No one knows where I am, then the phone helps find me."
  - "Cause being on the streets and homeless, you know, every day is a matter of survival."
- Symptom experience
  - "Like over the last two days, my breathing has been so bad I've nearly gone to the emergency room. Over the last 3 days actually. Every day I'd be thinking about it. I need to go, I need to go. My breathing's really bad. I didn't, but it's just hit or miss for me whether I do or not."
  - "I was hit by a van and run over. I already had a hernia pretty bad...and when I got hit by the van, it tore the skin open and parts of my small intestines was coming out. I ignored it for about two months and the infection got so bad, the pain was so intense that I was blacking out in pain so I had to head to the ER."
- Resilience
  - "...no we're not that dumb...we're really smart because we have to go through all this stuff to still be living."

Empowered by technology

- Connecting with others
  - "I had a phone number people could reach me at, so that was definitely a good, good thing."
  - "the new phone is so amazing. It will take you around the world."
- Regaining control
  - "It's like a life assistant."
  - "This phone actually helps me out though cause I can contact people that I need to talk to on a daily basis and get the information done and get situations taken care of."
- Getting things done
  - "phones are so important to have nowadays. You really have to have one to be able to keep track of everything."

Context of Homelessness

Overview

The majority of participants had multiple morbidities and were actively seeking to manage their health. However, the context of homelessness often precluded participants from successfully doing so. Participants described factors directly related to lack of housing and safe shelter that included being stigmatized, encountering barriers such as antiloitering policies and complex bureaucracies to access health care and social services, living in precarious environments, and experiencing symptoms such as disabling pain that contributed to their inability to successfully manage their health. Participants also revealed resilience that allowed them to reconcile their current circumstances with hope for a better future.

Being Stigmatized

Participants uniformly reported being treated differently—usually poorly—by members of the community during daily interactions and by health care professionals during clinical encounters. One participant described an ongoing battle with a gastric ulcer stemming from his physician refusing to prescribe antibiotics out of the fear that the participant's homelessness would prevent him from finishing the entire regimen. Another participant shared his experience of asking a front desk clerk at a downtown hotel to call an ambulance for him during an emergency health situation:

> *They wouldn't call 911. They would not call me an ambulance...instead, they called security on me because I was homeless.*

Encountering Barriers

In addition to the differential treatment they received, the stigma experienced by participants also created barriers that prevented them from accessing needed goods and services and forced their attention away from higher-order tasks needed for self-management. For example, participants had to carefully plan their days to complete relatively mundane tasks such as charging their cellphones. Participants described difficulty in keeping their phones charged because of many businesses instituting antiloitering policies that require individuals to purchase a product to use electrical outlets. Thus, participants learned which businesses would allow them to access electrical outlets and strategized to ensure that they could safely spend time, often spending hours, each day doing so.

Another frequently encountered barrier was the inability to maintain contact with important others, including family, friends, and service providers. Participants shared how the lack of a mailing address created barriers to, for example, receiving important communications regarding disability benefits or receiving a government-issued cell phone. They also cited the lack of a phone number where they could be reached as a substantial impediment to obtaining employment and housing. One participant succinctly stated this barrier when he said as follows:

> *You know, my only other option would have been to give them the phone number for the [shelter], which is a message phone to a homeless shelter, you know. Not fun convincing an employer to hire a homeless guy.*

Precarious Environments

Instability is a hallmark characteristic of homelessness, and participants clearly articulated the precarious environments in which they found themselves. Participants lived in a state of constant worry about their personal safety and the safety of their possessions. One participant said:

> *people that may be so desperate to make a few bucks off of something they see that phone and they maybe follow you or do things to try obtain that phone including, but not limited to-to you know threatening your own health.*

Given the precarity of their situations, participants struggled to assert or maintain control of their daily lives, possessions, and health. Participants often had to rely on others to meet their needs and described the goodwill of bus drivers who sometimes let them ride for free and the willingness of some service providers to bend rules on their behalf. Participants also described being taken advantage of, especially because they lacked options.

Symptom Experience

The context of homelessness also defined participants' symptom experience, influencing not only the perception and evaluation of symptoms but also the management of such experiences. Participants often attributed their current homelessness to symptoms stemming from their chronic health conditions. One participant clearly explained this as follows:

> *I am homeless because I had...four brain surgeries. I mean, I was fine. I had a home and everything...Now I lost everything.*

Another participant described his reliance on medication after experiencing a brain injury when he was younger:

> *how important it is for me...to be on medicine that allows me to function or do stuff cause otherwise I can't keep a job, I can't keep a house, I can't you know.*

Pain was a nearly universal symptom. One participant described how he had experienced so much pain for so long that he considered his pain to be a part of him. Participants also described the creative ways in which they managed their symptoms while experiencing homelessness. One participant who suffered from chronic pain stemming from injuries incurred during his time spent fighting in Afghanistan described how carrying a large backpack full of his personal belongings doubled as therapeutic as it helped to compress his back and alleviate some of his lower back pain.

Anxiety and depression were also widespread, and participants described various ways in which they managed their symptoms. When asked how she prevented herself from succumbing to misery, one participant simply said *drugs*. She went on to recount how she would prefer a prescription medication that allows her to function, but in lieu of that, she uses small amounts of illicit substances to help her overcome her anxiety, focus on tasks she needs to do, and engage with other people. Importantly,

participants recognized that prescribed medications can interact with alcohol and other substances and took steps to manage potential adverse outcomes.

Resilience

Participants also revealed a certain resilience that allowed them to persevere despite the stigmatizing attitudes, barriers, precarious environments, and unrelenting symptoms of chronic illness and homelessness. This means participants were able to cope with the context of homelessness by relying on their personal ingenuity and determination and on their social connections, however fragile they may have been. One participant described installing solar panels at his camp, providing him with electricity and the benefits associated with it.

Although other participants had more straightforward strategies for navigating their circumstances, participants' collective fortitude was revealed succinctly by one participant who had moved into housing during the study time frame when she said, "when I was living on the streets, I always found a way."

Empowered by Technology

Overview

Despite the circumstances of homelessness, participants persevered in efforts to meet their health and social needs and experienced varying degrees of success. However, when provided with a smartphone activated with a plan for unlimited text, calling, and data as well as access to public transportation, participants demonstrated how these basic tools equipped them to more easily navigate the context of homelessness. During exit interviews, participants revealed how they went above and beyond the expectations of the research studies and incorporated the technology into their daily lives. This consistent access to the tools provided by the phone as well as the information and resources available through the internet caused a shift in participants' behaviors, thoughts, and perceptions, resulting in participants being empowered by the technology and, subsequently, better able to engage in self-management activities. Specifically, the subthemes that support this global theme are *connecting with others*, *getting things done*, and *regaining control*.

Connecting With Others

A dedicated phone number and consistent phone service facilitated participants' ability to stay connected with family and friends and, in some instances, served as a catalyst for reconnection after long periods of no contact. One participant shared how he was able to talk with his daughter and hear the cry of his first grandchild over the phone. A different participant tearfully recounted saying goodbye to his mother who fell ill and died suddenly during the course of the research study. He said:

> *I would have never gotten to say goodbye to my mom if I didn't have that phone.*

For some participants, the mobile bus pass and resultant access to transportation was as important for connecting with others as was the ability to make phone calls. Participants described traveling to other parts of the city to talk with friends and acquaintances that they had not seen or communicated with in a while. Conversely, some participants availed themselves of their transportation options to leave social situations that they judged to be unsafe or unhealthy. The phone and access to transportation also facilitated participants' ability to reciprocate social support. For example, one participant took the opportunity to teach his peers who were also participants how to navigate the bus system and locate various food pantries in town.

Importantly, the tools available via smartphones also allowed participants to maintain connections with service providers, including case managers, physicians, and others involved in helping them navigate the complex world of homeless services. One participant who moved into housing during the course of the study counted 4 case managers with whom he worked. He said as follows:

> *I had so many appointments and doctor visits and this and that and the other. So the phone was integral. I would not be in that place [housing] I can guarantee right now if I did not have that phone.*

Transportation access also allowed participants to more easily travel to appointments, facilitating both their ability to adhere to health advice and some relief to their health symptoms. Indeed, being able to navigate the urban landscape via bus instead of foot resulted in less pain and easier breathing for many participants who struggled to manage chronic respiratory conditions.

Getting Things Done

Getting things done was defined as using the phone for tasks or activities that went above and beyond the requirements or expectations of the research studies. Given access to the tools available through smartphones, participants were able to accomplish relatively mundane tasks that were nearly impossible to do without these tools. For example, participants described being able to call providers and schedule appointments. One participant succinctly stated:

> *I am able to talk to people that I need to talk to like social security, the food stamp place, and I can get doctors' appointments squared away with it and when I wasn't doing that before...it helps me to be more alert and it helps me to take care of my business...when I need to because everything is on the phone.*

Having a phone meant that not only did they have access to information and to others, but they also always knew what time it was, which allowed them to establish routines that facilitated self-management. Participants described using prompts and time schedules to get "day-to-day stuff done" and how "a bus pass ready to go...any type of information that is needed on the information highway...[and] entertainment" available at their fingertips kept them moving toward their goals.

This ability to get things done by using the tools available on the smartphone facilitated the attainment of several social needs among participants enrolled in the 4-month study. Specifically, 6 participants moved into housing, 5 gained employment, and 1 obtained disability benefits during their time in the study, and participants credited their access to both the study phone and

reliable transportation with their ability to maintain contact and attend appointments that were required to attain these social needs.

Regaining Control

Ultimately, access to the tools available on the smartphone, including transportation and internet access, allowed participants to regain control over aspects of their environments and their everyday lives. For example, one participant said as follows:

> *It's too hot so I stayed on. I mean, I just always had a place to go with that bus pass. I didn't have to beg the bus driver to let me on.*

Another participant described the phone thusly, "...it gave me a little bit more extra freedom and kept me reminded of things I needed to do and stuff and things I wanted to do."

Others described shifting perceptions of themselves and their circumstances that empowered them to reassert control. One participant described how, by virtue of the smartphone and consistent access to transportation, he was able to think beyond his own survival needs and how this shifted his perceptions of his circumstances. He described the experience of having a smartphone and participating in the research study as helping him get back on track and said:

> *makes you feel a little bit more important, like I mean especially out on the street it makes you feel like you're doing something. Makes you feel a little bit better than what you are out here.*

He credited his decision to re-establish a connection with his case manager and enroll in a temporary housing program to this shift in his self-perception.

Participants also attributed the tools made available via the smartphone with decreasing their worry and improving their general sense of well-being:

> *It was a godsend. It really was. I mean because I didn't have to worry about a lot of things. I could make phone calls when I needed to...It just took a lot of burden off me knowing that I had a bus pass. I had a phone I could use, you know, if I got in trouble or something or was in a bad situation and so it was very helpful for that.*

Discussion

Principal Findings

Findings from this study extend our understanding of how mobile technology—specifically, the predictable and consistent access to the tools available on a smartphone—can facilitate self-management, including attainment of social needs within homeless populations. Smartphones allowed participants to have multiple resources, including internet access, mobile apps, email, maps, and entertainment, on one device that overcomes barriers to storing and keeping track of multiple items. This robust functionality, as compared with older phones without these features, is important to facilitate the connection between people experiencing homelessness and their communities.

Effective self-management requires individuals to manage the symptoms and consequences of chronic illness by regular monitoring and management of physical, cognitive, behavioral, and emotional changes [19]. Lorig and Holman [34] suggested that these activities can be classified as medical, emotional, and role management tasks, and our findings indicate that access to unlimited data, text, calling, and transportation can support people experiencing homelessness to successfully engage in these activities. This means that participants revealed how they were able to use the tools on the smartphone to make decisions, take action, solve problems, and use resources—skills necessary for fulfilling the tasks required for self-management and meeting social needs [34].

A hallmark characteristic of self-management is a collaborative approach in which patients partner with professionals to devise strategies related to the ongoing management of symptoms and illness [35] This type of patient-provider partnership enables decision-making and problem solving, but these activities are predicated on having access to appropriate information. Thus, without consistent and reliable access to communication technologies, it is nearly impossible for people experiencing homelessness to engage in this collaborative approach because of an inability to maintain contact with providers or access the required information and resources. However, our findings support earlier evidence suggesting that people experiencing homelessness perceive self-management as important as participants in this study sought to be active participants in their health care and proactively sought information about their health conditions and engaged with providers when possible [21]. Thus, when equipped with smartphone technology, participants were empowered to actively engage in self-management tasks and activities.

Unfortunately, without a phone, many people experiencing homelessness simply miss opportunities for housing and employment because they cannot be located in a timely manner and do not have reliable contact information for an employer. At least six participants stayed engaged with case managers via the smartphone to complete the requirements to get into housing during the course of the study, and five participants reported gaining employment. This indicates that a smartphone is a powerful and necessary tool for people experiencing homelessness to meet social needs.

It should be noted that the context of homelessness presents challenges for everyday use of technology. Similar to prior research, participants in this study indicated that maintaining charged devices and keeping devices protected from the elements and safe from others require ongoing efforts [36]. Participants described several methods used to navigate these circumstances to charge their phones and access the tools available to them. However, developing and implementing those strategies was time-consuming and inefficient. Furthermore, in this study, some participants mentioned that being responsible for the study-assigned smartphone caused added worries in their daily lives. These barriers suggest that community partners and social service providers should consider allowing people experiencing homelessness to spend time in their facilities safely charging their devices.

In addition to connecting them with professionals, access to the tools on the smartphone also facilitated connections to friends and family. Emotional management is an essential component of self-management [34], and this connection to others enables social support and reciprocal relationships that not only support self-management but also enhance resilience [37]. Earlier evidence has identified the mental strength that people experiencing homelessness with chronic illness need to survive while living on the street [38], and our findings indicate that connection to others can enhance resilience and mental strength in this population. Specifically, findings from this study extend our understanding of resilience in the presence of adversity and how equipping people with proper tools can return a sense of agency and autonomy—both of which are needed for successful self-management. Indeed, the accomplishment of the daily tasks described in this study transcended the study expectations, indicating that people experiencing homelessness are motivated and ready to *do what needs to be done* to self-manage their lives. Thus, an important implication of the current work is the need to use asset-based approaches when working with homeless populations. Evidence reveals that people experiencing homelessness value health and understand the importance of self-management [21,22], and the resilience of people experiencing homelessness has been acknowledged by prior authors [39,40]. Capitalizing on these assets by equipping homeless populations with consistent and reliable tools for communication, access to information, and transportation could yield important benefits, including a sense of security and autonomy that can empower people experiencing homelessness to navigate the context of homelessness and reassert control over their health and lives.

Findings from this study also suggest that access to the tools and information available via smartphones serve more practical purposes as well. Participants were more easily able to access food and other humanitarian aid satisfying daily subsistence needs and, subsequently, increasing their capacity to pursue higher-order tasks, including scheduling and attending appointments with health and social services providers. Gobeil-Lavoie et al [41] identified that one of the main challenges to self-management is the prioritization of self-care and that patients with multiple conditions must choose which care activities to prioritize. They also found that most individuals with multiple chronic conditions identify a single disorder on which to focus their efforts and that social and economic conditions will influence how patients prioritize self-care [41]. Our participants were no different and reported multiple competing demands on their time. Similarly, prior research has identified a prevailing present time orientation among people experiencing homelessness, the unpredictability of time while experiencing homelessness, and a sense of not having enough time [42], precluding the ability of many individuals in this population to meet the expectations of our health and social services systems as they are currently structured. Indeed, most participants described how consistently knowing the time of day helped them create routines in a way that was not possible without the phone. Many also used the tools on smartphones such as alarms for reminders to take medications and manage appointments.

Limitations

There are limitations to this study that should be considered. First, participants in both of these studies were recruited from church-based locations. It is possible that people experiencing homelessness who access services from community-based locations such as churches may have different needs or different abilities than individuals who access services from more traditional locations such as clinics or shelters. In addition, the relative success that the participants had with managing their smartphones, including keeping them safe, keeping them charged, and using them to manage everyday tasks, may indicate that this sample is more cognitively intact than many people experiencing homelessness. Thus, the findings of this study should be interpreted with caution.

Conclusions

Homelessness is a complex problem that presents barriers to the provision of appropriate health and social services, and people experiencing homelessness must navigate complex bureaucratic systems while also managing complex comorbidities without much, if any support to do so. Resolving homelessness and providing appropriate services will require multifaceted systems approaches achieved, largely, through policy change. However, findings from this analysis reveal that individual-level interventions also play an important role in building capacity among people experiencing homelessness. Indeed, our findings indicate that equipped with the basic tools of smartphone technology and reliable access to transportation, people experiencing homelessness can and do engage in activities needed to self-manage their health and meet social needs. As communities across the United States and globally define and incorporate system-wide strategies to prevent and end homelessness, the relatively straightforward and comparatively inexpensive tools of smartphone technology and public transportation are smart investments to help individuals navigate homelessness and self-manage their own health and social needs. Our findings indicate that access to communication, information via the internet, and transportation facilitates the use of these higher-order skills to self-manage their health and attain social needs.

Conflicts of Interest

None declared.

References

1. State of homelessness: 2021 edition. National Alliance to End Homelessness. 2020. URL: https://endhomelessness.org/homelessness-in-america/homelessness-statistics/state-of-homelessness-2021/ [accessed 2020-12-02]

2. Yurcaba J. Homelessness during the COVID-19 pandemic: What to know. Verywell Mind. 2020. URL: https://www.verywellmind.com/homelessness-during-the-covid-19-pandemic-5087059 [accessed 2020-12-03]
3. Fazel S, Geddes JR, Kushel M. The health of homeless people in high-income countries: descriptive epidemiology, health consequences, and clinical and policy recommendations. Lancet 2014 Oct 25;384(9953):1529-1540 [FREE Full text] [doi: 10.1016/S0140-6736(14)61132-6] [Medline: 25390578]
4. O'Toole TP, Gibbon JL, Hanusa BH, Freyder PJ, Conde AM, Fine MJ. Self-reported changes in drug and alcohol use after becoming homeless. Am J Public Health 2004 May;94(5):830-835. [doi: 10.2105/ajph.94.5.830] [Medline: 15117708]
5. Garibaldi B, Conde-Martel A, O'Toole TP. Self-reported comorbidities, perceived needs, and sources for usual care for older and younger homeless adults. J Gen Intern Med 2005 Aug;20(8):726-730 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.0142.x] [Medline: 16050882]
6. 2017 AHAR: Part 1 - Point in time estimates of homelessness in the United States. HUD Exchange. URL: https://www.hudexchange.info/resource/5639/2017-ahar-part-1-pit-estimates-of-homelessness-in-the-us/ [accessed 2021-01-13]
7. Alegría M, NeMoyer A, Bagué IF, Wang Y, Alvarez K. Social determinants of mental health: where we are and where we need to go. Curr Psychiatry Rep 2018 Sep 17;20(11):95 [FREE Full text] [doi: 10.1007/s11920-018-0969-9] [Medline: 30221308]
8. Social determinants of health: Predictors of health among people without homes. National Health Care for the Homeless Council. 2016. URL: http://councilbackup.flywheelsites.com/wp-content/uploads/2011/09/fact-sheet_2016_social-determinants-of-health1.pdf [accessed 2020-12-04]
9. Klop HT, Evenblij K, Gootjes JR, de Veer AJ, Onwuteaka-Philipsen BD. Care avoidance among homeless people and access to care: an interview study among spiritual caregivers, street pastors, homeless outreach workers and formerly homeless people. BMC Public Health 2018 Sep 05;18(1):1095 [FREE Full text] [doi: 10.1186/s12889-018-5989-1] [Medline: 30185163]
10. Zlotnick C, Zerger S, Wolfe PB. Health Care for the Homeless: What we have learned in the past 30 years and what's next. Am J Public Health 2013 Dec;103(S2):199-205. [doi: 10.2105/ajph.2013.301586]
11. Martins D. Experiences of homeless people in the health care delivery system: a descriptive phenomenological study. Public Health Nurs 2008;25(5):420-430. [doi: 10.1111/j.1525-1446.2008.00726.x] [Medline: 18816359]
12. Nickasch B, Marnocha S. Healthcare experiences of the homeless. J Am Acad Nurse Pract 2009 Jan;21(1):39-46. [doi: 10.1111/j.1745-7599.2008.00371.x] [Medline: 19125894]
13. Fitzpatrick KM, Willis DE. Homeless and hungry: food insecurity in the land of plenty. Food Secur 2020 Nov 04;13(1):1-10 [FREE Full text] [doi: 10.1007/s12571-020-01115-x] [Medline: 33169087]
14. Keene DE, Guo M, Murillo S. "That wasn't really a place to worry about diabetes": Housing access and diabetes self-management among low-income adults. Soc Sci Med 2018 Jan;197:71-77 [FREE Full text] [doi: 10.1016/j.socscimed.2017.11.051] [Medline: 29222997]
15. Lombardi K, Pines J, Mazer-Amirshahi M, Pourmand A. Findings of a national dataset analysis on the visits of homeless patients to US emergency departments during 2005-2015. Public Health 2020 Jan;178:82-89. [doi: 10.1016/j.puhe.2019.09.003] [Medline: 31644986]
16. Coe A, Moczygemba L, Harpe S, Gatewood S. Homeless patients' use of urban emergency departments in the United States. J Ambul Care Manage 2015;38(1):48-58. [doi: 10.1097/jac.0000000000000034]
17. Saab D, Nisenbaum R, Dhalla I, Hwang SW. Hospital readmissions in a community-based sample of homeless adults: a matched-cohort study. J Gen Intern Med 2016 Sep 19;31(9):1011-1018 [FREE Full text] [doi: 10.1007/s11606-016-3680-8] [Medline: 27197973]
18. Richard AA, Shea K. Delineation of self-care and associated concepts. J Nurs Scholarsh 2011 Sep;43(3):255-264. [doi: 10.1111/j.1547-5069.2011.01404.x] [Medline: 21884371]
19. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Edu Counsel 2002 Oct;48(2):177-187. [doi: 10.1016/S0738-3991(02)00032-0]
20. Hartweg DL. Dorothea Orem: Self-Care Deficit Theory. Thousand Oaks, CA: Sage Publications; 1991.
21. Merdsoy L, Lambert S, Sherman J. Perceptions, needs and preferences of chronic disease self-management support among men experiencing homelessness in Montreal. Health Expect 2020 Dec;23(6):1420-1430 [FREE Full text] [doi: 10.1111/hex.13106] [Medline: 32965730]
22. Henwood BF, Cabassa LJ, Craig CM, Padgett DK. Permanent supportive housing: Addressing homelessness and health disparities? Am J Public Health 2013 Dec;103(S2):188-192. [doi: 10.2105/ajph.2013.301490]
23. Hulton LJ, Strang S, Brooks S, Hostetter M. A pilot study of chronic disease self-management in a homeless population. GSTF J Nurs HealthC 2015 Nov 30;3(1):30-34 [FREE Full text]
24. Savage C, Xu Y, Richmond MM, Corbin A, Falciglia M, Gillespie G. A pilot study: retention of adults experiencing homelessness and feasibility of a CDSM diabetes program. J Community Health Nurs 2014 Oct 30;31(4):238-248. [doi: 10.1080/07370016.2014.958406] [Medline: 25356993]
25. Moczygemba L, Thurman W, Tormey K, Hudzik A, Welton-Arndt L, Kim E. GPS mHealth Intervention amongpPeople experiencing homelessness: A feasibility study. JMIR MHealth and UHealth. November 2021:2020. [doi: 10.2196/preprints.25553]

26. Burda C, Haack M, Duarte A, Alemi F. Medication adherence among homeless patients: a pilot study of cell phone effectiveness. J Am Acad Nurse Pract 2012 Nov;24(11):675-681 [FREE Full text] [doi: 10.1111/j.1745-7599.2012.00756.x] [Medline: 23088699]
27. Glover AC, Schueller SM, Winiarski DA, Smith DL, Karnik NS, Zalta AK. Automated mobile phone-based mental health resource for homeless youth: pilot study assessing feasibility and acceptability. JMIR Ment Health 2019 Oct 11;6(10):e15144 [FREE Full text] [doi: 10.2196/15144] [Medline: 31605516]
28. McInnes DK, Petrakis BA, Gifford AL, Rao SR, Houston TK, Asch SM, et al. Retaining homeless veterans in outpatient care: A pilot study of mobile phone text message appointment reminders. Am J Public Health 2014 Sep;104(S4):588-594. [doi: 10.2105/ajph.2014.302061]
29. Rhoades H, Wenzel S, Winetrobe H, Ramirez M, Wu S, Carranza A, et al. A text messaging-based intervention to increase physical activity among persons living in permanent supportive housing: Feasibility and acceptability findings from a pilot study. Digit Health 2019 Feb 27;5:2055207619832438 [FREE Full text] [doi: 10.1177/2055207619832438] [Medline: 30834135]
30. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec 16;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
31. Taylor SP. Int J Sci 2018;37(2):216-222. [doi: 10.9734/bpi/sthss/v3/1736c]
32. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
33. Morse JM. Critical analysis of strategies for determining rigor in qualitative inquiry. Qual Health Res 2015 Sep;25(9):1212-1222. [doi: 10.1177/1049732315588501] [Medline: 26184336]
34. Lorig KR, Holman HR. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]
35. Holman H, Lorig K. Patients as partners in managing chronic disease. Partnership is a prerequisite for effective and efficient health care. Br Med J 2000 Feb 26;320(7234):526-527 [FREE Full text] [doi: 10.1136/bmj.320.7234.526] [Medline: 10688539]
36. McInnes DK, Li AE, Hogan TP. Opportunities for engaging low-income, vulnerable populations in health care: a systematic review of homeless persons' access to and use of information technologies. Am J Public Health 2013 Dec;103 Suppl 2:11-24. [doi: 10.2105/AJPH.2013.301623] [Medline: 24148036]
37. Vest BM, Kahn LS, Danzo A, Tumiel-Berhalter L, Schuster RC, Karl R, et al. Diabetes self-management in a low-income population: impacts of social support and relationships with the health care system. Chronic Illn 2013 Jun 12;9(2):145-155 [FREE Full text] [doi: 10.1177/1742395313475674] [Medline: 23585634]
38. Håkanson C, Öhlén J. Illness narratives of people who are homeless. Int J Qual Stud Health Well-being 2016 Nov 30;11(1):32924 [FREE Full text] [doi: 10.3402/qhw.v11.32924] [Medline: 27914194]
39. Dorsen C. Vulnerability in homeless adolescents: concept analysis. J Adv Nurs 2010 Dec;66(12):2819-2827 [FREE Full text] [doi: 10.1111/j.1365-2648.2010.05375.x] [Medline: 21073499]
40. Shankar S, Gogosis E, Palepu A, Gadermann AM, Hwang SW. "I Haven't Given Up and I'm Not Gonna": A phenomenographic exploration of resilience among individuals experiencing homelessness. Qual Health Res 2019 Nov 26;29(13):1850-1861. [doi: 10.1177/1049732318798353] [Medline: 30253692]
41. Gobeil-Lavoie A, Chouinard M, Danish A, Hudon C. Characteristics of self-management among patients with complex health needs: a thematic analysis review. BMJ Open 2019 May 24;9(5):e028344 [FREE Full text] [doi: 10.1136/bmjopen-2018-028344] [Medline: 31129599]
42. Fraenkel P. Time as a source of struggle and resilience in homeless families. Time Soc 2020 Apr 11;29(2):581-603. [doi: 10.1177/0961463x20908082]

Abbreviations

ED: emergency department

Edited by R Kukafka; submitted 06.02.21; peer-reviewed by S Melbye, R Grainger; comments to author 02.05.21; revised version received 23.06.21; accepted 05.07.21; published 29.09.21.

Please cite as:
Thurman W, Semwal M, Moczygemba LR, Hilbelink M
Smartphone Technology to Empower People Experiencing Homelessness: Secondary Analysis
J Med Internet Res 2021;23(9):e27787
URL: https://www.jmir.org/2021/9/e27787
doi:10.2196/27787
PMID:34586073

Original Paper

Exploring the Associations Between Self-reported Tendencies Toward Smartphone Use Disorder and Objective Recordings of Smartphone, Instant Messaging, and Social Networking App Usage: Correlational Study

Davide Marengo[1], PhD; Rayna Sariyska[2], PhD; Helena Sophia Schmitt[2], MSc; Eva-Maria Messner[3], PhD; Harald Baumeister[3], PhD; Matthias Brand[4], PhD; Christopher Kannen[2], MSc; Christian Montag[2], PhD

[1]Department of Psychology, University of Turin, Turin, Italy

[2]Department of Molecular Psychology, Institute of Psychology and Education, Ulm University, Ulm, Germany

[3]Department of Clinical Psychology and Psychotherapy, Institute of Psychology and Education, Ulm University, Ulm, Germany

[4]Department of General Psychology: Cognition and Center for Behavioral Addiction Research (CeBAR), Faculty of Engineering, University of Duisburg-Essen, Duisburg, Germany

Corresponding Author:
Christian Montag, PhD
Department of Molecular Psychology
Institute of Psychology and Education
Ulm University
Helmholtzstraße 8/1
Ulm, 89081
Germany
Phone: 49 07315026550
Email: christian.montag@uni-ulm.de

Abstract

Background: Social communication via instant messaging (IM) and social networking (SN) apps makes up a large part of the time that smartphone users spend on their devices. Previous research has indicated that the excessive use of these apps is positively associated with problematic smartphone use behaviors. In particular, image-based SN apps, such as Instagram (Facebook Inc) and Snapchat (Snap Inc), have been shown to exert stronger detrimental effects than those exerted by traditional apps, such as Facebook (Facebook Inc) and Twitter (Twitter Inc).

Objective: In this study, we investigated the correlation between individuals' tendencies toward smartphone use disorder (SmUD) and objective measures of the frequency of smartphone usage. Additionally, we put to test the hypothesis that the pathway linking the frequency of actual smartphone usage to self-reported tendencies toward SmUD was mediated by the increased frequency of IM and SN app usage.

Methods: We recruited a sample of 124 adult smartphone users (females: 78/124, 62.9%; age: mean 23.84 years, SD 8.29 years) and collected objective information about the frequency of smartphone and SN app usage over 1 week. Participants also filled in a self-report measure for assessing the multiple components of tendencies toward SmUD. Bivariate associations were investigated by using Spearman correlation analyses. A parallel mediation analysis was conducted via multiple regression analysis.

Results: The frequency of smartphone usage, as well as the use of IM apps (Messenger, Telegram, and WhatsApp [Facebook Inc]), Facebook, and image-based apps (Instagram and Snapchat), had significant positive associations with at least 1 component of SmUD, and the cyberspace-oriented relationships factor exhibited the strongest associations overall. We found support for an indirect effect that linked actual smartphone usage to SmUD tendencies via the frequency of the use of image-based SN apps.

Conclusions: Our novel results shed light on the factors that promote SmUD tendencies and essentially indicate that image-based SN apps seem to be more strongly associated with problematic smartphone behaviors compared to IM apps and traditional SN apps, such as Facebook.

(J Med Internet Res 2021;23(9):e27093) doi:10.2196/27093

KEYWORDS

smartphone use disorder; smartphone use; social media; objective measures; mediation model; smartphone; web-based communication; social networking; mobile phone

Introduction

Web-based social communication via instant messaging (IM) and social networking (SN) apps makes up a large part of the time that smartphone users spend on their devices, and this time exceeds the time that smartphone users spend on other activities, including being productive and gaming [1]. Among the most popular IM apps are WhatsApp (Facebook Inc), Facebook Messenger (Facebook Inc), WeChat (Tencent Holdings Limited), Telegram, and Signal (Signal Messenger LLC), with WhatsApp being the most popular IM app worldwide [2]. IM apps allow for the one-to-one and group-based transmission of messages and vary considerably in terms of their specific characteristics and affordances (eg, file sharing and voice calling). Popular SN apps include Facebook (Facebook Inc), Instagram (Facebook Inc), Twitter (Twitter Inc), Snapchat (Snap Inc), and TikTok, with Facebook being the most popular SN app [3]. SN apps feature microblogging capabilities that allow smartphone users to broadcast messages, pictures, or videos to an extensive network of smartphone users and receive feedback (ie, comments and likes) on the shared content—a feature that is distinctive from those offered by IM apps and may play a crucial role in triggering problematic behaviors [4,5]. Currently, SN apps are shifting toward image-based ephemeral media sharing (ie, media that are available for a limited time)—a feature that was introduced by image-based apps such as Snapchat and Instagram but has also been implemented on more traditional platforms (eg, the recent introduction of Facebook's time-limited stories feature). Notably, such a time restriction on content availability might pressure smartphone users to watch or read a story instantly. Otherwise, one might miss crucial information in one's social network.

Previous findings have indicated that the excessive use of SN apps on smartphones is positively associated with problematic behaviors related to the use of devices in general [6-8]. These findings suggest that the link between problematic smartphone use and smartphone use disorder (SmUD; a discussion of terminology was conducted by Montag and colleagues [9] in 2021) can, in part, be explained by the use and overuse of SN apps on smartphones. It is worthy to note that many of the available inventories that assess problematic behaviors in this area of research use an addiction framework. However, the question regarding the actual nature of problematic smartphone and SN use has not been ultimately settled until now. In line with the term *disorder*, which is used to describe behavioral addictions in the *International Classification of Diseases, 11th edition* (eg, gaming disorder) as well as those in previous literature related to this area of research [7,10], we chose to use the term SmUD in this paper. Still, we explicitly stress the importance of refraining from pathologizing everyday life behaviors [11]. Moreover, we explicitly mention that so far, SmUD does not represent an official diagnosis, and in this paper, we aimed to unify the terminology used in related literature without judging on the nature of the construct underlying the term itself. It must also be mentioned that the concept of SmUD is being fiercely discussed by researchers, as smartphones are only vehicles for accessing various apps, which in turn may trigger the addiction to the device itself [12], similar to how someone diagnosed with an alcohol use disorder is not dependent on the bottle but on the beverage. Beyond this, it is very likely that for many smartphone users, SmUD might resemble social networks use disorder to a high degree due to the substantial overlap of concepts [7,13]. Further, we also want to hint at recent works that discuss what prerequisites need to be met to speak of a disorder that results from problematic behaviors and might be classified within the residual category of "other specified disorders due to addictive behaviors" in the *International Classification of Diseases, 11th edition*. The criteria suggested are (1) scientific relevance for clinical relevance, (2) theoretical embedding within addiction frameworks, and (3) empirical evidence for the validity of mechanisms that underlie addictive behaviors. In terms of a potential social networks use disorder, these criteria are widely met, although the scientific evidence for clinical relevance and functional impairment are still less convincing compared to the evidence for other disorders resulting from addictive behaviors, such as gaming disorder [14].

It is noteworthy that although we distinguish between messaging and SN apps in this paper, SN can be seen as an umbrella term that also includes messenger apps [15]. For our research, it is of interest that increasing amounts of evidence indicate that image-based apps such as Instagram and Snapchat may exert stronger detrimental effects on adverse health-related outcomes, including sleep problems, psychological distress, and problematic behaviors, compared to those exerted by more traditional SN platforms such as Facebook and Twitter [16-18]. Interestingly, a new study by Rozgonjuk et al [19] reported that the problematic use of WhatsApp and Instagram was the factor that was the most strongly linked to productivity loss in everyday life (ie, compared to Facebook and Snapchat use). Another work has also shown that these two platforms might be associated with the most problematic consequences for smartphone users [20]. These studies are based on self-report data; however, recent literature has explored the associations between smartphone and IM and SN app usage and SmUD based on objective measures [17,21,22] (a related review was conducted by Ryding and Kuss [23]). Some of the studies in this area of research focus on overall measures of smartphone usage, such as screen time and screen unlock events, but do not explore app usage measures [24,25]. Additionally, other studies have provided evidence of the links between SmUD variables and broad app usage categories (eg, chatting [26] and SN use [21,22]). We could only find 1 study [17] that distinguished between traditional and image-based SN platforms when examining the link between app usage and SmUD. The results provided by Noë and colleagues [17] showed that Snapchat use was more strongly related to self-reported tendencies toward SmUD than the use of other IM and SN apps (eg, WhatsApp

and Facebook)—a finding that the authors interpreted as the result of the in-built design features in Snapchat that promote high-frequency usage. The findings of Rozgonjuk et al [20] differ from those of Noë et al [17]; by using self-report data, Rozgonjuk et al [20] reported that the lowest magnitude of correlations was found between Snapchat use and SmUD incidence and that the highest correlations were found between SmUD and problematic WhatsApp and Instagram use. Nevertheless, both of these findings point toward the need to distinguish between IM, traditional, and image-based SN apps when examining their associations with SmUD.

To build on these findings, in this study, we aimed to examine the existing associations between objective records of smartphone, IM platform, and SN platform activity and individual differences in self-reported SmUD. More specifically, we tested the following hypotheses:

- Hypothesis 1: Significant associations exist between self-reported tendencies toward SmUD and objective measures of the usage of the smartphone and IM and SN platforms.
- Hypothesis 2: The association between self-reported tendencies toward SmUD and overall smartphone usage is mediated by the frequency of IM and SN app usage (both traditional and image-based apps).

Based on the existing literature about theoretical considerations and previous empirical findings [6-8], we expected that a positive link would emerge between actual smartphone usage and users' perceptions of being at risk for SmUD. We also expected that this link might, in part, be mediated by the usage of both IM and SN apps and that image-based SN apps would have a stronger mediation effect [17].

Methods

Sample

Our sample consisted of individuals who were recruited among students at a German university campus by using a convenience sampling strategy. The inclusion criteria for this study were legal age (18 years and over) and current ownership of a smartphone. Participants were recruited between March 2017 and August 2018 and consisted of 124 individuals with a mean age of 23.84 years (SD 8.29 years; range 18-63 years) and a majority of female individuals (78/124, 62.9%). Participants gained university credits or monetary compensation for their participation. With regard to education level, of the 124 participants, 2 (1.6%) reported having finished secondary school, 10 (8.1%) reported holding a vocational baccalaureate diploma, 92 (74.2%) reported holding an A-level diploma, 3 (2.4%) reported holding a degree from a university of applied sciences, and 17 (13.7%) reported holding a university degree. Prior to their participation in this study, participants were provided with detailed information on the aim of this study, all variables that were tracked by the Insights app (see also the *Objective Smartphone Activity Data* section), how data were recorded, who was granted access to their data, and where data were saved and analyzed. Participants were informed about their right to request the deletion of their data at any time during this study. All participants provided digital informed consent via the app prior to their participation in this study. In order to facilitate the installation process, participants were invited to the laboratory, where scientists guided the participants through the Insights app installation process. Therefore, no problems with the installation process were observed. The study was approved by the local ethics committee of Ulm University, Ulm, Germany. It should be noted that in this study's data set, there was an overlap among other publications—there was 1 study that investigated links between call variables and personality [27]; 1 study that investigated the molecular genetics of social network size [28]; and 1 study that investigated links between smartphone variables and self-reported mood, drive, and stress levels [29].

Objective Smartphone Activity Data

In order to collect participants' smartphone activity data, the Insights app [30]—an Android-based smartphone app—was installed on participants' phones by either the examiner or the participants themselves. The app allowed for the recording of different variables, such as the number of calls per day (incoming, outgoing, and missed calls), the frequency of specific smartphone events (eg, screen unlock and screen on events), and data related to users' sessions on smartphone apps. An in-depth review of the Insights app was conducted by Montag et al [27].

Participants varied considerably in terms of the duration of study participation, which ranged from 1 week to 2 months (based on recorded activity). Based on considerations about the potential incomparability of data that are collected during different time frames and the need to maximize our sample size for this study, we decided to limit our analyses to smartphone data that were collected during the first week (7 days) of smartphone data collection. Given this time frame, we retrieved the following information: the number of smartphone sessions that were started by the user unlocking the smartphone screen (ie, the number of screen unlock events) and the number of sessions on IM apps (ie, Facebook Messenger, Telegram, and WhatsApp), traditional SN apps (ie, Facebook and Twitter), and image-based SN apps (ie, Instagram, Snapchat, and TikTok). Participants were coded as having 0 sessions on the app if they either had the app installed on their smartphones and did not use it or did not have the app installed on their smartphones. We followed such a strategy to conduct analyses on the complete sample and to take into account that not each app of interest was installed on each phone. The Insights app allowed for the tracking of both the screen unlock events and screen on events of Android smartphones; each event represented an indicator of users' active use of their smartphone. However, while both screen unlock events and screen on events were triggered through users' interaction with their smartphone's hardware (ie, the screen or buttons), screen on events could also be triggered by app notifications (depending on the smartphone settings). Due to this potential confounding factor, for the purpose of this paper, analyses were presented by using the screen unlock event variable. However, for full transparency, we presented and commented on the associations between screen on and screen unlock events and app session and SmUD variables in Multimedia Appendix 1.

Based on the collected data, WhatsApp was the most used app during the investigated time frame (participants: 122/124, 98.4%), followed by Facebook (70/124, 56.5%), Messenger (60/124, 48.8%), Instagram (56/124, 45.2%), Snapchat (41/124, 33.1%), and Telegram (13/124, 10.50%). Of the 124 participants, 123 (99.2%) used IM apps and 67 (54%) used image-based SN apps. In turn, we found that no participants used the TikTok app; however, it should be noted that at the time data collection started for the present study, the TikTok app was called "musical.ly" [31], and was later renamed TikTok starting on November 2017. Further, only 3 participants reported using the Twitter app. Due to the small number of users, the TikTok and Twitter apps were not considered further in this study.

Individual Differences in Tendencies Toward SmUD

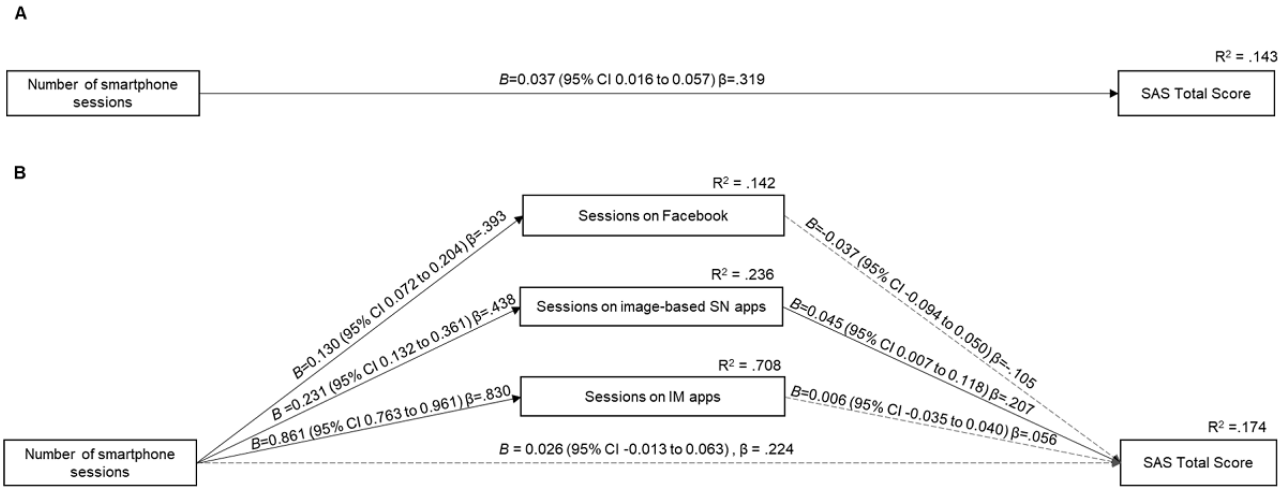
We administered a German version of the Smartphone Addiction Scale (SAS) [32], which consists of 33 items that are rated on a 6-point Likert scale (1=strongly disagree; 6=strongly agree). It is worthy to note that the SAS scale is not a diagnostic tool for SmUD but is instead meant to provide an assessment of users' addictive tendencies toward using smartphones. To keep in line with previous studies [7,10] however, in this study, the SAS total score was treated as an indicator of a general tendency toward SmUD. The SAS items assess 6 components of smartphone addiction, namely daily life disturbances (eg, "Missing planned works due to smartphone usage"), positive anticipation (eg, "Feeling pleasant or excited while using a smartphone"), withdrawal (eg, "Having my smartphone in my mind even when I'm not using it"), cyberspace-oriented relationships (eg, "Feeling that my smartphone buddies understand me better than my real-life friends"), overuse (eg, "Using my smartphone longer than I had intended"), and tolerance (eg, "The people around me tell me that I use my smartphone too much"). The items were summed to compute an overall indicator of the tendency toward SmUD (SAS total score: mean 70.10, SD 22.84). For the purpose of this study, the SAS total score had excellent reliability (Cronbach α=.94). The reliability of the 6 components was also adequate (daily life disturbances: α=.80; positive anticipation: α=.80; withdrawal: α=.80; cyberspace-oriented relationships: α=.77; overuse: α=.69; tolerance: α=.86). It should be noted that for most of the participants (74/124, 59.7%), the SAS was administered within 1 month of the beginning of smartphone data collection; the remaining participants filled in the SAS within 6 months (14/124, 11.3%) or more than 6 months (36/124, 29%) from the beginning of smartphone data collection. In regression analyses, we controlled for this effect by using a covariate that represented the time (in months) between self-reports and smartphone data collection.

Strategy of Analysis

First, we computed descriptive statistics (the mean and SD of the number of app and smartphone sessions). Second, we calculated Spearman correlation values between the number of app and smartphone sessions and age and between the number of such sessions and indicators of tendencies toward SmUD (ie, SAS subscales and total score). Mann-Whitney U tests were performed to evaluate the association between gender and smartphone and app session variables. In all inferential tests, results were considered significant when P was <.05.

Third, we performed a mediation analysis via multiple linear regression models for predicting the overall tendency toward SmUD. The diagrams for the examined models are shown in Figure 1. As a first step, we examined the total effect that participants' frequency of smartphone use (ie, the number of smartphone sessions started by a screen unlock) had on tendencies toward SmUD. In the mediation model, participants' number of sessions on Facebook, image-based SN apps (ie, Instagram and Snapchat), and IM apps (ie, WhatsApp, Telegram, and Messenger) were investigated as multiple parallel mediators of the link between the frequency of smartphone use (ie, the number of smartphone sessions started by a screen unlock) and tendencies toward SmUD. In all analyses, we controlled for the effects of gender, age, and the time of the administration of the SAS (ie, the number of months after the collection of smartphone data). Analyses were performed by using 10,000 bootstrap samples to compute 95% CIs. Bootstrapping was used, as it did not impose assumptions on the distribution of residuals. Moreover, the use of bootstrapping is appropriate when potentially dealing with nonnormally distributed variables [33]. Estimated effects were deemed significant if 95% (bootstrap) CIs did not span 0. Finally, pairwise contrasts of indirect effects were used to test whether differences in the size of indirect effects were statistically significant (P<.05). Analyses were performed with SPSS 23 (IBM Corporation) by using the Process macro [34].

Figure 1. Total effect (A) and mediation (B) models for smartphone use disorder. For ease of visualization, the results of covariates are not reported in the figure. IM: instant messaging; SAS: Smartphone Addiction Scale; SN: social networking.



Results

Descriptive Statistics and Correlations

Table 1 depicts descriptive statistics for app and smartphone session variables as well as the associations between these variables and gender. Table 2 provides information about the associations among app and smartphone session variables, age, and indicators of tendencies toward SmUD (SAS subscales and total score). Based on the number of weekly app sessions (Table 1), WhatsApp was the most frequently used app in the overall sample, followed by Facebook, Instagram, Snapchat, Messenger, and Telegram.

Table 1. The number of weekly sessions on apps and smartphones: descriptive statistics of the overall sample stratified by gender[a].

App or device	Number of sessions/week, mean (SD)			*U*	*P* value
	All participants (N=124)	Males (n=46)	Females (n=78)		
Facebook	43.25 (65.78)	45.61 (79.78)	41.86 (56.45)	1843.50	.79
Image-based social networking apps	61.81 (104.96)	50.09 (126.24)	68.72 (90.31)	2098.50	.10
Snapchat	21.86 (52.03)	22.24 (60.3)	21.64 (46.91)	1841.50	.77
Instagram	39.34 (70.35)	27.85 (74.98)	47.08 (66.94)	2098.50	.06
Instant messaging apps	327.31 (219.56)	336.26 (207.92)	322.03 (207.41)	1767.50	.89
Telegram	6.41 (40.10)	4.37 (13.88)	7.62 (49.52)	1773.00	.84
WhatsApp	319.10 (219.56)	332.96 (226.82)	310.92 (216.24)	1742.50	.79
Messenger	8.05 (22.57)	6.57 (15.53)	8.92 (25.9)	1976.00	.31
Smartphone use	347.26 (199.43)	331.54 (184.20)	356.53 (208.48)	1945.00	.44

[a]Gender was coded (male=0; female=1).

Table 2. The correlations between app and smartphone usage variables and age and between usage variables and the individual differences in tendencies toward smartphone use disorder (N=124).

App or device	Variables, Spearman ρ (*P* values)							
	Age	Life disturbances	Positive anticipation	Withdrawal	CY[a]	Overuse	Tolerance	Total SAS[b] score
Facebook	−0.06 (.53)	−0.01 (.96)	0.15 (.10)	0.14 (.12)	0.31 (<.001)	0.07 (.48)	−0.01 (.95)	0.14 (.11)
Image-based social networking apps	−0.36 (<.001)	0.31 (<.001)	0.29 (.001)	0.27 (.003)	0.40 (<.001)	0.28 (.002)	0.35 (.001)	0.40 (<.001)
Snapchat	−0.33 (<.001)	0.23 (.009)	0.19 (.03)	0.23 (.009)	0.30 (<.001)	0.24 (.007)	0.25 (.006)	0.30 (.006)
Instagram	−0.26 (.003)	0.26 (.003)	0.25 (.006)	0.18 (.049)	0.39 (<.001)	0.22 (.01)	0.30 (<.001)	0.33 (<.001)
Instant messaging apps	−0.26 (.004)	0.18 (.048)	0.33 (<.001)	0.24 (.007)	0.39 (<.001)	0.13 (.15)	0.20 (.03)	0.30 (<.001)
Telegram	−0.03 (.72)	−0.08 (.37)	0.16 (.07)	−0.03 (.73)	0.13 (.14)	−0.01 (.95)	0.09 (.31)	0.03 (.72)
WhatsApp	−0.24 (.006)	0.15 (.10)	0.28 (.002)	0.21 (.02)	0.36 (<.001)	0.09 (.33)	0.16 (.08)	0.26 (.003)
Messenger	−0.11 (.21)	0.12 (.17)	0.23 (.009)	0.25 (.005)	0.33 (<.001)	0.21 (.02)	0.15 (.10)	0.26 (.004)
Smartphone use	−0.30 (<.001)	0.23 (.01)	0.38 (<.001)	0.30 (<.001)	0.44 (<.001)	0.20 (.03)	0.23 (.01)	0.38 (<.001)

[a]CY: cyberspace-oriented relationship.

[b]SAS: Smartphone Addiction Scale.

The results of Mann-Whitney U (Table 1) tests indicated that there was no significant association between gender and smartphone and app session variables. With regard to correlations (Table 2), we found that all of the weekly app and smartphone session variables, except those for Facebook (*P*=.53), Messenger (*P*=.21), and Telegram (*P*=.72) sessions, had a significant negative correlation with age. We also analyzed correlations between the objectively recorded SN app, IM app, and smartphone session variables and the indicators of tendencies toward SmUD variables. The daily life disturbances score had positive correlations with the number of Instagram sessions, Snapchat sessions, image-based SN app sessions, IM app sessions, and overall smartphone sessions, but not with the number of Facebook and each specific IM app sessions. Positive anticipation, withdrawal, and the general tendency toward SmUD (ie, the SAS total score) had positive correlations with all of the app and smartphone session variables except for the number of Facebook and Telegram sessions. Cyberspace-oriented relationships had moderate positive correlations with all app and smartphone session variables except for the Telegram session variable. The overuse subscale scores positively correlated with the number of participants' Snapchat, Instagram, image-based SN app, and Messenger sessions and the number of overall smartphone sessions. Finally, tolerance positively correlated with the number of participants' Snapchat, Instagram, image-based SN app, and IM app sessions (tolerance did not show significant correlations with sessions on specific IM apps) and the number of overall smartphone sessions.

Mediation Analyses

We also examined the total effect and mediation models for analyzing the association between the number of smartphone sessions and participants' general tendency toward SmUD (ie, the SAS total score). The results of total and direct effects are shown in Figure 1. For ease of visualization, the results of control variables (ie, age, gender, and the time of the administration of the SAS) are not shown in the figure but are only discussed in the text. In the total effect model, we found that the number of smartphone sessions had a positive association with participants' tendency toward SmUD. In the mediation model, after including Facebook, image-based SN app, and IM app session variables in the model, the number of smartphone sessions (ie, screen unlocks) no longer showed a significant association with tendencies toward SmUD. In turn, participants' number of sessions on image-based SN apps had a significant positive association with tendencies toward SmUD. In contrast, the number of sessions on Facebook and IM apps had no significant association with tendencies toward SmUD. In all tested models, associations between control variables and tendencies toward SmUD were nonsignificant.

The examination of indirect effects (Table 3) showed that the association between the number of smartphone sessions and tendencies toward SmUD was mediated by the number of sessions spent on image-based SN apps (effect=0.010; 95% CI 0.002-0.024; standardized effect=0.090). The remaining indirect effects were not significant. Finally, based on the pairwise contrast of unstandardized indirect effects, there was a significant difference between the indirect effects in the Facebook and image-based SN app session routes (contrast=−0.015; 95% CI −0.032 to −0.001); however, the remaining contrasts were not significant (Facebook vs IM apps contrast: −0.010; 95% CI −0.043 to 0.029; image-based SN vs IM app contrast: −0.005; 95% CI −0.046 to 0.025).

Table 3. Indirect effects linking smartphone use to tendencies toward smartphone use disorder stratified by routes of indirect effects.

Route of indirect effects	Effect (95% CI)	Standardized effect (95% CI)
Session on Facebook	−0.005 (−0.016 to 0.006)	−0.041 (−0.132 to 0.049)
Session on image-based apps	0.010 (0.002 to 0.024)	0.090 (0.015 to 0.209)
Session on instant messaging apps	0.005 (−0.029 to 0.036)	0.046 (−0.250 to 0.319)

Discussion

Principal Findings

This study investigated the associations between individual differences in self-reported tendencies toward SmUD and a set of objectively measured smartphone activity frequency and IM and SN app usage indicators. First, we explored the hypothesis that objectively measured indicators of smartphone and app usage would have a significant association with self-reported tendencies toward SmUD. Our data support this hypothesis. From a purely quantitative standpoint, based on our findings, the association between actual smartphone usage (assessed via screen unlock events) and overall tendencies toward SmUD is moderate (ρ=0.38) and appears to be close to the upper limit of what has been reported in previous studies that used objective smartphone data [24,25]. Further, our results showed that an indicator of the usage of image-based SN apps (ie, Instagram and Snapchat) had the strongest association with SmUD tendencies when compared with indicators of the use of IM apps and Facebook. It is noteworthy that among the image-based SN apps, the use of Instagram had a relatively stronger association with tendencies toward SmUD variables (ρ: range 0.22-0.39) when compared with the use of Snapchat (ρ: range 0.19-0.30).

The overall use of IM apps was positively related to all of the components of tendencies toward SmUD except for the overuse component. In turn, Facebook use was only associated with the cyberspace-oriented relationship component of SmUD. Moreover, among the different components of tendencies toward SmUD, the cyberspace-oriented relationship score had the strongest association with the usage of both IM and SN apps as well as overall smartphone usage. It should be noted that in our previous work in investigating SmUD among the users of different IM and SN platforms via self-report questionnaires, the highest associations were observed between SmUD and problematic Instagram and WhatsApp use, and the lowest (but still significant) associations were observed between SmUD and Snapchat and Facebook use [20].

We also examined our second hypothesis and tested the potential role of the usage of IM and SN apps as a mediator of the link between smartphone usage (ie, the frequency of screen unlocks) and self-reported tendencies toward SmUD. We go beyond previous studies by exploring the associations between smartphone and app usage variables and self-reported tendencies toward SmUD by using a parallel mediation model. Our findings supported the hypothesis that the association between the frequency of overall smartphone use and tendencies toward SmUD could be mediated by the frequency of the usage of apps that allow for web-based social communication (ie, IM and SN apps). In particular, our analyses indicated the existence of an indirect effect that links smartphone use and tendencies toward SmUD via the image-based SN app session route (ie, the frequency of the use of such apps). However, this indirect effect does not exist in the IM app session and traditional SN app session routes.

Overall, combined with the findings that emerged from the correlation analysis, the results from the mediation analyses are in line with those of previous studies that suggest the importance of social motives for smartphone use as predictors of tendencies toward SmUD [6,8] as well as the relative stronger association between tendencies toward SmUD and the use of image-based SN apps, such as Instagram and Snapchat, when compared with the association between such tendencies and the use of other IM and SN platforms [17]. Further, by highlighting the importance of social motives for both increased smartphone use and SN app use, our findings seem to be in line with a compensation-seeking hypothesis for SN app use (ie, using SN platforms to compensate for perceived social deficits [35]) and, possibly, smartphone use.

Strengths and Limitations

This study has strengths. Since we used a validated instrument for the assessment of tendencies toward SmUD and smartphone activity data were obtained by the use of an objective assessment procedure (ie, a specifically devised smartphone app), our results appear to be more reliable and valid than those reported by studies that relied solely on self-report measures of smartphone activity. The fact that our results were, in several ways, in line with those reported by previous studies further supports the validity of our findings. Future studies should also use measures for directly assessing individual differences in social networks use disorder and contrast their findings with those that we observed in our study (ie, in the context of tendencies toward SmUD).

This study also has several limitations. First, analyses were performed on a relatively small convenience sample that consisted mostly of young adults aged equal or below 30 years old (116/124, 93.5%). Thus, our results cannot be generalized to the general population. The relatively small sample size may also have had a negative impact on statistical power; based on the posthoc computation, our sample size appears to be large enough to detect medium to large correlations and regression effects, but it was not large enough to detect small associations among the examined variables in a robust way. These limitations are common in psychological studies that are based on objective assessments of smartphone use [23]. Still, replication studies with larger representative samples are needed to improve the quality and generalizability of our results.

Conclusion

This study provides novel results that shed light on the association between individuals' use of popular SN and IM apps

and tendencies toward SmUD. The results from nonzero correlations and regression analyses seem to suggest that the features implemented in image-based SN apps could be more strongly related to individual differences in tendencies toward SmUD compared to those available in IM apps and traditional SN apps, such as Facebook. Further, we also mention our previous work [20] in using a self-report methodology to demonstrate the robust associations between SmUD and problematic WhatsApp use. However, this study supports the idea that smartphones are just vehicles for accessing many different apps. Our data further indicate that reducing the problematic use of social media and messaging apps (particularly image-based SN app use) might help to reduce the overall SmUD tendencies of a person. Of interest is a recent study [36] that suggests that switching the smartphone mode from color mode to gray mode helps to reduce the time spent on smartphones and social media apps. Image-based SN apps might be particularly less interesting when images are only presented in gray mode.

Conflicts of Interest

The authors have no conflicts of interest to declare. Nevertheless, for reasons of transparency, CM mentions that he has received grants from agencies such as the German Research Foundation. CM has performed grant reviews for several agencies, edited journal sections and articles, given academic lectures in clinical or scientific venues or companies, and generated books or book chapters for publishers of mental health texts. For some of these activities, he received royalties, but he has never received royalties from gaming or social media companies. CM mentions that he is part of a discussion circle (Digitalität und Verantwortung [37]) for debating ethical questions linked to social media, digitalization, society, and democracy at Facebook. In this context, he receives no salary for his activities. Finally, he mentions that he currently functions as an independent scientist on the scientific advisory board of the Nymphenburg group. This activity is financially compensated.

Multimedia Appendix 1
Spearman correlations among features extracted from smartphone log data, age, and smartphone addiction variables.
[DOCX File , 18 KB - jmir_v23i9e27093_app1.docx]

References

1. The state of mobile 2020. App Annie. URL: https://www.appannie.com/en/go/state-of-mobile-2020/ [accessed 2021-08-10]
2. Most popular messaging apps. Statista. 2021. URL: https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/ [accessed 2021-08-10]
3. Most used social media 2021. Statista. URL: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ [accessed 2021-08-10]
4. Marengo D, Poletti I, Settanni M. The interplay between neuroticism, extraversion, and social media addiction in young adult Facebook users: Testing the mediating role of online activity using objective data. Addict Behav 2020 Mar;102:106150. [doi: 10.1016/j.addbeh.2019.106150] [Medline: 31706139]
5. Montag C, Lachmann B, Herrlich M, Zweig K. Addictive features of social media/messenger platforms and freemium games against the background of psychological and economic theories. Int J Environ Res Public Health 2019 Jul 23;16(14):2612 [FREE Full text] [doi: 10.3390/ijerph16142612] [Medline: 31340426]
6. Elhai JD, Levine JC, Dvorak RD, Hall BJ. Non-social features of smartphone use are most related to depression, anxiety and problematic smartphone use. Comput Human Behav 2017 Apr;69:75-82. [doi: 10.1016/j.chb.2016.12.023]
7. Sha P, Sariyska R, Riedl R, Lachmann B, Montag C. Linking internet communication and smartphone use disorder by taking a closer look at the Facebook and WhatsApp applications. Addict Behav Rep 2018 Nov 23;9:100148 [FREE Full text] [doi: 10.1016/j.abrep.2018.100148] [Medline: 31193857]
8. Zhitomirsky-Geffet M, Blau M. Cross-generational analysis of predictive factors of addictive behavior in smartphone usage. Comput Human Behav 2016 Nov;64:682-693. [doi: 10.1016/j.chb.2016.07.061]
9. Montag C, Wegmann E, Sariyska R, Demetrovics Z, Brand M. How to overcome taxonomical problems in the study of internet use disorders and what to do with "smartphone addiction"? J Behav Addict 2019 Oct 31:1-7. [doi: 10.1556/2006.8.2019.59] [Medline: 31668089]
10. Gao Q, Jia G, Fu E, Olufadi Y, Huang Y. A configurational investigation of smartphone use disorder among adolescents in three educational levels. Addict Behav 2020 Apr;103:106231. [doi: 10.1016/j.addbeh.2019.106231] [Medline: 31862619]
11. Billieux J, Schimmenti A, Khazaal Y, Maurage P, Heeren A. Are we overpathologizing everyday life? A tenable blueprint for behavioral addiction research. J Behav Addict 2015 Sep;4(3):119-123 [FREE Full text] [doi: 10.1556/2006.4.2015.009] [Medline: 26014667]
12. Panova T, Carbonell X. Is smartphone addiction really an addiction? J Behav Addict 2018 Jun 01;7(2):252-259 [FREE Full text] [doi: 10.1556/2006.7.2018.49] [Medline: 29895183]
13. Leung H, Pakpour AH, Strong C, Lin YC, Tsai MC, Griffiths MD, et al. Measurement invariance across young adults from Hong Kong and Taiwan among three internet-related addiction scales: Bergen Social Media Addiction Scale (BSMAS),

Smartphone Application-Based Addiction Scale (SABAS), and Internet Gaming Disorder Scale-Short Form (IGDS-SF9) (Study Part A). Addict Behav 2020 Feb;101:105969. [doi: 10.1016/j.addbeh.2019.04.027] [Medline: 31078344]
14. Brand M, Rumpf HJ, Demetrovics Z, MÜller A, Stark R, King DL, et al. Which conditions should be considered as disorders in the International Classification of Diseases (ICD-11) designation of "other specified disorders due to addictive behaviors"? J Behav Addict. Epub ahead of print 2020 Jun 30. [doi: 10.1556/2006.2020.00035] [Medline: 32634114]
15. Carr CT, Hayes RA. Social media: Defining, developing, and divining. Atl J Commun 2015 Feb 06;23(1):46-65. [doi: 10.1080/15456870.2015.972282]
16. Marengo D, Longobardi C, Fabris MA, Settanni M. Highly-visual social media and internalizing symptoms in adolescence: The mediating role of body image concerns. Comput Human Behav 2018 May;82:63-69. [doi: 10.1016/j.chb.2018.01.003]
17. Noë B, Turner LD, Linden DEJ, Allen SM, Winkens B, Whitaker RM. Identifying indicators of smartphone addiction through user-app interaction. Comput Human Behav 2019 Oct;99:56-65 [FREE Full text] [doi: 10.1016/j.chb.2019.04.023] [Medline: 31582873]
18. #StatusOfMind: social media and young people's mental health and wellbeing. Royal Society for Public Health. 2017 May. URL: https://www.rsph.org.uk/static/uploaded/d125b27c-0b62-41c5-a2c0155a8887cd01.pdf [accessed 2021-08-17]
19. Rozgonjuk D, Sindermann C, Elhai JD, Montag C. Fear of missing out (FoMO) and social media's impact on daily-life and productivity at work: Do WhatsApp, Facebook, Instagram, and Snapchat use disorders mediate that association? Addict Behav 2020 Nov;110:106487. [doi: 10.1016/j.addbeh.2020.106487] [Medline: 32674020]
20. Rozgonjuk D, Sindermann C, Elhai JD, Montag C. Comparing smartphone, WhatsApp, Facebook, Instagram, and Snapchat: Which platform elicits the greatest use disorder symptoms? Cyberpsychol Behav Soc Netw 2021 Feb;24(2):129-134. [doi: 10.1089/cyber.2020.0156] [Medline: 32907403]
21. Choi J, Rho MJ, Kim Y, Yook IH, Yu H, Kim DJ, et al. Smartphone dependence classification using tensor factorization. PLoS One 2017 Jun 21;12(6):e0177629. [doi: 10.1371/journal.pone.0177629] [Medline: 28636614]
22. Shin M, Lee K. Measuring smartphone usage time is not sufficient to predict smartphone addiction. Journal of Theoretical and Applied Information Technology 2017 Oct;95(19):5296-5303 [FREE Full text]
23. Ryding FC, Kuss DJ. Passive objective measures in the assessment of problematic smartphone use: A systematic review. Addict Behav Rep 2020 Jan 27;11:100257 [FREE Full text] [doi: 10.1016/j.abrep.2020.100257] [Medline: 32467846]
24. Rozgonjuk D, Levine JC, Hall BJ, Elhai JD. The association between problematic smartphone use, depression and anxiety symptom severity, and objectively measured smartphone use over one week. Comput Human Behav 2018 Oct;87:10-17. [doi: 10.1016/j.chb.2018.05.019]
25. Ellis DA, Davidson BI, Shaw H, Geyer K. Do smartphone usage scales predict behavior? Int J Hum Comput Stud 2019 Oct;130:86-92. [doi: 10.1016/j.ijhcs.2019.05.004]
26. Lee M, Han M, Pak J. Analysis of behavioral characteristics of smartphone addiction using data mining. Appl Sci (Basel) 2018 Jul 20;8(7):1191 [FREE Full text] [doi: 10.3390/app8071191]
27. Montag C, Baumeister H, Kannen C, Sariyska R, Meßner EM, Brand M. Concept, possibilities and pilot-testing of a new smartphone application for the social and life sciences to study human behavior including validation data from personality psychology. J (Basel) 2019 Mar 27;2(2):102-115 [FREE Full text] [doi: 10.3390/j2020008]
28. Sariyska R, Rathner EM, Baumeister H, Montag C. Feasibility of linking molecular genetic markers to real-world social network size tracked on smartphones. Front Neurosci 2018 Dec 18;12:945 [FREE Full text] [doi: 10.3389/fnins.2018.00945] [Medline: 30618574]
29. Messner EM, Sariyska R, Mayer B, Montag C, Kannen C, Schwerdtfeger A, et al. Insights – Future implications of passive smartphone sensing in the therapeutic context. Verhaltenstherapie 2019 Aug 9:1-10. [doi: 10.1159/000501951]
30. Insights. insightsapp. URL: https://www.insightsapp.org/ [accessed 2021-08-10]
31. Montag C, Yang H, Elhai JD. On the psychology of TikTok use: A first glimpse from empirical findings. Front Public Health 2021 Mar 16;9:641673 [FREE Full text] [doi: 10.3389/fpubh.2021.641673] [Medline: 33816425]
32. Kwon M, Lee JY, Won WY, Park JW, Min JA, Hahn C, et al. Development and validation of a smartphone addiction scale (SAS). PLoS One 2013;8(2):e56936. [doi: 10.1371/journal.pone.0056936] [Medline: 23468893]
33. Efron B, Tibshirani RJ. An Introduction to the Bootstrap. Boca Raton, Florida: Chapman & Hall; 1994.
34. Preacher KJ, Hayes AF. SPSS and SAS procedures for estimating indirect effects in simple mediation models. Behav Res Methods Instrum Comput 2004 Nov;36(4):717-731. [doi: 10.3758/bf03206553] [Medline: 15641418]
35. Wegmann E, Brand M. A narrative overview about psychosocial characteristics as risk factors of a problematic social networks use. Curr Addict Rep 2019 Oct 28;6:402-409. [doi: 10.1007/s40429-019-00286-8]
36. Holte AJ, Ferraro FR. True colors: Grayscale setting reduces screen time in college students. Soc Sci J 2020 Apr 03:1-17. [doi: 10.1080/03623319.2020.1737461]
37. Gesprächskreis Digitalität und Verantwortung: Austausch zu ethischen Fragestellungen der Digitalisierung. Facebook. 2018 Jan 24. URL: https://about.fb.com/de/news/h/gespraechskreis-digitalitaet-und-verantwortung/ [accessed 2021-08-24]

Abbreviations

IM: instant messaging

SAS: Smartphone Addiction Scale
SmUD: smartphone use disorder
SN: social networking

Edited by R Kukafka; submitted 11.01.21; peer-reviewed by A Gulliver, R Pryss, JR Bautista, E Said-Hung; comments to author 16.03.21; revised version received 01.06.21; accepted 05.07.21; published 30.09.21.

Please cite as:
Marengo D, Sariyska R, Schmitt HS, Messner EM, Baumeister H, Brand M, Kannen C, Montag C
Exploring the Associations Between Self-reported Tendencies Toward Smartphone Use Disorder and Objective Recordings of Smartphone, Instant Messaging, and Social Networking App Usage: Correlational Study
J Med Internet Res 2021;23(9):e27093
URL: https://www.jmir.org/2021/9/e27093
doi:10.2196/27093
PMID:34591025

Original Paper

Roles of Psychosocial Factors on the Association Between Online Social Networking Use Intensity and Depressive Symptoms Among Adolescents: Prospective Cohort Study

Ji-Bin Li[1,2*], PhD; Li-Fen Feng[3], MD; Anise M S Wu[4], PhD; Jin-Chen Mai[5], BSc; Yu-Xia Chen[5], BSc; Phoenix K H Mo[2*], PhD; Joseph T F Lau[2*], PhD

[1]Department of Clinical Research, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, China

[2]Center for Health Behaviours Research, The Jockey Club School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, China

[3]Department of Statistics, Government Affairs Service Center of Health Commission of Guangdong Province, Guangzhou, China

[4]Department of Psychology, Faculty of Social Sciences, University of Macau, Macau, China

[5]Department of Psychological Health Research, Center for Health Promotion of Primary and Secondary School of Guangzhou, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Joseph T F Lau, PhD
Center for Health Behaviours Research
The Jockey Club School of Public Health and Primary Care
The Chinese University of Hong Kong
5/F, School of Public Health
Prince of Wales Hospital
Hong Kong
China
Phone: 86 26376606
Email: jlau@cuhk.edu.hk

Abstract

Background: The potential mechanisms underlying the association between online social networking use intensity and depressive symptoms are unclear and underresearched.

Objective: We aimed to investigate the potential roles of interpersonal psychosocial factors on the association between online social networking use intensity and depressive symptoms among early adolescents.

Methods: A total of 4237 adolescents from a 9-month longitudinal study were included. Score changes (indicated as △) for the social function use intensity (SFUI) and entertainment function use intensity (EFUI) subscales of the Online Social Networking Activity Intensity Scale and for friendship quality, perceived family support, perceived friend support, parent–adolescent conflict, social nonconfidence, and depressive symptoms were analyzed. The potential mediation effects of unfavorable psychosocial factors and suppression effects of favorable psychosocial factors on the association of △SFUI with △CES-D and the association of △EFUI with △CES-D were tested using hierarchical regression models.

Results: The association between △SFUI and △CES-D was partially mediated by △mother–adolescent conflict (mediation effect size 5.11%, *P*=.02) and △social nonconfidence (mediation effect size 20.97%, *P*<.001) but partially suppressed by △friendship quality, △perceived family support, and △perceived friend support, with suppression effects of –0.011 (*P*=.003), –0.009 (*P*=.003), and –0.022 (*P*<.001), respectively. The association between △EFUI and △CES-D was partially mediated by △social nonconfidence (mediation effect size 30.65%, *P*<.001) but partially suppressed by △perceived family support and △perceived friend support, with suppression effects of –0.036 (*P*<.001) and –0.039 (*P*<.001), respectively.

Conclusions: The association between online social networking use intensity and depressive symptoms was partially mediated through the indirect increase in social nonconfidence and mother–adolescent conflict; however, better perceived social support and friendship quality would partially compensate for the harmful impact of online social networking use intensity on depressive symptoms among early adolescents.

(J Med Internet Res 2021;23(9):e21316) doi:10.2196/21316

KEYWORDS

online social networking use intensity; depressive symptoms; psychosocial factors; mediation and suppression; longitudinal study

Introduction

Adolescent depression is a global challenge that is significantly associated with serious physical and psychological problems (eg, substance use, eating disorder, and suicide) [1-4]. A meta-analysis [5] estimated a pooled prevalence of depression of 24.3% among adolescents in mainland China (range 6.2%-64.8%); this prevalence was 24.6% at baseline and 26.6% 9 months later in our previous study [6] (which used the same data set as this paper). Various types of problematic internet use (eg, internet addiction and internet gaming disorders) aggravate adolescent depression [7-10]. The intensity of online social networking use also is positively associated with depressive symptoms among adolescents in several cross-sectional studies [11-13], and while it is necessary to have an understanding of the mechanisms underlying the association to address prevention and intervention, such an understanding is lacking. As cross-sectional studies are unable to establish temporal relationships, they cannot be used to investigate potential underlying mechanisms of the relationships between online social networking use intensity and depressive symptoms. Longitudinal studies are therefore warranted. Given that adolescents use online social networking mainly for social purposes (eg, communicating, seeking social support, and maintaining or establishing friendships) [14-16], we investigated whether favorable (friendship quality and social support) and unfavorable (conflict with parents and lack of social confidence) interpersonal attributes would mediate or suppress the association between online social networking use intensity measures and depressive symptoms in early adolescence, by using a prospective cohort study design [17-19].

Online social networking use can potentially lead to unfavorable and stressful interpersonal relationships. According to the displacement theory, internet use reduces adolescents' communication with family members and social involvement, which then increases psychiatric symptoms [20]. Consistently, heavy online social networking use has been shown to reduce adolescents' intimacy and time spent with their parents and families and to increase conflicts with parents [20,21]. For instance, college students who used online social networking to communicate with their parents reported higher levels of conflicts with their parents [22]. Parental control over adolescents' online social networking use may also cause conflicts with adolescents [23]. Furthermore, heavy online social networking use was associated with fewer face-to-face interactions and social confidence among adolescents [24], possibly due to habitual deprivation of nonverbal cues (ie, voice tone, eye contact) and reduction in social presence, which may reduce self-efficacy in handling social interactions (social nonconfidence) [25]. As parent–adolescent conflict [26,27] and social nonconfidence [28] have also been positively associated with depressive symptoms, we contended that these 2 unfavorable interpersonal relationship factors would mediate the association between online social networking use intensity and depressive symptoms among adolescents.

Nevertheless, online social networking use may enhance friendship quality among adolescents; a study [23] showed that 43% of high school students perceived that online social networking use increased their closeness with friends; and another study [21] reported that 49% of the college students' closest friends whom they met offline were also their closest friends in their online social networks. Online social networking use facilitates individuals' social connectedness and self-disclosure [29,30]; it is a convenient and essential source of social support [14]. For instance, Facebook users reported better perceived social support than nonusers [31], and the length of time spent on Facebook and the number of Facebook friends were positively associated with perceived social support [32,33]; however, nonsignificant associations have also been reported [34,35]. The frequency of online social networking use has also been associated with higher perceived social support and enhanced friendship quality [31,36]. As friendship quality [37,38] and perceived social support [39,40] are protective against depressive symptoms, we contended that these 2 favorable interpersonal relationship factors would potentially suppress the positive association between online social networking use intensity and depressive symptoms among adolescents.

It has been reported that the patterns of social relationships and depressive symptoms are likely to differ between the younger and older adolescents [41]. As a key stage of early adolescence, the rapid physical (eg, early puberty), psychological (eg, concrete thinking but early moral concepts) and social (eg, emotional separation from parents, start of strong peer identification; early exploratory behaviors) developmental changes produce specific disease patterns, unusual presentations of symptoms, unique communication and management challenges, and they gradually establish identity and autonomy in the context of the social cultural environment [42,43]. Moreover, it has been observed that the risk of depression is low in childhood but increasing substantially with adolescence [44] and that meaningful risk begins in the early teens [45]. Therefore, from the perspective of early prevention, we chose young adolescents as the target population.

In this 9-month prospective cohort study, we aimed to test the significance of 3 potential mediators (ie, father–adolescent conflict, mother–adolescent conflict, and social nonconfidence) and 3 potential suppressors (ie, friendship quality, perceived family support, and perceived friend support) for the association between online social networking use intensity measures and depressive symptoms in early adolescence. Specifically, it was hypothesized that the associations between changes in 2 online social networking use intensity measures—social function use intensity (SFUI) and entertainment function use intensity (EFUI) and depressive symptoms would be partially mediated or suppressed by the aforementioned unfavorable or favorable

interpersonal psychosocial factors. To the best of our knowledge, no studies have tested such mediation and suppression effects.

Methods

Participants and Procedures

This 9-month prospective cohort study was conducted in Guangzhou, South China. The methods have previously been described [6]. Briefly, 9 public secondary schools, 3 from each of the city's 3 regions (ie, core, suburb, and outer suburb regions), were selected by convenience sampling, and all students of the seventh and eighth grade (ie, 7 and 8 years of formal education) were invited to participate in the study. Grade 9 students were not included into the study given that they were too busy preparing for the public examination for entrance to senior high schools and would leave their schools before the end of the 9-month follow-up period; it would thus be practically difficult to follow up such students. The anonymous structured questionnaire was self-administered in classroom settings without the presence of teachers. We announced that the return of the completed questionnaire implied consent to participate in the study; this information was also printed on the cover page of questionnaire. No incentive was given to the participants. Permission to conduct the survey was granted by the school principals. The Survey and Behavioral Research Ethics Committee of the Chinese University of Hong Kong approved the study.

A total of 5365 students (response rate 98.0%) completed the baseline questionnaire. The baseline and follow-up questionnaires completed by each student were matched based on the last 4 digits of the student's home telephone number, the last 4 digits of the parents' mobile phone number, the last 4 digits of the student's identity card number, the student's birthday, and the last letter of the student's and parents' name. A total of 4871 (follow-up rate 90.8%) of the 5365 students provided matched completed questionnaires at the follow-up survey. Students who were lost to follow-up were more likely to be females, be in their junior year of school, to not live with both parents, to have a perceived poor/very poor family financial situation, and to have lower self-reported academic performance (Multimedia Appendix 1). Students who did not use online social networking (n=643) were excluded from the data analysis. The effective sample size was 4237 in this study.

Measurements

Background Variables

Participant sociodemographic information (ie, gender, grade, parental education levels, perceived family financial situation, and living arrangement with parents) and information about self-reported academic performance and perceived academic pressure were collected in the baseline survey.

Depression

Depressive symptoms were assessed using the Chinese version of the 20-item Center for Epidemiological Studies–Depression Scale (CES-D), which is an epidemiological screening tool, not a clinical diagnosis tool. All items were rated on a 4-point Likert scale from 0, representing none/rarely (less than 1 day), to 3, representing almost/all of the time (5-7 days), during the past week, with higher total scores (range 0 to 60) representing more depressive symptoms. Psychometric properties of the Chinese version of the CES-D have been demonstrated among Chinese adolescents [46]. The CES-D is commonly used in mental health studies [47], and depression defined by such scale is strongly associated with clinical diagnosis of depression [48]. In this study, the Cronbach α was 0.86 at baseline and 0.87 at follow-up, indicating good internal reliability in this study.

Online Social Networking Use Intensity

The intensity of online social networking use was measured using the Online Social Networking Activity Intensity Scale on a 5-point response scale ranging from 0 (never) to 4 (always). The scale was developed and validated among Chinese adolescents [49]. It measures SFUI (10 items; range: 0-40) and EFUI (4 items; range: 0-16). A higher score indicates a higher intensity of online social networking use. In this study, the Cronbach α for SFUI and EFUI were 0.88 and 0.60 at baseline and 0.89 and 0.62 at follow-up, which showed good internal reliability.

Friendship Quality

Friendship quality was measured using the 6-item peer-relationship subscale of the Children and Adolescent Quality of Life Scale, which has been widely used among Chinese adolescents [50,51]. Each item has 4 responses, and a higher score (range 6 to 24) indicates better friendship quality. In this study, the Cronbach α was 0.82 at baseline and 0.83 at follow-up, indicating good internal reliability.

Perceived Social Support

Perceived social support was measured using the Chinese version of the 12-item Multidimensional Scale of Perceived Social Support [52], including the dimensions of perceived family support (4 items) and perceived friends support (8 items). Items were rated on a 7-point Likert scale, from 1 (very strongly disagree) to 7 (very strongly agree). Higher scores indicated better perceived social support. In this study, Cronbach α values for both subscales were >.90 at both baseline and follow-up, indicating good internal reliability.

Parent–Adolescent Conflicts

The Chinese parent–adolescent conflict questionnaire includes 7 dimensions (ie, study, chores, expenses, living arrangement, appearance, relationship with other family members, and privacy) [53,54]. Frequency of conflict, from 1 (never) to 5 (almost every day), in the previous month, with the father and mother were rated separately. Higher scores (range 7-35) indicated more conflict with parents. In this study, Cronbach α values were 0.84 for father–adolescent conflict and 0.85 for mother–adolescent conflict at baseline and 0.86 and 0.87 at follow-up, indicating good internal reliability.

Social Nonconfidence

Social nonconfidence was measured using the 4-item social nonconfidence subscale of the Chinese version of the Social Skills Scale, which showed good psychometric properties among college students [28]. Social nonconfidence refers to a strong concern about what others think about oneself and the

experience of nervous feelings during social occasions [55]. Items were ranked on a 5-point Likert scale, from 1 (definitely not like me) to 5 (exactly like me), with higher scores (range 4-20) indicating stronger social nonconfidence. In this study, the Cronbach α was 0.60 at baseline and 0.64 at follow-up, indicating acceptable internal reliability.

Statistical Analysis

The associations between changes (indicated as △) in online social networking use intensity (SFUI and EFUI) and depressive symptoms were tested. Potential mediators of the associations included changes in father–adolescent conflict, mother–adolescent conflict, and social nonconfidence; potential suppressors included changes in friendship quality, perceived family support, and perceived friends support. Changes were determined by subtracting the scale scores at baseline from those at the 9-month follow-up.

Potential mediation or suppression effects were tested using hierarchical modeling based on the methods proposed by Baron and Kenny [56], which required the presence of significant associations between (1) △SFUI (or △EFUI) and △CES-D; (2) △SFUI (or △EFUI) and changes in potential mediators or suppressors; and (3) changes in potential mediators or suppressors and △CES-D after controlling for △SFUI (or △EFUI). The strength of the indirect effects (mediation or suppression) was estimated by the product $a \times b$, where a is the regression coefficient that relates the independent variable to the potential mediator (or suppressors) and b is the regression coefficient that relates the mediators (or suppressors) to the dependent variable, with significance determined using Sobel Z [57,58]; $a \times b / (a \times b + c')$ reflects the magnitude of mediation effect size [57,59], where c' is the regression coefficients relating the mediators (or suppressors) to the dependent variable and the independent variable to the dependent variable in the same model. As the total effect of the suppressors could not be decomposed into indirect and direct proportions because it comprised associations in opposite directions, the proportion of the indirect effect could not be calculated; thus, only the suppressing effects ($a \times b$) were reported. The interaction effects of gender for the associations between △SFUI (or △EFUI), changes in potential mediators (or suppressors) and △CES-D were further tested by including the interaction terms (gender multiplied by △SFUI, △EFUI, changes in potential mediators, or changes in potential suppressors).

Multilevel linear regression models (level 1: students, level 2: schools) with random intercepts were fitted to take into account the potential clustering effects from the schools. All models were adjusted for baseline background factors that were significantly associated with △CES-D in the univariate models (P<.05), and the unstandardized regression coefficients with standard errors are presented. Analyses were conducted using SAS software (version 9.4, SAS Institute). A 2-sided P<.05 was considered statistically significant.

Results

Sample Characteristics

In the baseline sample, 49.7% (2105/4237) of the participants were male, and 47.5% (2011/4237) were in the seventh grade. Approximately 48.3% (2047/4237) of students self-reported a good or very good family financial situation. The majority of students (88.4%, 3747/4237) lived with both parents. Approximately one-fifth (20.1%, 852/4237) reported lower academic performance, and 23.4% (993/4237) perceived heavy or very heavy academic pressure (Table 1).

Table 1. Background characteristics and their associations with change in Center for Epidemiological Studies–Depression Scale score by Univariate linear regression (n=4237).

Variables	n (%)	β (SE)	*P* value
Sociodemographic variables			
Gender			
Male	2105 (49.7)	—[a]	
Female	2132 (50.3)	–0.12 (0.26)	.65
Grade			
Seven	2011 (47.5)	—	
Eight	2226 (52.5)	0.54 (0.26)	.04
Father's education level			
Primary school or below	273 (6.4)	—	
Junior middle school	1425 (33.6)	0.53 (0.57)	.35
High middle school	1312 (31.0)	0.59 (0.58)	.31
University or above	1053 (24.9)	0.47 (0.60)	.43
Don't know	174 (4.1)	0.27 (0.84)	.74
Mother's education level			
Primary school or below	445 (10.5)	—	
Junior middle school	1507 (35.6)	1.24 (0.47)	.008
High middle school	1199 (28.3)	1.16 (0.49)	.02
University or above	913 (21.5)	0.65 (0.51)	.20
Don't know	123 (4.1)	0.51 (0.77)	.51
Family financial situation			
Very good/good	2047 (48.3)	—	
Medium	2072 (48.9)	0.02 (0.27)	.93
Poor/very poor	118 (2.8)	0.27 (0.81)	.73
Living with both parents			
Yes	3747 (88.4)	—	
No	490 (11.6)	–0.71 (0.41)	.09
School-related variables			
Academic performance			
Upper	1465 (34.6)	—	
Medium	1920 (45.3)	–0.68 (0.30)	.02
Lower	852 (20.1)	–1.24 (0.37)	<.001
Perceived academic pressure			
Nil/light	811 (19.1)	—	
Average	2433 (57.5)	–0.36 (0.35)	.30
Heavy/very heavy	993 (23.4)	–1.46 (0.40)	<.001

[a]Reference group.

During a 9-month follow-up period, the mean △CES-D was 0.48 (95% CI 0.22, 0.74; *P*<.001). There were slight significant reductions in the SFUI (*P*=.01) and EFUI scores (*P*<.001) and the scores of 3 potential suppressors (friendship quality: *P*=.001; perceived family support: *P*=.009; and perceived friend support: *P*<.001) and 2 potential mediators (father–adolescent conflict: *P*=.02; social nonconfidence: *P*<.001), whereas △mother–adolescent conflict from baseline to follow-up was not statistically significant (*P*=.21) (Table 2).

Table 2. Changes in Center for Epidemiological Studies–Depression, Social Function Use Intensity, Entertainment Function Use Intensity, and psychosocial factors scores from baseline to follow-up.

Variables	Baseline	Follow-up	Mean difference (95% CI)	*P* value for time
Center for Epidemiological Studies–Depression score	15.1 (9.5)	15.6 (9.5)	0.48 (0.22, 0.74)	<.001
Online social networking use intensity				
Social Function Use Intensity	18.4 (8.3)	18.1 (8.4)	–0.31 (–0.55, –0.08)	.01
Entertainment Function Use Intensity	8.2 (3.1)	7.8 (3.2)	–0.46 (–0.56, –0.36)	<.001
Psychosocial factors				
Friendship quality	19.0 (3.4)	18.8 (3.4)	–0.17 (–0.27, –0.08)	.001
Perceived family support	20.4 (5.6)	20.2 (5.4)	–0.21 (–0.37, –0.05)	.009
Perceived friend support	42.0 (10.1)	40.7 (10.5)	–1.31 (–1.63, –1.00)	<.001
Father–adolescent conflict	12.9 (5.8)	12.7 (5.8)	–0.25 (–0.45, –0.05)	.02
Mother–adolescent conflict	14.5 (6.3)	14.6 (6.7)	0.15 (–0.08, 0.37)	.21
Social nonconfidence	12.8 (3.1)	12.6 (3.1)	–0.22 (–0.33, –0.12)	<.001

Associations Between △SFUI (or △EFUI) and △CES-D (or Changes in Psychosocial Factors)

After adjustment for grade, academic performance, and perceived academic pressure, we found that (1) both △SFUI and △EFUI were positively associated with △CES-D (Table 3); (2) all 3 favorable interpersonal variables (potential suppressors) were negatively associated with △CES-D, while all 3 unfavorable interpersonal variables (potential mediators) were positively associated with △CES-D (Table 4); (3) △SFUI was positively associated with all interpersonal variables, except △father–adolescent conflict (*P*=.12); and (4) △EFUI was positively associated with 2 favorable interpersonal variables (△perceived family support and △perceived friend support) and one unfavorable interpersonal variables (△social nonconfidence), but not with △friendship quality, △father–adolescent conflict, and △mother–adolescent conflict (Table 3).

Interaction effects of gender on (1) the association between △SFUI (or △EFUI) and △CES-D (*P*=.60; *P*=.41), (2) the association between △SFUI (or △EFUI) and change in psychosocial factors (*P*=.09-.94; *P*=.18-.93), and (3) the associations between psychosocial factors and △CES-D (*P* range: .30–.68) were nonsignificant, except for the interaction between gender and △SFUI on △father–adolescent conflict (interaction *P*=.006), between gender and △friendship quality on △CES-D (*P*=.03), and between gender and △perceived family support on △CES-D (interaction *P*=.04) (Multimedia Appendix 2 and Multimedia Appendix 3).

Table 3. Associations between △SFUI (or △EFUI) and △CES-D (or psychosocial factors) from multilevel linear regression models (n=4237).

Dependent variable	Independent variable							
	△Social Function Use Intensity				△Entertainment Function Use Intensity			
	Univariate β (SE)	*P* value	Adjusted[a] β (SE)	*P* value	Univariate β (SE)	*P* value	Adjusted[a] β (SE)	*P* value
△CES-D[b]	0.08 (0.02)	<.001	0.07 (0.02)	<.001	0.14 (0.04)	<.001	0.14 (0.04)	<.001
Psychosocial factors								
△Friendship quality	0.02 (0.01)	.006	0.02 (0.01)	.004	0.02 (0.01)	.22	0.02 (0.02)	.06
△Perceived family support	0.03 (0.01)	.003	0.03 (0.01)	.002	0.15 (0.02)	<.001	0.15 (0.02)	<.001
△Perceived friend support	0.18 (0.02)	<.001	0.18 (0.02)	<.001	0.32 (0.05)	<.001	0.32 (0.05)	<.001
△Father–adolescent conflict	0.02 (0.01)	.08	0.02 (0.01)	.12	0.02 (0.03)	.56	0.01 (0.03)	.71
△Mother–adolescent conflict	0.04 (0.02)	.004	0.04 (0.02)	.006	0.06 (0.03)	.06	0.06 (0.03)	.10
△Social nonconfidence	0.04 (0.01)	<.001	0.04 (0.01)	<.001	0.12 (0.02)	<.001	0.12 (0.02)	<.001

[a]Models were adjusted by grade, academic performance, and perceived study pressure.

[b]CES-D: Center for Epidemiological Studies–Depression scale.

Table 4. Associations between changes in psychosocial factors and △Center for Epidemiological Studies–Depression Scale (n=4237).

Psychosocial factors	Univariate		Adjusted[a]	
	β (SE)	*P* value	β (SE)	*P* value
△Friendship quality	–0.62 (0.04)	<.001	–0.61 (0.04)	<.001
△Perceived family support	–0.29 (0.02)	<.001	–0.29 (0.02)	<.001
△Perceived friend support	–0.11 (0.01)	<.001	–0.12 (0.01)	<.001
△Father–adolescent conflict	0.10 (0.02)	<.001	0.09 (0.02)	<.001
△Mother–adolescent conflict	0.10 (0.02)	<.001	0.09 (0.02)	<.001
△Social nonconfidence	0.38 (0.04)	<.001	0.37 (0.04)	<.001

[a]Models were adjusted by grade, academic performance, and perceived study pressure.

Mediation and Suppression Effects Between △SFUI and △CES-D

When a potential mediator or suppressor was entered into the model that contained only △SFUI as the independent variable and had △CES-D as the dependent variable, a significant mediator resulted in a significant decrease in the regression coefficient, while a significant suppressor resulted in a significant increase in the regression coefficient between △SFUI and △CES-D. Accordingly, we found 2 significant mediators of the association between △SFUI and △CES-D: △mother–adolescent conflict (mediation effect size 5.11%, *P*=.02; model m1, Table 5) and △social nonconfidence (mediation effect size 20.97%, *P*<.001; model m2, Table 5). The combined mediation effect was approximately 25.33% (model m3).

In addition, we found that all 3 favorable interpersonal relationship variables were significant suppressors: △friendship quality (suppression effect: –0.011, *P*=.003; model s1, Table 6), △perceived family support (suppression effect: –0.009, *P*=.003; model s2), and △perceived friend support (suppression effect: –0.022, *P*<.001; model s3).

Table 5. Mediation effects of changes in negative psychosocial factors on the association between △SFUI (or △EFUI) and △Center for Epidemiological Studies–Depression Scale (n=4237) using multilevel linear regression models. All models were adjusted by grade, academic performance and perceived study pressure.

Model	β (SE)	*P* value	Mediation effect (the ratio of indirect effect to total effect)		
			Effect size (%)	Sobel *Z*	*P* value
Model m1			5.11	2.43	.02
△SFUI[a]	0.07 (0.02)	<.001			
△Mother–adolescent conflict	0.09 (0.02)	<.001			
Model m2			20.97	5.14	<.001
△SFUI	0.06 (0.02)	<.001			
△Social nonconfidence	0.36 (0.04)	<.001			
Model m3			25.33	N/A[b]	N/A
△SFUI	0.06 (0.02)	<.001			
△Mother–adolescent conflict	0.09 (0.02)	<.001			
△Social nonconfidence	0.35 (0.04)	<.001			
Model m4			30.65	6.07	<.001
△EFUI[c]	0.10 (0.04)	<.001			
△Social nonconfidence	0.36 (0.04)	<.001			

[a]SFUI: Social Function Use Intensity.

[b]N/A: not applicable.

[c]EFUI: Entertainment Function Use Intensity.

Table 6. Suppression effects of changes in positive psychosocial factors on the association between △SFUI (or △EFUI) and △Center for Epidemiological Studies–Depression Scale (n=4237) using multilevel linear regression models. All models were adjusted by grade, academic performance and perceived study pressure.

Model	β (SE)	*P* value	Suppression effect		
			Effect	Sobel *Z*	*P* value
Model s1			–0.011	–2.95	.003
△SFUI[a]	0.08 (0.02)	<.001			
△Friendship quality	–0.62 (0.04)	<.001			
Model s2			–0.009	–3.01	.003
△SFUI	0.08 (0.02)	<.001			
△Perceived family support	–0.30 (0.02)	<.001			
Model s3			–0.022	–6.47	<.001
△SFUI	0.10 (0.02)	<.001			
△Perceived friend support	–0.12 (0.01)	<.001			
Model s4			–0.036	–5.57	<.001
△EFUI[b]	0.18 (0.04)	<.001			
△Perceived family support	–0.30 (0.02)	<.001			
Model s5			–0.039	–5.52	<.001
△EFUI	0.18 (0.04)	<.001			
△Perceived friend support	–0.12 (0.01)	<.001			

[a]SFUI: Social Function Use Intensity.

[b]EFUI: Entertainment Function Use Intensity.

Mediation and Suppression Effects Between △EFUI and △CES-D

△Social nonconfidence (mediation effect size: 30.65%, *P*<.001; model m4) was the only significant mediator for the association between △EFUI and △CES-D. Significant suppressors included △perceived family support (suppression effect –0.036, *P*<.001; model s4) and △perceived friend support (suppression effect –0.039, *P*<.001; model s5).

Discussion

Principal Findings

In this study, we found a significant increase in depressive symptoms (*P*<.001) among young adolescents during the 9-month period and positive associations between △SFUI (or △EFUI) and △CES-D. Moreover, associations were simultaneously partially mediated by some unfavorable psychosocial factors (ie, social nonconfidence and mother–adolescent conflict) and partially suppressed by some favorable psychosocial factors (ie, perceived social support and quality of friendship).

Consistent with our findings, previous studies conducted in the United States [60] and Iceland [61] had also reported a gradual increase in depressive symptoms among adolescents. The findings suggest that adolescents are vulnerable to gradual deterioration of their mental health and even chronic depression. Attention is required, given that depression may have lasting negative consequences for adolescents.

We found that, with increases in depressive symptoms, adolescents may have less favorable relationships (ie, reduced friendship quality and perceived social support) over time. Given that adolescents need to establish an autonomous self-identity, they may spend less time with their families, resulting in reduced perceived family support [62]. Our findings indicate that adolescents seem to face an additional challenge of having their relationships with their peers deteriorate over time, possibly due to the low quality of friendships formed online, the difference between online and general offline social connectedness [63], and important developmental changes during adolescence (eg, decrease in parental attachment, learning to form attachment relationships with peers, exploration of new environments) [64]. Fortunately, adolescents may gain social confidence over time, possibly because they have more practice in social interactions with other people. We observed that changes in interpersonal relationships were associated with changes in depressive symptoms. As social relationships are important determinants of mental, physical, and behavioral health among adolescents, longitudinal studies are warranted to understand other consequences of changes in such interpersonal relationships. In addition, we found a significant and slight decline in online social networking use intensity (both SFUI with *P*=.01 and EFUI with *P*<.001) over the 9-month period. To the best of our knowledge, our study is the first to document a longitudinal trend of online social networking use intensity among adolescents. We show that such changes were associated with changes in interpersonal relationships and depressive symptoms.

This study identified a number of important partial mediators and suppressors of the associations between △SFUI (or △EFUI) and △CES-D. First, △social nonconfidence partially mediated the associations of both △SFUI and △EFUI with △CES-D. This is a novel finding. An increase in online social networking use may thus increase, rather than reduce, social nonconfidence, which in turn increases depressive symptoms. While online social networking sites provide convenient platforms for establishing communication, contacts, and relationships among adolescents [65], they might simultaneously deprive adolescents of skills, time allocation, and experiences related to effective direct offline social interactions, resulting in social nonconfidence. Online social networking use does not involve the use of social and nonverbal cues; it may also weaken individuals' desire for involvement in offline social activities and pleasure seeking [66]. Furthermore, social nonconfidence is potentially associated with isolation and other risk factors for depression, which might mediate the association between △SFUI (or △EFUI) and depression. Future research should distinguish between social nonconfidence in face-to-face versus online social interactions and reveal the relationship of social nonconfidence with online and offline social relationships.

Another important finding was that △father–adolescent conflict mediated the association of neither △SFUI nor △EFUI with △CES-D, while △mother–adolescent conflict partially mediated the association between △SFUI and △CES-D (a relatively small effect of 5.11%) but not between △EFUI and △CES-D. Thus, overall, conflict with parents is not a major mechanism explaining the associations between △SFUI (or △EFUI) and depressive symptoms. Furthermore, mother–adolescent conflict, but not father–adolescent conflict, played a role as a mediator. The difference between the 2 types of parental conflict was evident in the correlation results, which showed that △SFUI was positively associated with △mother–adolescent conflict, but not with △father–adolescent conflict, while △EFUI was not associated with either △father–adolescent conflict or △mother–adolescent conflict. We speculate that mothers tend to be more engaged in parenting than fathers; they also interact more frequently with adolescents concerning personal issues (eg, intensive online social networking use and poor time management) [67,68]. Such parenting may increase adolescent–mother conflict [69]. Additionally, heavy online social networking use may negatively impact adolescents' academic performance [70,71] and result in adolescent conflicts with their mothers rather than with their fathers, given that mothers may be more involved and concerned about their children's personal development (eg, academic performance) than fathers [72,73]. Thus, an increase in SFUI might result in adolescents' more conflict with their mothers than with fathers.

△Friendship quality showed a suppression effect between △SFUI and △CES-D, but not between △EFUI and △CES-D, which is understandable, as SFUI reflects the social functions of online social networking use, while EFUI reflects the entertainment functions. Positive friendship is a protective factor against depressive symptoms among adolescents [74]. Online social networking is doubtlessly one of the most preferred platforms for communication among adolescents, through which they can be socially connected with peers regardless of time and place [75]. For instance, online social networking use can supplement offline interactions and communications with friends to improve friendship quality among adolescents. Mutual and interactive updating of information (eg, messages and photos) online increases perceived closeness with friends [76] and hence supports the maintenance of friendships, especially between friends that are geographically separated.

Equally important, both △perceived family support and △perceived friend support suppressed the association between changes in online social networking use intensity (both △SFUI and △EFUI) and △CES-D. Increases in SFUI and EFUI may enhance perceived support from family and friends, which in turn decreases the level of depressive symptoms. In this study, △SFUI and △EFUI were positively associated with perceived social support. Conceptually, online social networking use allows adolescents to express their needs for support readily by means of online self-disclosure (eg, communicating personal information, thoughts, and feelings with others online) [77,78], which has been shown to be correlated with perceived social support [29,79]. Simultaneously, online social networking use provides channels for others to provide social support to adolescents and for adolescents to receive such social support. Interestingly, △perceived family support, similar to △perceived friend support, was a suppressor. It is plausible that online social networking use is not limited to interaction with friends but also includes interaction with family members. The relationship of online social networking use with family support and its mechanisms require future research attention.

Importantly, △SFUI (or △EFUI) possessed both mediators (social nonconfidence and mother–adolescent conflict) that were risk factors for △CES-D and suppressors (friendship quality and perceived social support) that were protective factors against △CES-D. Stakeholders should be made aware that online social networking use by adolescents is neither good nor bad; related interventions should maximize protective suppressors and minimize harmful mediators of online social networking use among adolescents to transform harms into benefits. Attention should be paid to reducing harm that arises from social nonconfidence as a mediator, which was the strongest among all the mediators that were studied. Such interventions should provide training on social skills and communication skills. Conflict resolution strategies for adolescents and their mothers (eg, conciliation and positive problem-copying strategies) might also be useful [80]. In addition, given the high popularity of online social networking among adolescents, web-based platforms might be an innovative avenue to identify individuals at high risk for depression and can be a cost-effective tool to improve psychological well-being among adolescents [81].

Limitations

This study has some limitations. Reporting bias (eg, social desirable bias and recall bias) might exist although most studies [21,22] used self-reported data on online social networking use. The survey may not be nationally representative, due to large geographic variations and its restriction to secondary school students. Our study focused on early adolescence (ie, secondary school students), caution needs to be used in generalizing the

results to older adolescents. Future comparisons of this study's findings with those of older adolescents are warranted. There were differences in some sociodemographic factors (ie, gender, grade, family financial situation, lives with both parents, and academic performance) between those who were followed up and lost to follow-up in our study. Although the influence of some of these factors (ie, grade, academic performance) have been considered in the multivariable analysis, interpretation of the findings should be made cautiously by considering the potential influence of these differences. Online and offline psychosocial relationships were not differentiated, whereas a study conducted in Hong Kong reported that perceived online social support failed to buffer stress for adolescents [82]. We investigated only perceived social support but no other forms of social support. The psychosocial factors represent self-perceptions, which were measured by widely used and validated scales. The measurements of parent–adolescent conflict could not identify whether the parent–adolescent conflict arose from online social networking use; further refinements in future studies were warranted. The relatively short follow-up time and involvement of only one follow-up (in the same school year) is a limitation; the interpretation of the findings should hence take the design into account (ie, that findings reflect short-term but not long term changes). We also did not include other types of interpersonal and personal mediators (or suppressors). The causal mediational sequence could not be derived from the 2-level study, and the follow-up period was relatively short. Moreover, the mediation and suppression effects found in this study were relatively small. Future multilevel research considering more comprehensive factors is warranted. Despite these limitations, our study has the strengths of being novel and using a longitudinal study design and well-validated instruments.

Conclusion

Consistent with previous findings, an increase in online social networking use was associated with an increase in adolescents' depressive symptoms. The risk effect was, however, partially a balance between the combination of mediation and suppression effects of psychosocial factors. The increase in online social networking use was associated with both unfavorable interpersonal situations (eg, increase in social nonconfidence) that increased depressive symptoms (mediators) and improvement in favorable interpersonal relationships (eg, perceived social support, friendship quality) that reduced depressive symptoms (suppressors). Future studies should clarify the relative contributions of the various interpersonal effects, and the findings of this study should be considered when designing interventions. Furthermore, we found that an increase in online social networking use also improved perceived family support. However, the mechanism, as well as the degree and role of the involvement of parents in adolescents' online social networking use, have not been well studied in the literature; this involvement may be a potential moderator of the relationship between online social networking use and depressive symptoms, but the direction is unclear. Future studies are warranted. Finally, our findings provide a reminder for various stakeholders to eliminate the stigma against adolescents who are heavy online social networking users, given that online social networking use can generate both harmful and protective effects against depressive symptoms. The understanding or perception that online social networking use is associated with depressive symptoms needs to be reviewed and elaborated.

Acknowledgments

The authors would like to express their appreciation to all participants, their families, and their schools as well as all field workers for their support. We would also like to thank Dr. Xue-fen Su for her assistance in the coordination of the data collection. This study was partially supported by the National Natural Science Foundation of China (81373021) and was also supported by the Jockey Club School of Public Health and Primary Care Research Postgraduate Students' Research Grants and Chinese University of Hong Kong Research Postgraduate Student Grants for Overseas Academic Activities. The funding sources had no role in the design or conduct of the study collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Authors' Contributions

JBL, PKHM, and JTFL conceived and designed the study. JBL, JCM, and YXC acquired the data. JBL and LFF performed the statistical analyses. JBL, LFF, AMSW, PKHM, and JTFL drafted and revised the manuscript. All authors contributed to the interpretation of the results and the critical revision of the manuscript for important intellectual content and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Attrition analysis for participants lost to follow-up.
[DOCX File , 16 KB - jmir_v23i9e21316_app1.docx]

Multimedia Appendix 2

Interaction effects of gender for the associations between △SFUI (or △EFUI) and △CES-D (or changes in psychosocial factors).
[DOCX File , 15 KB - jmir_v23i9e21316_app2.docx]

Multimedia Appendix 3
Interaction effects of gender for the associations between changes in psychosocial factors and △CES-D.
[DOCX File , 14 KB - jmir_v23i9e21316_app3.docx]

References

1. Birmaher B. Course and outcome of child and adolescent major depressive disorder. Child Adolesc Psychiatr Clin N Am 2002 Jul;11(3):619-637. [doi: 10.1016/s1056-4993(02)00011-1] [Medline: 12222086]
2. Howard AL, Kennedy TM, Macdonald EP, Mitchell JT, Sibley MH, Roy A, et al. Depression and ADHD-related risk for substance use in adolescence and early adulthood: concurrent and prospective associations in the MTA. J Abnorm Child Psychol 2019 Dec;47(12):1903-1916 [FREE Full text] [doi: 10.1007/s10802-019-00573-y] [Medline: 31273568]
3. Brown M, Hochman A, Micali N. Emotional instability as a trait risk factor for eating disorder behaviors in adolescents: sex differences in a large-scale prospective study. Psychol Med 2020 Aug;50(11):1783-1794. [doi: 10.1017/S0033291719001818] [Medline: 31379310]
4. Wild LG, Flisher AJ, Lombard C. Suicidal ideation and attempts in adolescents: associations with depression and six domains of self-esteem. J Adolesc 2004 Dec;27(6):611-624. [doi: 10.1016/j.adolescence.2004.03.001] [Medline: 15561306]
5. Tang X, Tang S, Ren Z, Wong DFK. Prevalence of depressive symptoms among adolescents in secondary school in mainland China: a systematic review and meta-analysis. J Affect Disord 2019 Feb 15;245:498-507. [doi: 10.1016/j.jad.2018.11.043] [Medline: 30439677]
6. Li J, Mo PKH, Lau JTF, Su X, Zhang X, Wu AMS, et al. Online social networking addiction and depression: the results from a large-scale prospective cohort study in Chinese adolescents. J Behav Addict 2018 Sep 01;7(3):686-696 [FREE Full text] [doi: 10.1556/2006.7.2018.69] [Medline: 30203664]
7. Lau JTF, Walden DL, Wu AMS, Cheng K, Lau MCM, Mo PKH. Bidirectional predictions between Internet addiction and probable depression among Chinese adolescents. J Behav Addict 2018 Sep 01;7(3):633-643 [FREE Full text] [doi: 10.1556/2006.7.2018.87] [Medline: 30264608]
8. Moreno MA, Jelenchick LA, Breland DJ. Exploring depression and problematic internet use among college females: a multisite study. Comput Human Behav 2015 Aug;49:601-607. [doi: 10.1016/j.chb.2015.03.033]
9. González-Bueso V, Santamaría JJ, Fernández D, Merino L, Montero E, Ribas J. Association between internet gaming disorder or pathological video-game use and comorbid psychopathology: a comprehensive review. Int J Environ Res Public Health 2018 Apr 03;15(4):668 [FREE Full text] [doi: 10.3390/ijerph15040668] [Medline: 29614059]
10. Guo W, Tao Y, Li X, Lin X, Meng Y, Yang X, et al. Associations of internet addiction severity with psychopathology, serious mental illness, and suicidality: large-sample cross-sectional study. J Med Internet Res 2020 Aug 11;22(8):e17560 [FREE Full text] [doi: 10.2196/17560] [Medline: 32780029]
11. Frison E, Eggermont S. Browsing, posting, and liking on Instagram: the reciprocal relationships between different types of instagram use and adolescents' depressed mood. Cyberpsychol Behav Soc Netw 2017 Oct;20(10):603-609. [doi: 10.1089/cyber.2017.0156] [Medline: 29039700]
12. Pantic I, Damjanovic A, Todorovic J, Topalovic D, Bojovic-Jovic D, Ristic S, et al. Association between online social networking and depression in high school students: behavioral physiology viewpoint. Psychiatr Danub 2012 Mar;24(1):90-93 [FREE Full text] [Medline: 22447092]
13. Labrague L. Facebook use and adolescents' emotional states of depression, anxiety, and stress. Health Sci J 2014;8(1):80-89 [FREE Full text]
14. Kim Y, Sohn D, Choi SM. Cultural difference in motivations for using social network sites: a comparative study of American and Korean college students. Comput Human Behav 2011 Jan;27(1):365-372. [doi: 10.1016/j.chb.2010.08.015]
15. Tosun LP. Motives for Facebook use and expressing "true self" on the internet. Comput Human Behav 2012 Jul;28(4):1510-1517. [doi: 10.1016/j.chb.2012.03.018]
16. Pempek TA, Yermolayeva YA, Calvert SL. College students' social networking experiences on Facebook. J Appl Dev Psychol 2009 May;30(3):227-238. [doi: 10.1016/j.appdev.2008.12.010]
17. Tabernero C, Gutiérrez-Domingo T, Vecchione M, Cuadrado E, Castillo-Mayén R, Rubio S, et al. A longitudinal study on perceived health in cardiovascular patients: the role of conscientiousness, subjective wellbeing and cardiac self-efficacy. PLoS One 2019;14(10):e0223862 [FREE Full text] [doi: 10.1371/journal.pone.0223862] [Medline: 31622377]
18. Awadalla S, Davies EB, Glazebrook C. A longitudinal cohort study to explore the relationship between depression, anxiety and academic performance among Emirati university students. BMC Psychiatry 2020 Sep 11;20(1):448 [FREE Full text] [doi: 10.1186/s12888-020-02854-z] [Medline: 32917172]
19. Opdal IM, Morseth B, Handegård BH, Lillevoll K, Ask H, Nielsen CS, et al. Change in physical activity is not associated with change in mental distress among adolescents: the Tromsø study: Fit Futures. BMC Public Health 2019 Jul 09;19(1):916 [FREE Full text] [doi: 10.1186/s12889-019-7271-6] [Medline: 31288796]

20. Kraut R, Patterson M, Lundmark V, Kiesler S, Mukopadhyay T, Scherlis W. Internet paradox. a social technology that reduces social involvement and psychological well-being? Am Psychol 1998 Sep;53(9):1017-1031. [doi: 10.1037//0003-066x.53.9.1017] [Medline: 9841579]
21. Subrahmanyam K, Reich SM, Waechter N, Espinoza G. Online and offline social networks: use of social networking sites by emerging adults. J Appl Dev Psychol 2008 Nov;29(6):420-433. [doi: 10.1016/j.appdev.2008.07.003]
22. Gentzler AL, Oberhauser AM, Westerman D, Nadorff DK. College students' use of electronic communication with parents: links to loneliness, attachment, and relationship quality. Cyberpsychol Behav Soc Netw 2011;14(1-2):71-74. [doi: 10.1089/cyber.2009.0409] [Medline: 20973676]
23. Reich SM, Subrahmanyam K, Espinoza G. Friending, IMing, and hanging out face-to-face: overlap in adolescents' online and offline social networks. Dev Psychol 2012 Mar;48(2):356-368. [doi: 10.1037/a0026980] [Medline: 22369341]
24. Briones R, Janoske M. How American students perceive social networking sites: an application of uses and gratifications theory. In: Lssa T, Lsaias P, Kommers P, editors. Social Networking and Education: Global Perspectives. Cham: Springer International Publishing; 2016:239-253 URL: https://link.springer.com/content/pdf/10.1007%2F978-3-319-17716-8.pdf
25. Chan DK, Cheng GH. A comparison of offline and online friendship qualities at different stages of relationship development. J Soc Pers Relat 2016 Jun 30;21(3):305-320. [doi: 10.1177/0265407504042834]
26. O'Shea G, Spence SH, Donovan CL. Interpersonal factors associated with depression in adolescents: are these consistent with theories underpinning interpersonal psychotherapy? Clin Psychol Psychother 2014;21(6):548-558. [doi: 10.1002/cpp.1849] [Medline: 23801523]
27. Hu J, Ai H. Self-esteem mediates the effect of the parent-adolescent relationship on depression. J Health Psychol 2016 Jun;21(6):897-904. [doi: 10.1177/1359105314541315] [Medline: 25030796]
28. Zhuang M, Gan Y, Liu H. The revision and preliminary application of social skill scale [Chinese]. Chinese Mental Health Journal 2004;18(11):755-759. [doi: 10.3321/j.issn:1000-6729.2004.11.001]
29. Wang J, Jackson LA, Gaskin J, Wang H. The effects of Social Networking Site (SNS) use on college students' friendship and well-being. Comput Human Behav 2014 Aug;37:229-236. [doi: 10.1016/j.chb.2014.04.051]
30. Valkenburg PM, Peter J. Preadolescents' and adolescents' online communication and their closeness to friends. Dev Psychol 2007 Mar;43(2):267-277. [doi: 10.1037/0012-1649.43.2.267] [Medline: 17352538]
31. Hampton K, Goulet L, Rainie L, Purcell K. Social networking sites and our lives. Pew Research Center. 2011 Jun 16. URL: http://www.stat.columbia.edu/~gelman/surveys.course/HamptonGouletRainiePurcell2011.pdf [accessed 2021-08-30]
32. Olson D, Liu J, Shultz K. The influence of Facebook usage on perceptions of social support, personal efficacy, and life satisfaction. J Organ Psychol 2012;12:133-144 [FREE Full text] [doi: 10.1023/A:1017032020760]
33. Nabi RL, Prestin A, So J. Facebook friends with (health) benefits? exploring social network site use and perceptions of social support, stress, and well-being. Cyberpsychol Behav Soc Netw 2013 Oct;16(10):721-727. [doi: 10.1089/cyber.2012.0521] [Medline: 23790356]
34. Li X, Chen W, Popiel P. What happens on Facebook stays on Facebook? the implications of Facebook interaction for perceived, receiving, and giving social support. Comput Human Behav 2015 Oct;51:106-113. [doi: 10.1016/j.chb.2015.04.066]
35. Vitak J, Ellison N, Steinfield C. The ties that bond: re-examining the relationship between Facebook use and bonding social capital. 2011 Presented at: 44th International Conference on System Sciences; January 6-9; Hawaii p. 1-10 URL: https://www.researchgate.net/publication/224221312_The_Ties_That_Bond_Re-Examining_the_Relationship_between_Facebook_Use_and_Bonding_Social_Capital [doi: 10.1109/hicss.2011.435]
36. Quinn S, Oldmeadow JA. Is the igeneration a 'we' generation? social networking use among 9- to 13-year-olds and belonging. Br J Dev Psychol 2013 Mar;31(Pt 1):136-142. [doi: 10.1111/bjdp.12007] [Medline: 23331112]
37. Field T, Diego M, Sanders C. Adolescent depression and risk factors. Adolescence 2001;36(143):491-498. [Medline: 11817630]
38. Wilkinson RB. Best friend attachment versus peer attachment in the prediction of adolescent psychological adjustment. J Adolesc 2010 Oct;33(5):709-717. [doi: 10.1016/j.adolescence.2009.10.013] [Medline: 19910040]
39. He F, Zhou Q, Li J, Cao R, Guan H. Effect of social support on depression of internet addicts and the mediating role of loneliness. Int J Ment Health Syst 2014;8:34 [FREE Full text] [doi: 10.1186/1752-4458-8-34] [Medline: 25147581]
40. Smith NR, Clark C, Smuk M, Cummins S, Stansfeld SA. The influence of social support on ethnic differences in well-being and depression in adolescents: findings from the prospective Olympic Regeneration in East London (ORiEL) study. Soc Psychiatry Psychiatr Epidemiol 2015 Nov;50(11):1701-1711. [doi: 10.1007/s00127-015-1098-y] [Medline: 26205126]
41. Cyranowski JM, Frank E, Young E, Shear MK. Adolescent onset of the gender difference in lifetime rates of major depression: a theoretical model. Arch Gen Psychiatry 2000 Jan;57(1):21-27. [doi: 10.1001/archpsyc.57.1.21] [Medline: 10632229]
42. Sanders RA. Adolescent psychosocial, social, and cognitive development. Pediatr Rev 2013 Aug;34(8):354-8; quiz 358-9. [doi: 10.1542/pir.34-8-354] [Medline: 23908362]
43. Christie D, Viner R. Adolescent development. BMJ 2005 Mar 05;330(7486):301-304 [FREE Full text] [doi: 10.1136/bmj.330.7486.301] [Medline: 15695279]

44. Lewinsohn PM, Clarke GN, Seeley JR, Rohde P. Major depression in community adolescents: age at onset, episode duration, and time to recurrence. J Am Acad Child Adolesc Psychiatry 1994;33(6):809-818. [doi: 10.1097/00004583-199407000-00006] [Medline: 7598758]
45. Kessler RC, Avenevoli S, Ries MK. Mood disorders in children and adolescents: an epidemiologic perspective. Biol Psychiatry 2001 Jun 15;49(12):1002-1014. [doi: 10.1016/s0006-3223(01)01129-5] [Medline: 11430842]
46. Wang M, Armour C, Wu Y, Ren F, Zhu X, Yao S. Factor structure of the CES-D and measurement invariance across gender in mainland Chinese adolescents. J Clin Psychol 2013 Sep;69(9):966-979. [doi: 10.1002/jclp.21978] [Medline: 23775279]
47. Tang J, Yu Y, Du Y, Ma Y, Zhang D, Wang J. Prevalence of internet addiction and its association with stressful life events and psychological symptoms among adolescent internet users. Addict Behav 2014 Mar;39(3):744-747. [doi: 10.1016/j.addbeh.2013.12.010] [Medline: 24388433]
48. Pietsch K, Allgaier A, Frühe B, Hoyler A, Rohde S, Schulte-Körne G. Screening for adolescent depression in paediatric care: validity of a new brief version of the Center for Epidemiological Studies Depression Scale. Child Adolesc Ment Health 2013 May;18(2):76-81. [doi: 10.1111/j.1475-3588.2012.00650.x] [Medline: 32847288]
49. Li J, Lau JTF, Mo PKH, Su X, Wu AMS, Tang J, et al. Validation of the social networking activity intensity scale among junior middle school students in China. PLoS One 2016;11(10):e0165695 [FREE Full text] [doi: 10.1371/journal.pone.0165695] [Medline: 27798699]
50. Wu H, Liu P, Meng H. Norm, reliability and validity of childern and adolescents' QOL scale [Chinese]. Chinese Journal of School Health 2006;27(1):18-21. [doi: 10.3969/j.issn.1000-9817.2006.01.017]
51. Ye HH, Chen WR, Deng DM, Lin Z, Yang WH. [Quality of life assessment in children with congenital bilateral cataract]. Zhonghua Yan Ke Za Zhi 2007 Nov;43(11):996-999. [Medline: 18307942]
52. Chou K. Assessing Chinese adolescents' social support: the multidimensional scale of perceived social support. Pers Individ Dif 2000 Feb;28(2):299-307. [doi: 10.1016/s0191-8869(99)00098-7]
53. Fang XY, Dong Q. Parent-child conflicts during early adolescence [Chinese]. Psychology Research 1998;21:122-125. [doi: 10.3969/j.issn.1006-6020.2003.04.003]
54. Fang X, Zhang J, Sun L, Liu Z. Parent-adolescent conflict and adolescents' social adjustment [Chinese]. Chinese Journal of Applied Psychology 2003;9(4):14-21. [doi: 10.3969/j.issn.1006-6020.2003.04.003]
55. Baron RA, Markman GD. Beyond social capital: the role of entrepreneurs' social competence in their financial success. J Bus Ventur 2003 Jan;18(1):41-60. [doi: 10.1016/s0883-9026(00)00069-0]
56. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 1986 Dec;51(6):1173-1182. [doi: 10.1037//0022-3514.51.6.1173] [Medline: 3806354]
57. MacKinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. Annu Rev Psychol 2007;58:593-614 [FREE Full text] [doi: 10.1146/annurev.psych.58.110405.085542] [Medline: 16968208]
58. Teng M, Hsu L, Wu S, Sun Y, Juan S, Ko Y. Association of CDH13 genotypes/haplotypes with circulating adiponectin levels, metabolic syndrome, and related metabolic phenotypes: the role of the suppression effect. PLoS One 2015;10(4):e0122664 [FREE Full text] [doi: 10.1371/journal.pone.0122664] [Medline: 25875811]
59. Bentley J. An examination of statistical methods for longitudinal mediation modeling. The University of Alabama. 2011. URL: https://www.proquest.com/openview/0a218171ab0097f5209dbab79efb9e6d/1.pdf?pq-origsite=gscholar&cbl=18750 [accessed 2021-08-30]
60. Keyes KM, Gary D, O'Malley PM, Hamilton A, Schulenberg J. Recent increases in depressive symptoms among US adolescents: trends from 1991 to 2018. Soc Psychiatry Psychiatr Epidemiol 2019 Aug;54(8):987-996 [FREE Full text] [doi: 10.1007/s00127-019-01697-8] [Medline: 30929042]
61. Thorisdottir IE, Asgeirsdottir BB, Sigurvinsdottir R, Allegrante JP, Sigfusdottir ID. The increase in symptoms of anxiety and depressed mood among Icelandic adolescents: time trend between 2006 and 2016. Eur J Public Health 2017 Oct 01;27(5):856-861. [doi: 10.1093/eurpub/ckx111] [Medline: 28957485]
62. Meeus W, Iedema J, Maassen G, Engels R. Separation-individuation revisited: on the interplay of parent-adolescent relations, identity and emotional adjustment in adolescence. J Adolesc 2005 Feb;28(1):89-106. [doi: 10.1016/j.adolescence.2004.07.003] [Medline: 15683637]
63. Spies SLA, Margolin G. Growing up wired: social networking sites and adolescent psychosocial development. Clin Child Fam Psychol Rev 2014 Mar;17(1):1-18 [FREE Full text] [doi: 10.1007/s10567-013-0135-1] [Medline: 23645343]
64. Keizer R, Helmerhorst KOW, van Rijn-van Gelderen L. Perceived quality of the mother-adolescent and father-adolescent attachment relationship and adolescents' self-esteem. J Youth Adolesc 2019 Jun;48(6):1203-1217 [FREE Full text] [doi: 10.1007/s10964-019-01007-0] [Medline: 30887261]
65. Laghi F, Schneider BH, Vitoroulis I, Coplan RJ, Baiocco R, Amichai-Hamburger Y, et al. Knowing when not to use the internet: shyness and adolescents' on-line and off-line interactions with friends. Comput Human Behav 2013 Jan;29(1):51-57. [doi: 10.1016/j.chb.2012.07.015]
66. Li S, Chang YY, Chiou W. Things online social networking can take away: reminders of social networking sites undermine the desirability of offline socializing and pleasures. Scand J Psychol 2017 Apr;58(2):179-184. [doi: 10.1111/sjop.12348] [Medline: 27925217]

67. Collins W, Russell G. Mother-child and father-child relationships in middle childhood and adolescence: a developmental analysis. Developmental Review 1991 Jun;11(2):99-136. [doi: 10.1016/0273-2297(91)90004-8]
68. Wierson M, Armistead L, Forehand R, Thomas AM, Fauber R. Parent-adolescent conflict and stress as a parent: are there differences between being a mother or a father? J Fam Viol 1990 Sep;5(3):187-197. [doi: 10.1007/bf00980814]
69. Krishnan U. Parent-adolescent conflict and adolescent functioning in a collectivist, ethnically heterogenous culture: Malaysia. The Ohio State University/OhioLINK. 2004 May 20. URL: https://etd.ohiolink.edu/apexprod/rws_etd/send_file/send?accession=osu1079038747&disposition=inline [accessed 2021-08-30]
70. Tsitsika AK, Tzavela EC, Janikian M, Ólafsson K, Iordache A, Schoenmakers TM, et al. Online social networking in adolescence: patterns of use in six European countries and links with psychosocial functioning. J Adolesc Health 2014 Jul;55(1):141-147. [doi: 10.1016/j.jadohealth.2013.11.010] [Medline: 24618179]
71. Yan H, Zhang R, Oniffrey TM, Chen G, Wang Y, Wu Y, et al. Associations among screen time and unhealthy behaviors, academic performance, and well-being in Chinese adolescents. Int J Environ Res Public Health 2017 Jun 04;14(6):596 [FREE Full text] [doi: 10.3390/ijerph14060596] [Medline: 28587225]
72. Cabrera NJ, Volling BL, Barr R. Fathers are parents, too! widening the lens on parenting for children's development. Child Dev Perspect 2018 Jan 08;12(3):152-157. [doi: 10.1111/cdep.12275]
73. Tam VCW. A comparison of fathers' and mothers' contributions in the prediction of academic performance of school-age children in Hong Kong. Int J Psychol 2009 Apr;44(2):147-156. [doi: 10.1080/00207590801910242] [Medline: 22029456]
74. Brendgen M, Lamarche V, Wanner B, Vitaro F. Links between friendship relations and early adolescents' trajectories of depressed mood. Dev Psychol 2010 Mar;46(2):491-501. [doi: 10.1037/a0017413] [Medline: 20210508]
75. Davis K. Friendship 2.0: adolescents' experiences of belonging and self-disclosure online. J Adolesc 2012 Dec;35(6):1527-1536. [doi: 10.1016/j.adolescence.2012.02.013] [Medline: 22475444]
76. Grieve R, Indian M, Witteveen K, Anne Tolan G, Marrington J. Face-to-face or Facebook: can social connectedness be derived online? Comput Human Behav 2013 May;29(3):604-609. [doi: 10.1016/j.chb.2012.11.017]
77. Jelenchick LA, Eickhoff JC, Moreno MA. "Facebook depression?" social networking site use and depression in older adolescents. J Adolesc Health 2013 Jan;52(1):128-130. [doi: 10.1016/j.jadohealth.2012.05.008] [Medline: 23260846]
78. Tidwell LC, Walther JB. Computer‐mediated communication effects on disclosure, impressions, and interpersonal evaluations: getting to know one another a bit at a time. Human Comm Res 2002 Jul;28(3):317-348. [doi: 10.1111/j.1468-2958.2002.tb00811.x]
79. Lee K, Noh M, Koo D. Lonely people are no longer lonely on social networking sites: the mediating role of self-disclosure and social support. Cyberpsychol Behav Soc Netw 2013 Jun;16(6):413-418. [doi: 10.1089/cyber.2012.0553] [Medline: 23621716]
80. Zhao H, Xu Y, Wang F, Jiang J, Zhang X, Wang X. Chinese adolescents' coping tactics in a parent-adolescent conflict and their relationships with life satisfaction: the differences between coping with mother and father. Front Psychol 2015 Oct 15;6:1572 [FREE Full text] [doi: 10.3389/fpsyg.2015.01572] [Medline: 26528224]
81. Ploeg J, Markle-Reid M, Valaitis R, McAiney C, Duggleby W, Bartholomew A, et al. Web-based interventions to improve mental health, general caregiving outcomes, and general health for informal caregivers of adults with chronic conditions living in the community: rapid evidence review. J Med Internet Res 2017 Jul 28;19(7):e263 [FREE Full text] [doi: 10.2196/jmir.7564] [Medline: 28754652]
82. Ho T. Functional similarities and differences between online and offline social support on psychological well-being. The University of Hong Kong. 2014. URL: http://hub.hku.hk/handle/10722/209695 [accessed 2021-08-30]

Abbreviations

CES-D: Center for Epidemiological Studies–Depression Scale
EFUI: Entertainment Function Use Intensity
SFUI: Social Function Use Intensity

Edited by R Kukafka; submitted 11.06.20; peer-reviewed by S Byrne, E Said-Hung; comments to author 28.07.20; revised version received 21.09.20; accepted 04.08.21; published 21.09.21.

Please cite as:
Li JB, Feng LF, Wu AMS, Mai JC, Chen YX, Mo PKH, Lau JTF
Roles of Psychosocial Factors on the Association Between Online Social Networking Use Intensity and Depressive Symptoms Among Adolescents: Prospective Cohort Study
J Med Internet Res 2021;23(9):e21316
URL: https://www.jmir.org/2021/9/e21316
doi:10.2196/21316
PMID:34546173

Original Paper

Factors Influencing the Popularity of a Health-Related Answer on a Chinese Question-and-Answer Website: Case Study

Jinhui Li[1,2], PhD; Han Zheng[3], MSc; Xu Duan[3], MA

[1]School of Journalism and Communication, Jinan University, Guangzhou, China

[2]National Media Experimental Teaching Demonstration Center, Jinan University, Guangzhou, China

[3]Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore, Singapore

Corresponding Author:
Han Zheng, MSc
Wee Kim Wee School of Communication and Information
Nanyang Technological University
#05-10, WKWSCI Building, 31 Nanyang Link
Singapore, 637718
Singapore
Phone: 65 86691766
Email: han019@e.ntu.edu.sg

Abstract

Background: Social question-and-answer (Q&A) sites have become an important venue for individuals to obtain and share human papillomavirus (HPV) vaccine knowledge.

Objective: This study aims to examine how different features of an HPV vaccine–related answer are associated with users' response behaviors on social Q&A websites.

Methods: A total of 2953 answers and 270 corresponding questions regarding the HPV vaccine were collected from a leading Chinese social Q&A platform, Zhihu. Three types of key features, including content, context, and contributor, were extracted and coded. Negative binomial regression models were used to examine their impact on the vote and comment count of an HPV vaccine–related answer.

Results: The findings showed that both content length and vividness were positively related to the response behaviors of HPV vaccine–related answers. In addition, compared with answers under the question theme *benefits and risks*, answers under the question theme *vaccination experience* received fewer votes and answers under the theme *news opinions* received more votes but fewer comments. The effects of characteristics of contributors were also supported, suggesting that answers from a male contributor with more followers and no professional identity would attract more votes and comments from community members. The significant interaction effect between content and context features further showed that long and vivid answers about HPV vaccination experience were more likely to receive votes and comments of users than those about benefits and risks.

Conclusions: The study provides a complete picture of the underlying mechanism behind response behaviors of users toward HPV vaccine–related answers on social Q&A websites. The results help health community organizers develop better strategies for building and maintaining a vibrant web-based community for communicating HPV vaccine knowledge.

(J Med Internet Res 2021;23(9):e29885) doi:10.2196/29885

KEYWORDS

answer-response behaviors; Zhihu; HPV vaccine information; content features; context features; contributor features

Introduction

Background

With the prevalence of Web 2.0, social media has become a significant venue for individuals to exchange health-related information. Compared with traditional information sources (eg, television and newspapers), social media provides more customized information with higher efficiency for users to access. Health information–seeking behaviors and needs on social media have inspired a number of recent research streams [1,2], engendering crucial insights into the treatment of various health issues such as diabetes, HIV, chronic illnesses, and mental health. Among them, a significant number of studies have focused on information related to human papillomavirus (HPV)

vaccines on social media [3-6]. HPV is a common sexually transmitted infection responsible for most cervical cancer cases [7]. The HPV vaccine has been recognized as an effective method to prevent cervical cancer, but its vaccination rate among young women remains low in developing regions, such as 1.1% in Asia and 1.2% in Africa [8]. In China, only 3.1% of college students have been vaccinated against HPV [9].

The literature has suggested that HPV vaccine–related information on social media plays an important role in shaping attitudes and behaviors of individuals toward HPV vaccination [10-12]. For example, both a survey study [13] and a content analysis [14] showed that information about HPV vaccine safety concerns might affect HPV vaccination acceptance and behaviors. Although these studies provide valuable insights on knowledge sharing of HPV vaccines on different social media platforms, few of them have investigated what makes an HPV vaccine answer popular in social question-and-answer (Q&A) sites.

Social Q&A websites are a type of social media that were originally designed for knowledge creation and sharing. It presents the content with a *Q&A* structure, where users ask and answer questions of one another about a wide range of topics [15]. Recent years have witnessed the increasing popularity for the public to seek and share health information on social Q&A platforms [16,17]. Unlike other social media platforms that are often centered on tight social relationships, social Q&A websites are typically based on loose social connections between community members who share a common interest. It is especially important to understand what makes an answer popular, as it would facilitate the maintenance of a vibrant web-based community on social Q&A platforms [18]. However, little research effort has been made to examine the popularity of predictors of health-related answers, not to mention the answers about the HPV vaccine topic. To address this research gap, this study aims to advance the existing literature on social engagement of users on social Q&A platforms by investigating how different characteristics of an HPV vaccine–related answer affect response behaviors of users (ie, voting and commenting) on Zhihu, a Chinese social Q&A platform. The findings of this study could benefit academia and practice by identifying and formulating well-grounded strategies to encourage user engagement in health knowledge discussions on social Q&A websites, which could facilitate the amplification of HPV vaccine information on social media and eventually enhance vaccine acceptance and vaccination behavior.

Literature Review and Hypotheses

Behavioral engagement of users (eg, voting and commenting) on social Q&A platforms is a key driver of platform growth and its long-term success [18]. The quality of answers on these platforms is critical for users' engagement with the content and the platform. In this study, we are particularly interested in exploring various features of health-related answers that can shape the actual behavioral engagement of users on social Q&A websites. Previous studies have used existing theoretical frameworks to understand potential features. For example, based on the heuristic-systematic model, Wang et al [19] developed a conceptual framework consisting of content and sender factors that drive cancer information diffusion on social media. Some social Q&A studies [20,21] further emphasized the question features that may influence the user response rate, such as the topic and the posting time of a question. Inspired by the abovementioned studies, this study proposed three main features of a health-related answer that may affect response behaviors of users, including content (answer-related features), context (question-related features), and contributor (user-related features).

Content Features

In the context of web-based communities, textual content is the basic element for conveying information. Content length, which is defined in this study as the number of words that a contributor writes for an answer, is often considered as an important indicator for comprehension and helpfulness of textual content. A post with a longer length indicates a greater likelihood of involving detailed descriptions [22], increasing the perceived usefulness of the audience [23]. As such, an answer containing more words is more likely to satisfy the health knowledge needs of users in a social Q&A environment. For example, in an HPV vaccine–related web-based discussion forum, a longer answer might include more details, such as the risks and benefits of the HPV vaccine or vaccine uptake experience [24]. Thus, long answers could convey more relevant knowledge to readers and aid their understanding of HPV vaccination.

The content vividness of an answer is another important feature that facilitates information dissemination on social Q&A platforms. In this study, we define content vividness as the amount of sensory information (eg, image) provided in an answer. Fang et al [18] argued that content vividness is useful for attracting the attention of users by increasing information richness and visual richness. On social media platforms, the vividness of a post is commonly evoked by multimedia, such as pictures or videos. Yin et al [25] found that using multimedia to complement textual content can increase information credibility in sharing posts. Furthermore, Jiang and Benbasat [26] observed that content vividness enhances user enjoyment. Hence, content vividness can bring both functional and emotional benefits to readers in a social Q&A environment.

According to the social exchange theory [27,28], when people receive help and support from peers in a social network, they feel the need to reciprocate such help. Applying this theory to the social Q&A context, users might feel obligated to pay back the contributors who share knowledge in an answer, especially when they perceive that the answer successfully fulfills their information needs through a detailed description and some visual aids. Post replying is the most common repaying behavior because it can promote the recognition and reputation of contributors [18]. Many studies have demonstrated that health-related content on social media predicts user social engagement, such as sharing or commenting [4,24]. For example, a study by Rus and Cameron [29] indicated that diabetes-related messages with images receive higher rates of user engagement compared with those without images. Similarly, Fang et al [18] also found that post length and information vividness affect the number of reply posts in a

web-based community. Taken together, we propose the following hypothesis:

Hypothesis 1: content features, including (a) content length and (b) content vividness, are positively associated with response behaviors toward HPV vaccine–related answers.

Context Features

The core concept of social Q&A platforms is that users are allowed to express an information need in the form of a question and seek answers from fellow users in the community [30]. The themes of these questions vary from one to another. For example, Sharon et al [31] found that vaccine-related questions focus on different topics, such as the benefits and risks of vaccination and vaccine schedules. These questions have become an essential part of a diverse and vibrant whole on many platforms, reflecting and supporting users with different goals and expectations [32]. Considering the variety of question themes, the way in which knowledge contributors answer a question might be different [33]. There are several early attempts to investigate user preferences for answering certain themes of questions. For instance, Hsieh and Counts [34] found that *socially conducive* questions are more favored by users, whereas Harper et al [35] indicated that *advice-seeking* questions receive longer answers than *fact-seeking* questions. Depending on the theme of the question, the context of answering that question is also different. Therefore, in this study, we argue that the question theme might serve as the context feature of an answer on a social Q&A platform.

Despite a growing body of literature investigating users' response behaviors in web-based communities [18,29], limited research has examined how the situational context behind a question affects the popularity of its answers. In a social Q&A context, answers received from fellow users are often expected to address what the asker articulates in the question. In other words, users' evaluation of these answers is closely related to the asker's perceived expectations indicated by the questions [32]. In a study of the social Q&A site *Yahoo! Answers*, Kim et al [36] found that question themes shape the evaluation criteria of users to select the *best answer*. As an answer with a high evaluation score may trigger more social engagement (eg, liking and commenting) [18,37] and different question themes may be evaluated differently by the users, it is reasonable to assume that question themes can affect the number of likes and replies in an answer when people contribute HPV vaccine knowledge on a social Q&A platform. Therefore, we hypothesize the following:

Hypothesis 2: the question theme is associated with response behaviors toward HPV vaccine–related answers.

In addition to the main effects mentioned earlier, there might be an interaction effect between content and context features on response behaviors of users. It is possible that the effects of a long and vivid answer on response behaviors would be amplified within specific question themes. For instance, a more detailed answer with additional images to the question asking people's experience of HPV vaccination might appeal to more users, thus triggering more discussion from the community members when compared with those questions under another theme (such as costs of HPV vaccines). In other words, more vivid, comprehensive answers under certain question themes might better satisfy health information needs of users and generate more user engagement. Understanding this interaction effect is valuable in addressing the practical needs of community developers who want to maximize user engagement on their social Q&A websites. Considering that there is limited research examining such synergistic effects on social engagement, we propose the following research question:

Research question: does the question theme moderate the effects of content length and content vividness on response behaviors toward HPV vaccine–related answers?

Contributor Features

When browsing user-generated content on social media, users often evaluate the characteristics of a source to determine its credibility, which further affects their behavioral intentions. Source credibility theory [38] identifies source expertise and trustworthiness as key elements that affect information credibility. Whether a message sender is perceived as highly credible is determined by others' evaluation of their knowledge, occupation, social training, or experience [39]. In the web-based environment, such evaluations are based on the social cues provided by the contributor in the site network, such as gender, education level, profession, and the number of followers. These cues may serve as heuristics that impact source credibility, translating to the contributor's perceived expertise and trustworthiness.

For instance, compared with female sources, information coming from male sources is more likely to be perceived as credible, especially on information-focused topics [40]. Furthermore, those who have higher education levels are perceived to have sufficient expertise; thus, they are considered credible [41]. Profession also contributes to the perceived expertise of the contributor, which predicts perceived credibility [42,43]. Moreover, users who have more followers are more likely to be considered trustworthy than those with a low number of followers [44]. All these social cues of source credibility will eventually affect other social engagement of users with the information. Finally, Kareklas et al [45] conducted a web-based experiment among participants who viewed 4 fictitious individuals who either expressed pro- or antivaccination viewpoints. Their results suggest that individuals who are perceived as credible are more likely to shape consumers' vaccination attitudes and behavioral intentions. As such, users of social Q&A platforms constantly check the credibility of an answer's contributor through social cues (ie, gender, education level, follower account, and profession); such credibility evaluations are critical in affecting users' participation in blog-mediated communication [46]. Thus, we propose the following hypothesis:

Hypothesis 3: contributor features such as (1) gender, (2) education level, (3) follower count, and (4) profession are associated with response behaviors toward HPV vaccine–related answers.

Methods

Data Collection

Data were collected from a leading social Q&A website in Zhihu, China [1]. Launched in 2011, Zhihu is the largest Chinese social Q&A website with over 160 million registered users and 26 million daily active users as of July 2018 [47]. It is a platform that facilitates knowledge creation and sharing in the public. In Zhihu, questions are created, answered, edited, and managed by users. There is a wide range of topics discussed on the Zhihu platform, enabling users to have convenient access to specific questions and answers. For example, on the platform, HPV vaccine (Chinese "宫颈癌疫苗") is a unifying topic where one can find relevant questions and answers. A screenshot of the Zhihu interface is shown in Figure 1, which shows one question related to the HPV vaccine and the answers contributed by the community members. Under this particular topic, users could raise HPV vaccine–related questions while other community members with relevant knowledge (ie, contributors) answer the questions. We extracted all the questions with respective answers relating to the HPV vaccine available on Zhihu using the Python Web Crawler (Python Software Foundation) in March 2019. A total of 2953 answers with 270 corresponding questions were collected for the study. In addition, profile information of contributors (eg, gender and education) was also collected for data analysis.

Figure 1. Screenshot of the main interface of Zhihu.

Operationalization of Study Variables

We hypothesized that three types of features (ie, content, context, and contributor features) are determinants of response behaviors on social Q&A platforms. The concepts and measurements used in this study are summarized in Table 1. First, content features included content length and content vividness, which were operationalized, respectively, as word count (WC) and image count (IC) in a specific answer. Second, context features were assessed using the theme of each answer. In total, 3 research assistants who majored in communication were recruited to independently review all 270 questions and code them into different categories. They further discussed, based on the initial subject of the questions, and resolved the discrepancies in the coding results until a consensus was reached. To this end, a coding scheme including three types of question themes was developed: (1) benefits and risks, (2) vaccination experience, and (3) news opinions. Detailed descriptions and examples are provided in Table 1. Third, contributor features were measured by four types of personal information of each user on Zhihu, including gender (male or female), education (presented or not presented), profession (presented or not presented), and follower count. Finally, as the dependent variable of this study, response behaviors were assessed using two metrics of any given answer on Zhihu: (1) vote count (ie, the number of users who show *agreement* with one answer) and (2) comment count (ie, the number of comments in one answer).

Table 1. Measurements of constructs.

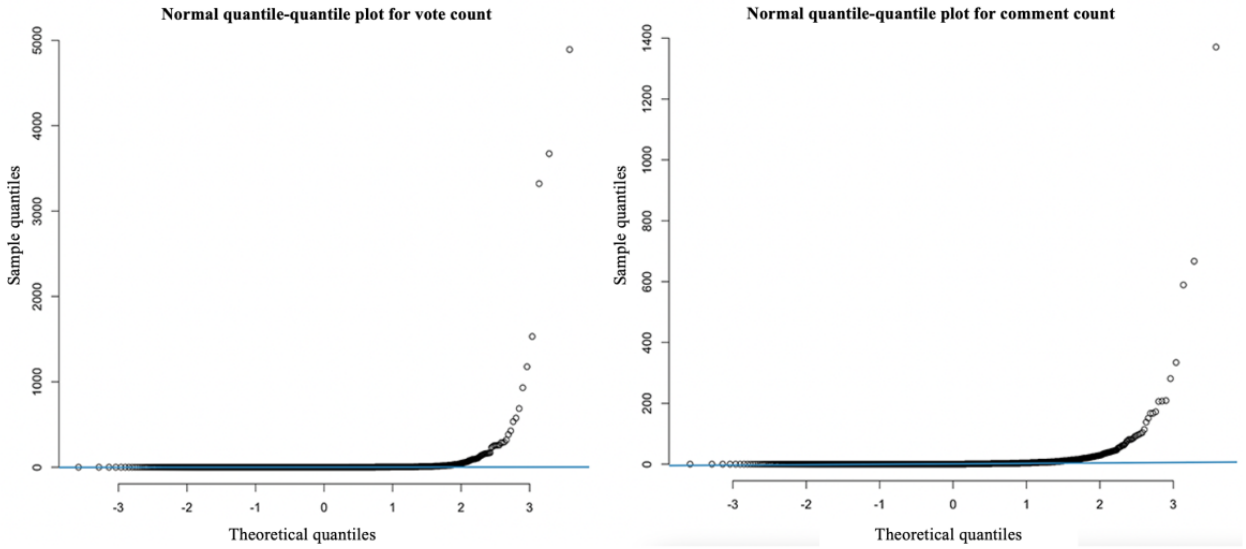
Construct and description	Value
Content features	
Word count, mean (SD)	
The number of words in answer i	307.39 (1088.77)
Image count, mean (SD)	
The number of images in answer i	0.42 (1.98)
Context features	
Question theme, n (%)	
Theme 1. Benefits and risks: referring to questions about positive and negative (side effects) effects of HPV[a] vaccine, for example "网传HPV疫苗将导致不孕及其他副作用是真的吗？ (Is the web-based information true that the HPV vaccine will cause infertility and other side effects)"	542 (18.43)
Theme 2. Vaccination experience: referring to questions about procedure and cost of HPV vaccine, for example "深圳或香港怎么预约HPV疫苗 (How to make an appointment for HPV vaccine in Shenzhen or Hong Kong)"	1917 (65.18)
Theme 3. News opinions: referring to questions about opinions or discussion of public news relating to HPV vaccine, for example "如何看待时隔11年后宫颈癌二价疫苗终于在中国大陆上市 (How do you think that the Cervarix vaccine has finally been launched in mainland China after 11 years)"	482 (16.39)
Contributor features	
Gender, n (%)	
The author's gender of answer i	
Male	1919 (65.25)
Female	1022 (34.75)
Education, n (%)	
If the author of answer i reveals educational background	
Presented	673 (22.88)
Not presented	2268 (77.12)
Profession, n (%)	
If the author of answer i reveals professional background	
Presented	443 (15.06)
Not presented	2498 (84.94)
Follower count, mean (SD)	
The number of fans of answer i's author at time t	2145.98 (27,095.82)
Response quantity	
Vote count, mean (SD)	
The number of votes of answer i at time t	10.16 (137.42)
Comment count, mean (SD)	
The number of comments of answer i at time t	4.44 (33.56)

[a]HPV: human papillomavirus.

Data Analysis

The two measures of response behaviors (ie, vote count and comment count) were count data, a type of data in which the observations can take only the nonnegative integers. As shown in Table 1, the variances of vote count and comment count (137.42 and 33.56, respectively) were significantly greater than their conditional means (10.16 and 4.44, respectively), indicating that the data distributions for the two variables were overdispersed. Furthermore, the quantile-quantile plots also show that vote count and comment count did not follow a normal distribution, and there were numerous extreme values (Figure 2). In this case, regression models using ordinary least squares might produce inconsistent, biased results [48]. Therefore, we used negative binomial regression modeling as it does not assume an equal mean and variance and introduces a parameter to correct for overdispersion.

Figure 2. Quantile-quantile plots for vote count and comment count.



We established hierarchical regression models to examine the influence of different factors on vote and comment counts. Model 1 included measures of content, context, and contributor features to test the main effects. Following this, the interactions between the measures of content and context features were added in model 2. In addition, our model included a variety of independent variables in one setting, and Bonferroni correction was used to avoid multiple comparisons fallacy. More specifically, the Bonferroni-corrected *P* value is defined as α/k *(number of tests)* [49,50]. Accordingly, we applied a significance level of *P*=.02 (*P*=.05 is divided by three themes) in our analysis. Finally, we performed the variance inflation factor test to check for multicollinearity, and the results indicated that the variance inflation factors for study variables ranged from 1.03 to 1.95, which were all below the recommended value of 10 [51]. Thus, there is no multicollinearity in the proposed empirical models.

Results

Main Effects

Tables 2 and 3 show the effects of various factors on vote and comment counts. Hypotheses 1(a) and 1(b) show that content length and content vividness are positively associated with response behaviors of users. As is evident from model 2 in the two tables, the WC of an answer was positively associated with both vote count (β=7.96E-04; incidence rate ratio [IRR]=1; *P*<.001) and comment count (β=1.35E-04; IRR=1; *P*=.004). This finding supports hypothesis 1(a). The IRRs or exponentiated values were calculated to interpret the regression coefficients of the indicator variables. For example, a one-unit change in WC increases the answer's rate of obtaining a vote and a comment by a factor of 1, while holding all other predictors constant in the model. Similarly, the IC in an answer was positively related to both vote count (β=3.96E-01; IRR=1.49; *P*<.001) and comment count (β=2.19E-01; IRR=1.25; *P*=.02). Thus, hypothesis 1(b) was supported.

Hypothesis 2 stated that the context feature of an answer (ie, a question theme) would affect response behaviors. We found that compared with the reference group *benefits and risks* (theme 1), *news opinions* (theme 3) was positively related to vote count (β=4.76E-01; IRR=1.61; *P*=.007), whereas *vaccination experience* (theme 2) was not significantly related to vote count (β=−2.87E-01; IRR=0.75; *P*=.04). In addition, compared with the reference group *benefits and risks*, *news opinions* had a negative effect on comment count (β=−4.35E-01; IRR=0.65; *P*=.004).

As for hypothesis 3, we stated that various contributor features would induce an HPV vaccine–related answer's response behavior. The results showed that male gender (β=6.68E-01; IRR=1.95; *P*<.001) and follower count (β=2.98E-01; IRR=1.35; *P*<.001) were positively associated with vote count, whereas presence of profession (*P*=.03) and education (*P*=.90) was not significantly associated with vote count. Similarly, male gender (β=2.21E-01; IRR=1.25; *P*=.01) and follower count (β=1.70E-01; IRR=1.19; *P*<.001) had similar positive effects on comment count, whereas presence of profession (*P*=.03) and education (*P*=.64) had no impact on comment count.

Table 2. Results of negative binomial regression model (vote count).

Variables	Model 1			Model 2		
	Estimate (SE)	*P* value	IRR[a]	Estimate (SE)	*P* value	IRR
Content features						
Word count	1.42E-03 (4.56E-05)	<.001	1.00	7.96E-04 (5.42E-05)	<.001	1.00
Image count	8.43E-01 (2.52E-02)	<.001	2.34	3.96E-01 (1.08E-01)	<.001	1.49
Context features (question theme)						
Theme 1[b] (reference)	—[c]	—	—	—	—	—
Theme 2[d]	2.89E-02 (1.34E-01)	.83	1.04	−2.87E-01 (1.38E-01)	.04	0.75
Theme 3[e]	5.39E-01 (1.70E-01)	.002	1.81	4.76E-01 (1.76E-01)	.007	1.61
Contributor features						
Gender (male)	6.33E-01 (1.06E-01)	<.001	1.97	6.68E-01 (1.05E-01)	<.001	1.95
Education (presented)	6.46E-02 (1.49E-01)	.67	1.10	4.71E-02 (1.48E-01)	.90	1.05
Profession (presented)	−3.75E-01 (1.69E-01)	.03	0.65	−3.57E-01 (1.67E-01)	.03	0.70
Ln(fans count)	3.07E-01 (2.74E-02)	<.001	1.31	2.98E-01 (2.73E-02)	<.001	1.35
Interaction						
WC[f] Theme 1[b] (reference)	—	—	—	—	—	—
WC Theme 2[d]	—	—	—	8.03E-04 (1.04E-04)	<.001	1.01
WC Theme 3[e]	—	—	—	1.98E-04 (2.06E-04)	.34	1.00
IC[g] Theme 1[b] (reference)	—	—	—	—	—	—
IC Theme 2[d]	—	—	—	4.88E-01 (1.11E-01)	<.001	1.63
IC Theme 3[e]	—	—	—	−4.69E-01 (1.95E-01)	.02	0.63
Model fit						
Null deviance (df[h])	3327.6 (2940)	—	—	3397.7 (2940)	—	—
Residual deviance (df)	2051.1 (2931)	—	—	2054.4 (2927)	—	—
AIC[i]	9433.7	—	—	9406.6	—	—

[a]IRR: incidence rate ratio.
[b]Theme 1: benefits and risks.
[c]Not available.
[d]Theme 2: vaccination experience.
[e]Theme 3: news opinions.
[f]WC: word count.
[g]IC: image count.
[h]df: degree of freedom.
[i]AIC: Akaike information criterion.

Table 3. Results of negative binomial regression model (comment count).

Variables	Model 1			Model 2		
	Estimate (SE)	*P* value	IRR[a]	Estimate (SE)	*P* value	IRR
Content features						
Word count	9.41E-04 (3.88E-05)	<.001	1.00	1.35E-04 (4.64E-05)	.004	1.00
Image count	1.96E-01 (2.14E-02)	<.001	1.22	2.19E-01 (9.26E-02)	.02	1.25
Context features (question theme)						
Theme 1 (reference)	—[b]	—	—	—	—	—
Theme 2	9.662E-02 (1.119E-01)	.38	1.10	−1.36E-01 (1.15E-01)	.24	0.87
Theme 3	−2.370E-01 (1.450E-01)	.10	0.79	−4.35E-01 (1.51E-01)	.004	0.65
Contributor features						
Gender (male)	1.90E-01 (9.01E-02)	.04	1.21	2.21E-01 (8.97E-02)	.01	1.25
Education (presented)	−1.05E-01 (1.26E-01)	.40	0.90	−1.04E-01 (1.26E-01)	.64	0.90
Profession (presented)	−2.98E-01 (1.43E-01)	.04	0.74	−3.04E-01 (1.43E-01)	.03	0.74
Ln(fans count)	1.67E-01 (2.31E-02)	<.001	1.18	1.70E-01 (2.32E-02)	<.001	1.19
Interaction						
WC[c] Theme 1[d] (reference)	—	—	—	—	—	—
WC Theme 2[e]	—	—	—	9.02E-04 (8.95E-05)	<.001	1.00
WC Theme 3[f]	—	—	—	1.01E-03 (1.77E-04)	<.001	1.00
IC[g] Theme 1 (reference)	—	—	—	—	—	—
IC Theme 2	—	—	—	−1.11E-02 (9.57E-02)	.91	0.99
IC Theme 3	—	—	—	−4.26E-01 (1.72E-01)	.01	0.65
Model fit						
Null deviance (df[h])	2996.0 (2940)	—	—	3025.0 (2940)	—	—
Residual deviance (df)	2398.4 (2931)	—	—	2401.1 (2927)	—	—
AIC[i]	10,711	—	—	10,700	—	—

[a]IRR: incidence rate ratio.

[b]Not available.

[c]WC: word count.

[d]Theme 1: benefits and risks.

[e]Theme 2: vaccination experience.

[f]Theme 3: news opinions.

[g]IC: image count.

[h]df: degree of freedom.

[i]AIC: Akaike information criterion.

Interaction Effects

Finally, the research question asked whether there was an interaction effect between the content and context features on response behaviors. First, for the moderating effect of WC and question theme, compared with the reference group *WC Theme 1: benefits and risks*, the interaction *WC Theme 2: vaccination experience* was positively related to vote count (β=8.03E-04; IRR=1.01; *P*<.001) and comment count (β=9.02E-04; IRR=1; *P*<.001). Moreover, the interaction between WC and news opinion was only positively associated with comment count (β=1.01E-03; IRR=1; *P*<.001). Second, for the moderating effect of IC and question theme, compared with the reference group *IC Theme 1: benefits and risks*, the interaction *IC Theme 2: vaccination experience* was positively associated with vote count (β=4.88E-01; IRR=1.63; *P*<.001). In contrast, the interaction *IC Theme 3: news opinions* was negatively associated with vote count (β=-4.69E-01; IRR=0.63; *P*=.02) and comment count (β=-4.26E-01; IRR=0.65; *P*=.01).

Discussion

Principal Findings

On social Q&A websites, social engagement of users toward an answer creates complex social feedback effects, which influence the knowledge-sharing behavior of a contributor in the future. Through the lens of existing theories, including social exchange theory and source credibility theory, we identified seven essential features from three distinct dimensions (ie, content, context, and contributor features) that may be associated with the heterogeneity in the answer-response behaviors, and evaluated the impacts of these features in a holistic and comprehensive research model that has not yet been investigated in the extant literature. The empirical results of this study lead to important findings in health communication and social media research. Overall, the results suggest that the content, context, and contributor features of an HPV vaccine–related answer are associated with other users' engagement on a social Q&A site. At the content level, the findings show that both post length and image number positively impact the number of votes and comments toward a certain answer. Although similar results have been found in user-generated content on social media or general web-based communities [18,52,53], this study further extends these key conclusions to the context of communicating health information on emerging social Q&A websites. Specifically, our findings highlight that richness and vividness of an HPV vaccine–related answer can increase the willingness of other members to respond to this post. The findings also support the applicability of social exchange theory to understand the effect of content features on the response behaviors of users. From this theoretical perspective, people tend to show support and gratitude to those who spend time and effort contributing to comprehensive knowledge [54]. Votes and comments are thus regarded as reciprocal behavior from the users who have successfully had their information and psychological needs addressed by the rich and vivid answers. As the awareness of people of an HPV vaccine has been reported to be low in China [55], this study highlights the need to include helpful and concrete messages with appropriate visual illustrations to promote relevant health knowledge sharing in the social Q&A context.

At the context level, the results reveal that the question theme can affect the response behaviors of users toward a corresponding answer on social Q&A websites. Several existing studies in the health promotion research domain [56,57] have shown that posts under certain health topics are more likely to trigger information sharing of users. This study contributes to this body of literature by suggesting that answers responding to the topics about benefits and risks of HPV vaccine have received fewer votes but more comments from web-based members, compared with those discussing relevant news of HPV vaccine. One possible explanation for this interesting observation is that, due to relatively low awareness about HPV vaccination in China, it may be difficult for the users to vote for an answer when they perceive the effects and safety of HPV vaccination as controversial. Instead, by having more latitude for self-expression than a simple vote, they are more likely to share their opinions regarding these controversial topics, thus generating more comments.

As for the characteristics of contributors, gender has a relatively large effect on response behaviors (both vote and comment count), followed by follower counts. Drawing support from source credibility theory, this study confirms the role of the demographic characteristics of a contributor as important credibility indicators in web-based knowledge-sharing communities. The findings have emphasized how these key accessible cues of source credibility of a contributor drive participation of other members in their answers on social Q&A websites. In general, an HPV vaccine–related answer posted by a male contributor with more followers but without professional identity would attract more users' engagement from community members. Gender is widely recognized as an important demographic factor in studies of credibility perceptions [58,59] and web-based behaviors [60,61]. A study conducted on Facebook [62] found that compared with women users who tend to share more personal issues, men discuss more general public topics, which ultimately elicit more comments. Consistent with the literature, our study shows that even for some *female-related* public topics such as HPV vaccine, a contributor's gender has the same influence on the response behaviors. Furthermore, contributors who have more followers would trigger more votes and comments regarding their posted answers. Compared with newcomers, they know how to create content that attracts attention, recognition, and favorable feedback from other users and are motivated to maintain their reputation in the web-based community [18]. However, this study did not support the notion that the presence of professional and educational profiles in an answer would affect its votes and comments. This conclusion is inconsistent with prior studies, which demonstrated that individuals often perceive opinions of experts as trustworthy and therefore are more likely to agree with them [63,64]. A possible explanation might be that nearly half of the Zhihu answer contributors who revealed their professional information in our data set are from the health insurance or commercial health industry. Users may not consider their posts as objective as actual health experts (such as doctors in public hospitals) because of the potential commercial interests of HPV vaccines.

More importantly, this study provides novel evidence related to the interaction effects between content and context features on response behaviors. In particular, long and vivid answers about HPV vaccination experience are more likely to receive users' votes and comments than those about risks and benefits. This finding suggests that community members prefer to read and discuss more comprehensive answers about personal vaccination stories rather than medical facts about vaccines [24]. Such stories might provide users with more meaningful and customized guidance on how to take up an HPV vaccine in terms of its costs, schedule, and feelings, especially for those who intend to do so. Interestingly, we found that when answering a question about vaccine news, answers with more images received fewer votes and comments. It is possible that users might expect answers that unpack the systemic causes and consequences of public news related to HPV vaccines, for example, the reasons and results of the delays in approving HPV

vaccines in mainland China. When it comes to HPV news discussion, users show a more favorable attitude toward answers using thematic framing instead of episodic framing [65]. Adding numerous images in this scenario might deteriorate the scientific value of such answers.

Limitations and Future Directions

This study has several limitations that may open directions for future research. First, because of the nature of a content analysis, this study only focuses on observed answer-response behaviors in social Q&A platforms, failing to explore the psychological processes underlying these behaviors. Although there is existing literature that discusses the cognitive mechanisms of how social media information shapes decision-making behaviors [66], future studies should use surveys or experimental methods to further empirically investigate the underpinning process of response behaviors in this particular health context. Second, the content features measured in this study are relatively abstract and broad, which might not be able to capture the dynamics of the concepts. It is worthwhile for future studies to explore more detailed content features, such as readability, narrativity, and coherence, which could inform public health professionals to take more specific and effective actions to promote user engagement on social Q&A websites. Third, as the primary aim of this study is to investigate web-based health answers on a Chinese social Q&A website, the findings may not be generalizable to social Q&A websites in other cultural contexts. As a result, more studies need to be conducted within different cultural contexts to further validate our findings and provide additional insights into the potential cultural influence on HPV vaccine answer-response behaviors. Finally, the wide spread of misinformation on social media is noteworthy. Although beyond the scope of this study, one future direction is to consider the veracity of information regarding HPV vaccination shared on social Q&A websites, for example, using a data mining method to examine the relationship between the truthfulness of an answer and its engagement metrics (ie, number of votes and comments).

Research and Practical Implications

This study contributes to the existing body of literature research in several ways. First, to our knowledge, this is one of the first attempts to investigate the factors predicting users' response behaviors toward health-related answers on social Q&A platforms, under a novel and comprehensive framework with features from three different layers: content, context, and contributor. The findings pave the way for understanding the social dynamics of emerging web-based knowledge-sharing communities. The framework established in this study has the potential to inspire future research to extend its applicability to other health issues on social media platforms. Second, the findings emphasize the critical role of the question theme in facilitating social engagement in the web-based knowledge-sharing context. It contributes to a complete account of contextual effects within social Q&A platforms, seldom covered in previous research. Finally, the study extends the current web-based information behavior literature by examining the interaction effect between content and context features on social Q&A websites. Although many studies have highlighted the need for comprehensive content in web-based communities [67], they neglected the confluence of comprehensive contents and contextual topics that collectively induce response behaviors. This study provides a clearer picture of why people respond to health-related information on social Q&A websites by considering the synergistic effects between content and question themes.

From a practical perspective, the results can also help health practitioners and knowledge contributors on social media to better understand the information needs and preferences of the public regarding HPV vaccines. By identifying the key factors and their interaction effects that drive users' response behaviors, this study helps health professionals develop better strategies for building and maintaining a vibrant web-based community for sharing HPV knowledge. For instance, health care providers and health communicators may use social Q&A platforms to develop campaigns to promote public awareness of HPV vaccination, especially to leverage the influence of opinion leaders on the platforms, as shown in the findings regarding contributors in this study. Specifically, the findings on characteristics of a contributor imply that male contributors are advisable to be more involved in the discussions of topics that are traditionally considered *female-only* topics, such as HPV vaccination. Perhaps, instead of framing it as a *woman's issue*, practitioners might consider involving more men as HPV vaccine spokespersons, educating the public that HPV infection is not a woman's issue but one that challenges both men and women as the disease can be transmitted across genders [68]. Health behaviors of both genders play a crucial role in the health of both men and women. As such, by engaging more men in cervical cancer prevention campaigns, practitioners are better able to erase the discrimination of *females-only issues*, increase HPV vaccination rate, and work toward a wider, healthier community in which both men and women are a part of.

In addition, social Q&A platform managers and developers can enact relevant policies and/or adopt useful tools to encourage comprehensive health content with high-quality images, such as applying optimal image-uploading technologies or offering extra incentives to informative answers. Furthermore, as implied in our findings, important members with a large group of followers (eg, opinion leaders or experts) on social media play a crucial role in engaging a wider public to discuss HPV vaccines. Therefore, social Q&A platform managers may incentivize knowledge contributors at the platform level to promote comprehensive, high-quality health content. For instance, they can apply certain algorithms to award informative answers with credits or privileges to encourage them to consistently share validated and useful information about HPV vaccines. General answer contributors should adapt different content strategies for different question topics to attract attention and feedback from other users. For instance, inserting more pictures in an answer is useful for questions about personal stories and experience, but it is not advised for questions about HPV news discussion.

Acknowledgments

This study was supported by Guangzhou City Spirit and City Image Research Center, under the Young Innovative Talent Grant of Guangdong Universities (grant 2020WQNCX005).

Conflicts of Interest

None declared.

References

1. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libr J 2017 Dec;34(4):268-283. [doi: 10.1111/hir.12192] [Medline: 29045011]
2. Hong Z, Deng Z, Evans R, Wu H. J Med Internet Res 2020 Apr 16;22(4):e13071 [FREE Full text] [doi: 10.2196/13071] [Medline: 32297872]
3. Sundstrom B, Aylor E, Cartmell KB, Brandt HM, Bryant DC, Hughes Halbert C, et al. Beyond the birds and the bees: a qualitative content analysis of online HPV vaccination communication. J Commun HealthC 2018 Jun 06;11(3):205-214. [doi: 10.1080/17538068.2018.1484984]
4. Basch CH, MacLean SA. A content analysis of HPV related posts on instagram. Hum Vaccin Immunother 2019;15(7-8):1476-1478 [FREE Full text] [doi: 10.1080/21645515.2018.1560774] [Medline: 30570379]
5. Massey PM, Leader A, Yom-Tov E, Budenz A, Fisher K, Klassen AC. Applying multiple data collection tools to quantify human papillomavirus vaccine communication on Twitter. J Med Internet Res 2016 Dec 05;18(12):e318 [FREE Full text] [doi: 10.2196/jmir.6670] [Medline: 27919863]
6. Lama Y, Hu D, Jamison A, Quinn SC, Broniatowski DA. Characterizing trends in human papillomavirus vaccine discourse on Reddit (2007-2015): an observational study. JMIR Public Health Surveill 2019 Mar 27;5(1):e12480 [FREE Full text] [doi: 10.2196/12480] [Medline: 30916662]
7. Baseman J, Koutsky L. The epidemiology of human papillomavirus infections. J Clin Virol 2005 Mar;32 Suppl 1:16-24 [FREE Full text] [doi: 10.1016/j.jcv.2004.12.008] [Medline: 15753008]
8. Bruni L, Diaz M, Barrionuevo-Rosas L, Herrero R, Bray F, Bosch FX, et al. Global estimates of human papillomavirus vaccination coverage by region and income level: a pooled analysis. Lancet Glob Health 2016 Jul;4(7):e453-e463 [FREE Full text] [doi: 10.1016/S2214-109X(16)30099-7] [Medline: 27340003]
9. Deng C, Chen X, Liu Y. Human papillomavirus vaccination: coverage rate, knowledge, acceptance, and associated factors in college students in mainland China. Hum Vaccin Immunother 2021 Mar 04;17(3):828-835 [FREE Full text] [doi: 10.1080/21645515.2020.1797368] [Medline: 32873128]
10. Ingledue K, Cottrell R, Bernard A. College women's knowledge, perceptions, and preventive behaviors regarding human papillomavirus infection and cervical cancer. Am J Health Stud 2004;19(1):28-34 [FREE Full text]
11. Wang X, Zhou X, Leesa L, Mantwill S. The effect of vaccine literacy on parental trust and intention to vaccinate after a major vaccine scandal. J Health Commun 2018;23(5):413-421. [doi: 10.1080/10810730.2018.1455771] [Medline: 29589807]
12. Liu Z, Zhang L, Yang Y, Meng R, Fang T, Dong Y, et al. Active surveillance of adverse events following human papillomavirus vaccination: feasibility pilot study based on the Regional Health Care Information Platform in the city of Ningbo, China. J Med Internet Res 2020 Jun 01;22(6):e17446 [FREE Full text] [doi: 10.2196/17446] [Medline: 32234696]
13. Margolis MA, Brewer NT, Shah PD, Calo WA, Gilkey MB. Stories about HPV vaccine in social media, traditional media, and conversations. Prev Med 2019 Jan;118:251-256. [doi: 10.1016/j.ypmed.2018.11.005] [Medline: 30414396]
14. Dunn AG, Surian D, Leask J, Dey A, Mandl KD, Coiera E. Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States. Vaccine 2017 May 25;35(23):3033-3040 [FREE Full text] [doi: 10.1016/j.vaccine.2017.04.060] [Medline: 28461067]
15. Oh S, Zhang Y, Park MS. Cancer information seeking in social question and answer services: identifying health-related topics in cancer questions on Yahoo! Answers. Information Research. 2016. URL: http://informationr.net/ir/21-3/paper718.html#.YTI9xo4zayI [accessed 2021-09-03]
16. Kim S, Oh S, Oh JS. Evaluating health answers in a social Q and A site. Proc Am Soc Info Sci Tech 2009 Jun 03;45(1):1-6 [FREE Full text] [doi: 10.1002/meet.2008.14504503134]
17. Oh S, Yi YJ, Worrall A. Quality of health answers in social Q and A. Proc Am Soc Info Sci Tech 2013 Jan 24;49(1):1-6 [FREE Full text] [doi: 10.1002/meet.14504901075]
18. Fang J, Chen L, Wang X, George B. Not all posts are treated equal: An empirical investigation of post replying behavior in an online travel community. Info Manag 2018 Nov;55(7):890-900 [FREE Full text] [doi: 10.1016/j.im.2018.04.003]
19. Wang X, Chen L, Shi J, Peng T. What makes cancer information viral on social media? Comput Hum Behav 2019 Apr;93:149-156 [FREE Full text] [doi: 10.1016/j.chb.2018.12.024]
20. Liu Z, Jansen B. Factors influencing the response rate in social question and answering behavior. In: Proceedings of the 2013 Conference on Computer Supported Cooperative Work. 2013 Feb Presented at: CSCW '13: Computer Supported Cooperative Work; February 23 - 27, 2013; San Antonio Texas USA p. 1263-1274. [doi: 10.1145/2441776.2441918]

21. Li L, Huang K, Ye EM, Zhang C. Proc Asso Info Sci Tech 2019 Oct 18;56(1):709-711 [FREE Full text] [doi: 10.1002/pra2.144]
22. Chen Y, Chai Y, Liu Y, Xu Y. Analysis of review helpfulness based on consumer perspective. Tinshhua Sci Technol 2015 Jun;20(3):293-305. [doi: 10.1109/tst.2015.7128942]
23. Pan Y, Zhang J. Born Unequal: A study of the helpfulness of user-generated product reviews. J Retail 2011 Dec;87(4):598-612 [FREE Full text] [doi: 10.1016/j.jretai.2011.05.002]
24. Li J, Zheng H. Coverage of HPV-related information on chinese social media: a content analysis of articles in Zhihu. Hum Vaccin Immunother 2020 Oct 02;16(10):2548-2554 [FREE Full text] [doi: 10.1080/21645515.2020.1729028] [Medline: 32159420]
25. Yin C, Sun Y, Fang Y, Lim K. Exploring the dual-role of cognitive heuristics and the moderating effect of gender in microblog information credibility evaluation. Info Technol People 2018 May 16;31(3):741-769. [doi: 10.1108/itp-12-2016-0300]
26. Jiang Z, Benbasat I. Research Note—investigating the influence of the functional mechanisms of online product presentations. Info Sys Res 2007 Dec;18(4):454-470. [doi: 10.1287/isre.1070.0124]
27. Peter B. Exchange and Power in Social Life. New York: Routledge; 2017.
28. Homans GC. Social Behavior: Its Elementary Forms. Oxford, England: Harcourt Brace Jovanovich; 1974:1-386.
29. Rus HM, Cameron LD. Health communication in social media: Message features predicting user engagement on diabetes-related facebook pages. Ann Behav Med 2016 Dec;50(5):678-689. [doi: 10.1007/s12160-016-9793-9] [Medline: 27059761]
30. Gazan R. Social Q and A. J Am Soc Inf Sci 2011 May 23;62(12):2301-2312. [doi: 10.1002/asi.21562]
31. Sharon A, Yom-Tov E, Baram-Tsabari A. Vaccine information seeking on social Q and A services. Vaccine 2020 Mar 10;38(12):2691-2699. [doi: 10.1016/j.vaccine.2020.02.010] [Medline: 32061384]
32. Choi E, Shah C. Asking for more than an answer: What do askers expect in online Q and A services? J Info Sci 2016 May 01;43(3):424-435. [doi: 10.1177/0165551516645530]
33. Pomerantz J. A linguistic analysis of question taxonomies. J Am Soc Inf Sci 2005 May;56(7):715-728. [doi: 10.1002/asi.20162]
34. Hsieh G, Counts S. Mimir: A market-based real-time question and answer service. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 4, 2009; Boston p. 769-778. [doi: 10.1145/1518701.1518820]
35. Harper F, Raban D, Rafaeli S, Konstan J. Predictors of answer quality in online Q and A sites. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2008 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 6, 2008; Florence p. 865-874. [doi: 10.1145/1357054.1357191]
36. Kim S, Oh JS, Oh S. Best-answer selection criteria in a social Q and A site from the user-oriented relevance perspective. In: Proceedings of the American Society for Information Science and Technology. 2008 Presented at: American Society for Information Science and Technology; 2007; Milwaukee p. 1-15. [doi: 10.1002/meet.1450440256]
37. Jin J, Li Y, Zhong X, Zhai L. Why users contribute knowledge to online communities: An empirical study of an online social Q and A community. Info Manag 2015 Nov;52(7):840-849 [FREE Full text] [doi: 10.1016/j.im.2015.07.005]
38. Hovland C, Janis I, Kelley H. Communication and Persuasion. New Haven: Yale University Press; 1953:1-327.
39. Pallavicini F, Cipresso P, Mantovani F. Beyond sentiment: How social network analytics can enhance opinion mining and sentiment analysis. In: Sentiment Analysis in Social Networks. Burlington: Morgan Kaufmann; 2017:13-29.
40. Armstrong C, McAdams M. Blogs of information: How gender cue and individual motivations influence perceptions of credibility. J Comp Mediat Commun 2009;14(3):435-456. [doi: 10.1111/j.1083-6101.2009.01448.x]
41. Tewksbury D, Jensen J, Coe K. Video news releases and the public: The impact of source labeling on the perceived credibility of television news. J Commun 2011:328-348 [FREE Full text] [doi: 10.1111/j.1460-2466.2011.01542.x]
42. De Meulenaer S, De Pelsmacker P, Dens N. Power distance, uncertainty avoidance, and the effects of source credibility on health risk message compliance. Health Commun 2018 Mar;33(3):291-298. [doi: 10.1080/10410236.2016.1266573] [Medline: 28094542]
43. Ismagilova E, Slade E, Rana NP, Dwivedi YK. The effect of characteristics of source credibility on consumer behaviour: A meta-analysis. J Retail Consum Ser 2020;53:101736. [doi: 10.1016/j.jretconser.2019.01.005]
44. Jin SA, Phua J. Following Celebrities' Tweets About Brands: The impact of Twitter-based electronic word-of-mouth on consumers' source credibility perception, buying intention, and social identification with celebrities. J Advert 2014;43(2):181-195. [doi: 10.1080/00913367.2013.827606]
45. Kareklas I, Muehling DD, Weber T. Reexamining health messages in the digital age: a fresh look at source credibility effects. J Advert 2015;44(2):88-104. [doi: 10.1080/00913367.2015.1018461]
46. Yang S, Kang M, Johnson P. Effects of narratives, openness to dialogic communication, and credibility on engagement in crisis communication through organizational blogs. Commun Res 2010;37(4):473-497. [doi: 10.1177/0093650210362682]
47. Smith C. Interesting Zhihu statistics and facts (2019). DMR Business Statistics. 2021 Jan 28. URL: https://expandedramblings.com/index.php/zhihu-statistics-and-facts/ [accessed 2021-01-28]

48. Son J, Lee J, Larsen K, Woo J. Understanding the uncertainty of disaster tweets and its effect on retweeting: The perspectives of uncertainty reduction theory and information entropy. J Asso Info Sci Technol 2020;71(10):1145-1161 [FREE Full text] [doi: 10.1002/asi.24329]
49. Lee S, Lee DK. What is the proper way to apply the multiple comparison test? Korean J Anesthesiol 2018;71(5):353-360 [FREE Full text] [doi: 10.4097/kja.d.18.00242] [Medline: 30157585]
50. Zheng H, Li J, Salmon C, Theng Y. The effects of exergames on emotional well-being of older adults. Comput Hum Behav 2020;110:106383 [FREE Full text] [doi: 10.1016/j.chb.2020.106383]
51. Hair J, Tatham R, Anderson R, Black W. Multivariate Data Analysis (5th Ed). London: Prentice-Hall; 1998:1-768.
52. Liu J, Li C, Ji Y, North M, Yang F. Like it or not: The Fortune 500's Facebook strategies to generate users' electronic word-of-mouth. Comput Hum Behav 2017;73:605-613 [FREE Full text] [doi: 10.1016/j.chb.2017.03.068]
53. Xu W, Zhang C. Sentiment, richness, authority, and relevance model of information sharing during social Crises—the case of #MH370 tweets. Comput Hum Behav 2018;89:199-206 [FREE Full text] [doi: 10.1016/j.chb.2018.07.041]
54. Thibaut J, Kelley H. The Social Psychology of Groups. Oxford, England: John Wiley; 1959:1-313.
55. Tang G, Zeng G, Zhao B, Li Y, Wang R. Awareness and knowledge of human papillomavirus vaccination and their influential factors among Chinese women: a systematic review. Front Nurs 2019;6(4):317-326 [FREE Full text] [doi: 10.2478/fon-2019-0048]
56. Chung J. Retweeting in health promotion: Analysis of tweets about Breast Cancer Awareness Month. Comput Hum Behav 2017;74:112-119 [FREE Full text] [doi: 10.1016/j.chb.2017.04.025]
57. A Rahim AI, Ibrahim MI, A Salim FN, Ariffin MA. Health information engagement factors in Malaysia: a content analysis of Facebook use by the Ministry of Health in 2016 and 2017. Int J Environ Res Public Health 2019;16(4):591 [FREE Full text] [doi: 10.3390/ijerph16040591] [Medline: 30781691]
58. Bigham A, Meyers C, Li N, Irlbeck E. The effect of emphasizing credibility elements and the role of source gender on perceptions of source credibility. J Appl Commun 2019;103(2):1. [doi: 10.4148/1051-0834.2270]
59. Flanagin AJ, Metzger MJ. The perceived credibility of personal web page information as influenced by the sex of the source. Comput Hum Behav 2003;19(6):683-701. [doi: 10.1016/s0747-5632(03)00021-9]
60. Lian J, Yen D. Online shopping drivers and barriers for older adults: Age and gender differences. Comput Hum Behav 2014;37:133-143 [FREE Full text] [doi: 10.1016/j.chb.2014.04.028]
61. Sun B, Mao H, Yin C. Male and female users' differences in online technology community based on text mining. Front Psychol 2020;11:806 [FREE Full text] [doi: 10.3389/fpsyg.2020.00806] [Medline: 32528342]
62. Wang Y, Burke M, Kraut R. Gender, topic, and audience response: an analysis of user-generated content on facebook. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI '13: CHI Conference on Human Factors in Computing Systems; April 27 - May 2, 2013; Paris France p. 31-34. [doi: 10.1145/2470654.2470659]
63. Shen Y, Huang C, Chu C, Liao H. Virtual community loyalty: an interpersonal-interaction perspective. Int J Electron Comm 2014;15(1):49-74. [doi: 10.2753/jec1086-4415150102]
64. Willemsen LM, Neijens PC, Bronner F. The ironic effect of source identification on the perceived credibility of online product reviewers. J Comput-Mediat Comm 2012;18(1):16-31. [doi: 10.1111/j.1083-6101.2012.01598.x]
65. Zhang Y, Jin Y. Thematic and episodic framing of depression: how chinese and american newspapers framed a major public health threat. Athens J Mass Media Commun 2017;3(2):91-106. [doi: 10.30958/ajmmc.3.2.1]
66. Reyna VF. A scientific theory of gist communication and misinformation resistance, with implications for health, education, and policy. Proc Natl Acad Sci 2021;118(15):e1912441117 [FREE Full text] [doi: 10.1073/pnas.1912441117] [Medline: 32312815]
67. Schreiner M, Fischer T, Riedl R. Impact of content characteristics and emotion on behavioral engagement in social media: literature review and research agenda. Electron Comm Res 2019;21(2):329-345. [doi: 10.1007/s10660-019-09353-8]
68. Giuliano AR, Anic G, Nyitray AG. Epidemiology and pathology of HPV disease in males. Gynecol Oncol 2010;117(2):15-19 [FREE Full text] [doi: 10.1016/j.ygyno.2010.01.026] [Medline: 20138345]

Abbreviations

HPV: human papillomavirus
IC: image count
IRR: incidence rate ratio
Q&A: question-and-answer
WC: word count

Edited by R Kukafka; submitted 24.04.21; peer-reviewed by X Wang, N Nama, D Hu, Y Lama; comments to author 28.06.21; revised version received 02.08.21; accepted 12.08.21; published 28.09.21.

Please cite as:
Li J, Zheng H, Duan X
Factors Influencing the Popularity of a Health-Related Answer on a Chinese Question-and-Answer Website: Case Study
J Med Internet Res 2021;23(9):e29885
URL: https://www.jmir.org/2021/9/e29885
doi:10.2196/29885
PMID:34581675

Original Paper

Understanding Mental Health App Use Among Community College Students: Web-Based Survey Study

Judith Borghouts[1], BSc, MSc, PhD; Elizabeth V Eikey[2,3], PhD; Gloria Mark[4], PhD; Cinthia De Leon[1], MPH; Stephen M Schueller[4,5], PhD; Margaret Schneider[6], PhD; Nicole Stadnick[7,8,9], MPH, PhD; Kai Zheng[4], PhD; Dana B Mukamel[1], PhD; Dara H Sorkin[1], PhD

[1]Department of Medicine, University of California Irvine, Irvine, CA, United States
[2]Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego, San Diego, CA, United States
[3]The Design Lab, University of California, San Diego, San Diego, CA, United States
[4]Department of Informatics, University of California, Irvine, Irvine, CA, United States
[5]Department of Psychological Science, University of California, Irvine, Irvine, CA, United States
[6]Department of Public Health, University of California, Irvine, Irvine, CA, United States
[7]Department of Psychiatry, University of California, San Diego, La Jolla, CA, United States
[8]Dissemination and Implementation Science Center, UC San Diego Altman Clinical and Translational Research Institute, La Jolla, CA, United States
[9]Child and Adolescent Services Research Center, San Diego, CA, United States

Corresponding Author:
Judith Borghouts, BSc, MSc, PhD
Department of Medicine
University of California, Irvine
100 Theory
Irvine, CA, 92617
United States
Phone: 1 9498240246
Email: jborghou@uci.edu

Abstract

Background: Mental health concerns are a significant issue among community college students, who often have less access to resources than traditional university college students. Mobile apps have the potential to increase access to mental health care, but there has been little research investigating factors associated with mental health app use within the community college population.

Objective: This study aimed to understand facilitators of and barriers to mental health app use among community college students.

Methods: A web-based survey was administered to a randomly selected sample of 500 community college students from April 16 to June 30, 2020. Structural equation modeling was used to test the relationships between the use of mental health apps, perceived stress, perceived need to seek help for mental health concerns, perceived stigma, past use of professional mental health services, privacy concerns, and social influence of other people in using mental health apps.

Results: Of the 500 participants, 106 (21.2%) reported use of mental health apps. Perceived stress, perceived need to seek help, past use of professional services, and social influence were positively associated with mental health app use. Furthermore, the effect of stress was mediated by a perceived need to seek help. Privacy concerns were negatively associated with mental health app use. Stigma, age, and gender did not have a statistically significant effect.

Conclusions: These findings can inform development of new digital interventions and appropriate outreach strategies to engage community college students in using mental health apps.

(J Med Internet Res 2021;23(9):e27745) doi:10.2196/27745

KEYWORDS

mHealth; mental health; community college; students; structural equation modeling; mobile apps; services; mental health services; stress; privacy

Introduction

Background

Mental health concerns are a significant issue among college students [1,2], and the last decade has seen a rise in mental health concerns among students [3]. Community college students, in particular, face a growing crisis of mental health concerns. A survey conducted by the Wisconsin HOPE Lab found that almost 49.4% of respondents across 10 community colleges in seven states reported mental health issues [4]. A survey among 39,930 community college students found that 60% of respondents were housing insecure in the previous year, and 19% of respondents had been homeless [5]. These basic needs insecurities not only impact students' performance at school but also have been associated with poorer physical and mental health [6,7].

College students can face multiple barriers to accessing support and resources, and these barriers may be more pronounced for students enrolled in community college. For example, the community college population has a higher proportion of students who are single parents and/or who are working jobs while attending classes, and the students are more likely to come from a lower socioeconomic background compared to university students [8]. Whereas high workload and academic stress are common issues raised among university students, financial or housing stress may be more common among community college students [9,10]. Furthermore, community college students have less access to mental health services than university students, and they may be in particular need of increased access to mental health resources [9]. Students may also not recognize a need to seek help, or they may feel uncomfortable discussing mental health problems [11].

Given college students' limited access to mental health resources, studies have explored the use of technology-based solutions, such as mobile apps [12-14]. Mobile apps for health purposes are one of the fastest growing categories of apps, and currently, more than 10,000 mental health apps are publicly available [15]. Mental health apps may be particularly useful for students: according to a Pew Research 2019 study on mobile phone ownership, 85% of people attending college own a smartphone, and 96% of people aged 18-29 years own a smartphone [16]. Furthermore, students and young adults are active users of the internet for seeking health information [2,10,17,18].

To inform development and deployment of mental health apps for community college students, an understanding of students' mental health needs and current use of apps is necessary. However, community college students are often underrepresented in the discussion of student mental health [8,9], and the mechanisms behind community college students' use of mental health apps are not yet fully understood, despite the fact that there are 11.8 million students enrolled in community colleges in the United States [19]. There are differences in university and community college environments with regards to demographics, culture, and mental health issues [9,10]. These differences between community college and university student populations may affect students' mental health needs and subsequent app use. To address this gap, this paper aimed to address what factors are associated with community college students' use of mental health apps. Specifically, our study focused on 6 factors that prior research has found to affect use of mental health resources: perceived need to seek help for mental health concerns, perceived stigma, past use of professional mental health services, privacy concerns, and social influence of other people in using mental health apps.

Mental Health App Interest and Use Among Students

With the growing development and ubiquity of mental health apps, research has been performed on both interest in and use of these apps among students. Kern et al [20] conducted a needs assessment among 741 university students about their attitude toward mental health apps. The researchers found that there is interest in mental health apps among students, especially among students who had received professional mental health services in the past 12 months. The primary reasons for interest in mental health apps were confidentiality, convenience, and immediate availability. However, despite interest, actual use was limited, highlighting the importance to understand what factors are associated with adoption and app use.

Perceived Stress

Experiencing stress can be a starting point for students to adopt mental health apps. Perceived stress has been linked with interest in, and acceptance of, stress management apps [21,22]. This link indicates that people with an actual need for support are more likely to accept and use mental health interventions.

Help-Seeking Behavior Among Students

For a person to use a health intervention, there must be a motivation and willingness to use it [23]. The Health Belief Model [23] explains the adoption of health interventions through several constructs related to an individual's beliefs about their condition. It argues that a person's belief in the severity of their illness or health symptoms, and the perceived benefits of seeking treatment for these symptoms, affect their adoption of health interventions. These constructs suggest that the mere presence of symptoms is insufficient: the individual also has to believe that there is benefit in seeking help for these symptoms.

Stigma

Previous studies found that a common barrier among young people and students in seeking traditional mental health resources is stigma associated with mental health [11,24]. A survey study with university students found that, despite overall poor mental well-being across the study sample, there was only a weak relationship between mental well-being and use of mental health resources, and this relationship was mediated by stigma [25]. In other words, if students are in need of help but experience a high level of stigma, that stigma may still prevent them from seeking resources such as apps.

It is argued that services delivered through technology can overcome the barrier of stigma because people can use the technology privately, and other people do not have to know they are seeking help [20]. However, it has not been evaluated if and how perceived stigma is associated with use of apps and

how it may mediate the relationship between stress and mental health app use.

Past Use of Professional Services

Past use of mental health services has also been associated with app use. A person's prior use of mental health services can increase their interest in using mental health technologies [20]. A positive experience with these services can increase the likelihood that people will be open to trying a digital intervention [26-28].

Privacy

An important aspect of mental health apps among university students is confidentiality. Students in some studies thought that mental health apps can facilitate confidentiality [20], and they did not think that privacy would be an issue as long as their data were anonymous [29]. University students in a focus group study, however, expressed concerns that their data would leak and that people such as their lecturers or tutors could access their data [14]. It is therefore expected that privacy concerns can be a challenge or barrier to mental health apps and will negatively affect app use.

Social Influence

Expectations and beliefs about mental health apps by other people close to the user, such as family, friends, or health providers, may also play a role in mental health app use. The Technology Acceptance Model (TAM) [30] includes the importance placed on other people's opinions, labelled social influence, as a predictor of technology adoption. Including social influence as a component of technology adoption is based on the theory of reasoned action (TRA), in which user actions are a result of the user's perception that significant others expect them to conduct these actions [31].

In general, although previous studies have identified what students perceive to be important aspects of mental health apps, it is less understood how these aspects are associated with use. Moreover, most studies included university students, but community college students can experience different mental health issues and barriers to accessing resources, which can affect app use.

Our study aimed to understand how perceived stress, perceived need to seek help, perceived stigma, past use of professional mental health services, privacy concerns, and social influence are associated with mental health app use among community college students. Our findings can be used to guide development and selection of digital mental health interventions that address community college students' needs, and they can inform effective outreach and engagement strategies to engage community college students with these interventions.

Hypotheses

Based on prior studies, the hypotheses of the study are:

1. Perceived stress is positively associated with students' use of mental health apps.
2. The association of perceived stress and use of mental health apps is mediated by a perceived need for help.
3. The association of perceived stress and use of mental health apps is mediated by perceived stigma.
4. Past use of professional mental health services is positively associated with students' use of mental health apps.
5. Privacy concerns are negatively associated with mental health app use.
6. Social influence is positively associated with students' use of mental health apps.

Methods

Study Design

A random sample of 5000 students at a community college in California were selected to participate in this study. To obtain a sample that was balanced for gender and race, sampling was in proportion to the demographics (gender/race) of California community colleges [32]. The demographics for California community colleges are as follows: 26% identify as White, 45% as Hispanic/Latinx, 12% as Asian, and 6% as African American. Approximately half of community college students (54%) identify as female.

Students' email addresses were obtained through the College's registrar's office. These addresses were provided to a staff member at the College, who distributed the invitational emails. The web-based survey took 20-30 minutes to complete. Participants received a US $10 gift card for completing the survey.

Participants

The email was received by 4985 students (15 emails bounced back). A total of 574 participants started the survey, resulting in a response rate of 11.5% (574/4985). A total of 500 participants completed the survey. The survey participants had a mean age of 23.8 years (SD 8.0), and 314/500 (62.8%) identified as female. Of the 500 students, 137 (27.4%) identified as White, 119 (23.8%) identified as Hispanic/Latinx, 66 (13.2%) identified as Asian, and 32 (6.4%) identified as African American. With the exception of the proportion of Hispanic/Latinx participants, our study sample broadly matches the breakdown of demographics for California community colleges.

Most reported their primary language as English (310/500, 62%), while 27% (135/500) reported Spanish as their primary language; the survey was only available in English. Of the 500 participants, 87 (17.4%) had an annual household income of less than US $10,000. The majority of participants (402/500, 80.4%) had health insurance, but only 23.9% (96/402) of those with insurance were sure that their plan provided coverage for mental health services. Further demographic characteristics are shown in Table 1.

Table 1. Demographics of the study participants (N=500).

Demographics	Value[a]
Age (years), mean (SD)	23.8 (8)
Gender[b], n (%)	
Male	171 (34.2)
Female	314 (62.8)
Sexual orientation[c], n (%)	
Heterosexual or straight	391 (78.2)
Bisexual	39 (7.8)
Questioning or unsure of sexual orientation	18 (3.6)
Enrollment status, n (%)	
Full-time	306 (61.2)
Part-time	167 (33.4)
Time of classes, n (%)	
Daytime	377 (75.4)
Evening	83 (16.6)
Employment status, n (%)	
Unemployed	213 (42.6)
Part-time	158 (31.6)
Full-time	63 (12.6)
Retired	5 (1)
Race, n (%)	
White	137 (27.4)
Hispanic/Latino/a/x	119 (23.8)
Asian	66 (13.2)
More than one race	44 (8.8)
Black or African American	32 (6.4)
American Indian or Alaska Native	3 (0.6)
Native Hawaiian or other Pacific Islander	3 (0.6)
Ethnicity, n (%)	
Mexican/Mexican-American/Chicano	183 (36.6)
More than one ethnicity	55 (11)
Asian	52 (10.4)
European	46 (9.2)
Central American	37 (7.4)
African	18 (3.6)
Middle Eastern	13 (2.6)
South American	12 (2.4)
Eastern European	10 (2)
Primary language, n (%)	
English	310 (62)
Spanish	135 (27)
Vietnamese	7 (1.4)

Demographics	Value[a]
Arabic	6 (1.2)
Mandarin	4 (0.8)
Russian	3 (0.6)
Marital status, n (%)	
Single	289 (57.8)
In a committed relationship	146 (29.2)
Married	35 (7)
Divorced or separated	9 (1.8)
Children, n (%)	
Yes	69 (13.8)
No	416 (83.2)
Living situation, n (%)	
Live with family	389 (77.8)
Live with spouse or partner	45 (9)
Live alone	18 (3.6)
Live with roommate(s)	14 (2.8)
Live with children	10 (2)
Homeless, n (%)	
Yes	12 (2.4)
No	472 (94.4)
Household income (US $), n (%)	
<10,000	87 (17.4)
10,0000-29,999	131 (26.2)
30,000-49,999	59 (11.8)
50,000-89,999	52 (10.4)
90,000 or above	49 (9.8)
Disability	
Yes	47 (9.4)
No	426 (85.2)
Veteran	
Yes	8 (1.6)
No	480 (96)
Health insurance	
Yes	402 (80.4)
No	61 (12.2)

[a]Not all respondents answered each question; hence, some percentages do not sum to 100%.

[b]Multiple genders were included as options on the survey; however, participants reported their genders as only male or female.

[c]Other sexual orientations were included as options on the survey; however, the results in the table reflect the participants' responses.

Measures

The complete survey instrument is included in Multimedia Appendix 1.

Barriers to Mental Health Resources, Important Aspects of Mental Health Apps, and Activities People Would Like to Do Using Mental Health Apps

Participants were asked to report on barriers they faced to accessing mental health–related resources, important aspects

about using mental health apps, and what they would like to do using mental health apps. For each of these questions, they were instructed to "select all that apply" from a list of options and/or give an answer in their own words. The list of barrier options was taken from the Healthy Minds Study, an annual web-based survey assessing mental health and service use among students [3].

COVID-19

Survey responses were collected between April 16 and June 30, 2020, during the COVID-19 pandemic. Participants were asked questions related to whether and how COVID-19 had impacted their lives, to gauge changes resulting from the pandemic. These results were not used in the analysis but are reported here to characterize the impact of COVID-19 on the participants. Of the 500 students, 107 (21.4%) knew someone who had been diagnosed with COVID-19. Of these 107 participants, 77 (72.0%) reported that the person diagnosed was a friend or acquaintance; for 37 (34.6%) it was a family member; for 11 (10.3%) it was a colleague; and 4 (3.7%) said they had been diagnosed themselves.

Mental Health

A single dichotomous (yes or no) item asked participants whether they were currently experiencing or had ever experienced a mental illness. In the survey, it was explained that while the term mental illness was used, there are many different terms that can be used, such as mental health problem, emotional distress, psychological disorder, and mental challenge.

Mental Health App Use

A single question was used to identify whether participants had used mental health apps. In this study, a mental health app was defined as "an application on your mobile phone or tablet device that helps you manage your mental, emotional, or psychological health or get access to resources to support your mental, emotional, or psychological health." Participants could respond by indicating whether they had used apps in the past, were currently using apps, had never used apps but would be interested in doing so, or had never used apps and were not interested in doing so. In this paper, those who had used or were currently using mental health apps were defined as users, and those who had never used mental health apps were defined as nonusers.

Past Use of Professional Mental Health Services

Two dichotomous (yes or no) response items from the California Health Interview Survey [33] were used to identify whether participants had sought help from their general practitioner or another professional, such as a counselor, for mental health concerns in the past 12 months.

Perceived Stress

The 7-item version of the College Student Stress Scale [34] was used to assess perceived stress. A higher score on this scale indicates a higher level of stress. The validity and reliability of this scale have been validated with college students (α=.87) [34].

Perceived Need to Seek Help

A dichotomous (yes or no) response item from the California Health Interview Survey [33] was used to identify whether participants felt they may have needed to see a professional because of problems with their mental health in the past 12 months.

Mental Health Concerns

Participants who answered they felt a need to see a professional in the past 12 months were asked to select what mental health concerns, if any, they experienced in these past 12 months. Participants could give an answer in their own words and/or select all options that applied from a list. Examples of mental health concerns were stress, depression, anxiety, loneliness, and concerns related to interpersonal relationships.

Perceived Stigma

The Perceived Stigma subscale of the Depression Stigma Scale [35] was used to assess the perceived stigma participants experience toward mental health. There are multiple types of stigma, and we chose to include this particular type because it can be measured for all participants, regardless of whether they have experienced a mental illness. Participants were asked to rate 9 statements related to perceived stigma (eg, "Most people believe that people with a mental illness could snap out of it if they wanted"), using a Likert scale from 1, strongly disagree, to 4, strongly agree. A higher score indicates greater perceived stigma. Internal consistency of reliability of the Perceived Stigma scale has been validated in prior work (Cronbach α=.82) [35].

Social Influence

Social influence was measured using three statements (eg, "People who are important to me think I should use mental health apps") based on the Unified Theory of Acceptance and Use of Technology (UTAUT) questionnaire [36], which is used to evaluate people's technology acceptance and adoption. Participants were asked to rate these statements using a Likert scale from 1, strongly disagree, to 5, strongly agree. A higher score indicates greater importance placed on other people's expectations.

Privacy Concerns

Privacy concerns were measured using 6 statements (eg, "I feel that as a result of my using mental health apps, others know more about me than I am comfortable with") based on the Scale on Mobile Users' Information Privacy Concerns [37]. The original items were adapted to refer to mental health apps specifically (eg, an original item was "I feel that as a result of my using mobile apps, others know more about me than I am comfortable with"). Participants were asked to rate these statements using a Likert scale from 1, strongly disagree, to 5, strongly agree. A higher score indicates greater concerns. The items were worded slightly differently for participants who answered they had not used mental health apps compared to participants who had used such apps (eg, "I feel that if I were to use mental health apps, others know more about me than I am comfortable with" vs "I feel that as a result of my using

mental health apps, others know more about me than I am comfortable with").

Factor Analysis and Checking for Multicollinearity

The structural model included two latent variables: privacy and social influence. These were included as latent variables because there were multiple survey items related to the concepts of privacy and social influence, and these were adapted from validated scales to refer to mental health apps specifically. Use of the total score of these adapted versions has not been tested in prior work. Combining the related items into a latent variable, rather than treating them as separate variables, reduces the dimensionality of the data.

The variables of perceived stress and perceived stigma were included in the structural model as observed variables. The reliability of using the total score of these scales been tested and validated in prior work, and they have each been used as one aggregated continuous score in previous models [38]. Aggregating these scales as observed variables, rather than latent variables, was therefore deemed appropriate, and this approach makes it easier to interpret the data.

The latent variables were first assessed using confirmatory factor analysis. Two indices were used to assess the fit of the measurement models: the comparative fit index (CFI) and Tucker-Lewis index (TLI) (the threshold value of acceptable model fit for the CFI and TLI is at least 0.90; see [39]). Maximum likelihood was used as an estimator.

Privacy was measured in the survey with 6 questions (see Multimedia Appendix 1). A test of the fit of an initial model including all 6 privacy questions showed incremental fit indices that were below the acceptable threshold of 0.90 (χ^2_{21}=326.712, CFI=0.85, TLI=0.75). A correlation matrix (see Table 2) identified two groups of correlated items, with intercorrelation values >0.7. The privacy construct was therefore split into two separate privacy constructs: one construct related to information being visible by others (Privacy Construct 1) and one construct related to how information is used (Privacy Construct 2). Two measurement models, one with the first 3 privacy questions and one with the remaining 3 privacy questions, did show an acceptable model fit (χ^2_6=162.443, CFI=1.0, TLI=1.0, and χ^2_6=147.822, CFI=1.0, TLI=1.0, respectively). The Cronbach α values for the two scales were .91, indicating good internal consistency.

The social influence construct comprised 3 questions (see Multimedia Appendix 1). Confirmatory factor analysis for the social influence construct also showed an acceptable model fit (χ^2_6=1155.902, CFI=1.0, TLI=1.0). The Cronbach α value for the scale was .95, showing good internal consistency. The questions were therefore deemed suitable to combine into a latent variable for the structural model.

Before the structural model was tested, the variance inflation factors of the model variables were assessed to detect potential multicollinearity. The variance inflation factors of the variables were all under 1.3, indicating no problematic multicollinearity. The correlation matrix of the variables is displayed in Table 3. The correlation values further show that there were no strong correlations (correlation values exceeding 0.70, see [40]) among the constructs.

Table 2. The correlation matrix of privacy items showing two correlated groups of items (Group 1, consisting of items 1, 2 and 3, and Group 2, consisting of items 4, 5, and 6).

	Privacy_1	Privacy_2	Privacy_3	Privacy_4	Privacy_5	Privacy_6
Privacy_1	1					
Privacy_2	0.70	1				
Privacy_3	0.73	0.74	1			
Privacy_4	0.44	0.56	0.61	1		
Privacy_5	0.47	0.56	0.61	0.86	1	
Privacy_6	0.43	0.56	0.62	0.87	0.88	1

Table 3. Correlation matrix of model variables showing no strong correlations between variables.

	Perceived stress	Perceived need to seek hep	Past use of professional mental health services	Stigma	Social influence	Privacy Construct 1	Privacy Construct 2	Age	Gender
Perceived stress	1								
Perceived need to seek help	0.40	1							
Past use of professional mental health services	0.06	0.33	1						
Stigma	0.23	0.20	0.11	1					
Social influence	0.08	0.24	0.09	0.08	1				
Privacy Construct 1	0.17	0.09	0.01	0.21	0.20	1			
Privacy Construct 2	0.17	0.12	–0.01	0.16	0.06	0.64	1		
Age	–0.07	–0.09	–0.03	0.01	–0.05	0.00	0.05	1	
Gender	0.21	0.28	0.11	0.10	0.12	0.05	0.07	0.03	1

Analysis

The dependent variable of the model was the participants' mental health app use as a dichotomous (yes or no) variable. The independent variables were perceived stress, perceived need to seek help (shortened in the *Results* section as perceived need), past use of professional mental health services (shortened in the *Results* section as past use of services), perceived stigma, social influence, and privacy concerns. Age and gender were added as covariables.

To address Hypotheses 1, 4, 5, and 6, the model tests the effects of stress, past use of services, privacy concerns, and social influence on app use. To address Hypotheses 3 and 4, the model tests whether the effect of stress on app use is mediated by perceived need and perceived stigma.

The relationships between measured variables were tested using structural equation modeling (SEM). SEM was used because it allows the flexibility to include privacy concerns and social influence, which were composed of a subselection and/or adapted items from validated scales as unobservable latent variables, and it allows testing for mediation effects of perceived need and perceived stigma [41]. Latent variables are unobserved constructs that are measured by a number of observed variables (ie, items).

Two indices were used to assess the fit of the structural model: the standardized root mean square residual (SRMR) and root mean square error of approximation (RMSEA) (a SRMR value of <0.08 or an RMSEA value of <0.06 indicates that the model fits the data well; see [39]). Maximum likelihood was used as estimator for the structural equation models. Bootstrapping was used to examine the significance of indirect mediating effects [42]. We used bias-corrected bootstrapping with 1000 samples.

We compared the fit of the full mediation model, which tests for mediation effects of perceived need and perceived stigma, with a direct effect model. The direct effect model tested the direct effects of the independent variables on the dependent variable without considering mediating effects.

In this study, the focus was on testing the mediating effect of stigma on app use. It is imaginable that stigma is related to other independent variables—for example, stigma may be related to higher privacy concerns, or a high level of stigma may be negatively related to past use of services. To understand these interrelationships, intercorrelations among independent variables were determined before the structural model was tested.

Full maximum likelihood was used to impute missing data on scales (ie, perceived stress, perceived stigma, social influence, and privacy concerns). Participants with missing data on dichotomous model variables (ie, perceived need and past use of services) were excluded from the model. The software environment R (R Foundation for Statistical Computing) was used for statistical analysis, and the R package lavaan was used for the structural equation models, bootstrapping, and confirmatory factor analysis [43].

Results

Demographic Information

Of the 500 participants, 189 (37.8%) reported that they had experienced a mental illness; 219 (43.8%) of participants reported mental health concerns related to stress. Over half of the participants (262/500, 52.4%) had a stigma score of 21 or higher, suggesting moderate to severe perceived stigma. Of the 500 participants, 106 (21.2%) had used a mental health app. Other descriptive statistics are summarized in Table 4.

Table 4. Overview of responses to survey items (N=500).

Variable	Value
Technology ownership, n (%)	
Smartphone	443 (88.6)
Desktop or laptop computer	463 (92.6)
Tablet	138 (27.6)
Mobile/cell phone but not a smartphone	32 (6.4)
Technology use, n (%)	
Access to Wi-Fi	448 (89.6)
Access to a mobile data plan	440 (88)
Use of internet constantly or many times per day	452 (90.4)
Mental health app use, n (%)	
Current user	34 (6.8)
Past user	72 (14.4)
Nonuser, interested in using apps	199 (39.8)
Nonuser, not interested in using apps	180 (36)
Mental illness (self-reported), n (%)	
Yes	189 (37.8)
No	260 (52)
Prefer not to answer	51 (10.2)
Most common mental health concerns, n (%)	
Stress	219 (43.8)
Anxiety	207 (41.4)
Depression	172 (34.4)
Use of professional services in the past 12 months, n (%)	
Yes	115 (23)
No	374 (74.8)
Perceived need to seek help, n (%)	
Yes	221 (44.2)
No	221 (44.2)
Prefer not to answer	21 (4.2)
Most common barriers to accessing mental health resources, n (%)	
I prefer to deal with issues on my own	260 (52)
Financial reasons (eg, too expensive)	141 (28.2)
I am concerned about privacy	129 (25.8)
I question how serious my needs are	126 (25.2)
I worry what other people will think of me	116 (23.2)
Most important aspects of using mental health apps, n (%)	
The app is free	429 (85.8)
Personal information will be kept private	397 (79.4)
No negative effect on device (eg, drain phone battery)	264 (52.8)
Parts of the app can be used offline	256 (51.2)
People on the app have similar mental health experiences to mine	228 (45.6)
Most common activities participants would like to do using mental health apps	

Variable	Value
Work through negative emotions and thoughts	329 (65.8)
Identify or recognize symptoms	290 (58)
Talk with other people to get/give support	244 (48.8)
Track symptoms	239 (47.8)
Distract myself from negative thoughts or emotions	228 (45.6)
Stress score, mean (SD)	21.9 (5.3)[a]
Stigma score, mean (SD)	22.4 (6.2)[b]
Privacy score, mean (SD)	17.8 (7.6)[c]
Social influence score, mean (SD)	6.6 (3.4)[d]

[a]The score could range from 7 to 35.

[b]The score could range from 4 to 36.

[c]The score could range from 6 to 30.

[d]The score could range from 3 to 15.

Barriers to Mental Health Resources, Important Aspects about Mental Health Apps, and Activities People Would Like to Do With Mental Health Apps

The community college studied provided some mental health resources, such as counseling and workshops on stress management. The most common barrier to accessing mental health resources was that participants preferred to deal with issues on their own (260/500, 52%). Other barriers were concerns about privacy (129/500, 25.8%), questioning how serious their needs were (126/500, 25.2%), and worries about what other people will think of them (116/500, 23.2%). The most important aspect, which was selected by 429 of the 500 participants (85.8%), was that the app was free. Other important aspects were that personal information would be kept private (397/500, 79.4%) and that people on the app had similar mental health experiences to theirs (228/500, 45.6%). In terms of what respondents would like to do with mental health resources, the most common answers were work through negative emotions and thoughts (329/500, 65.8%), identify or recognize symptoms (290/500, 58%), and talk with others to get or give support (244/500, 48.8%).

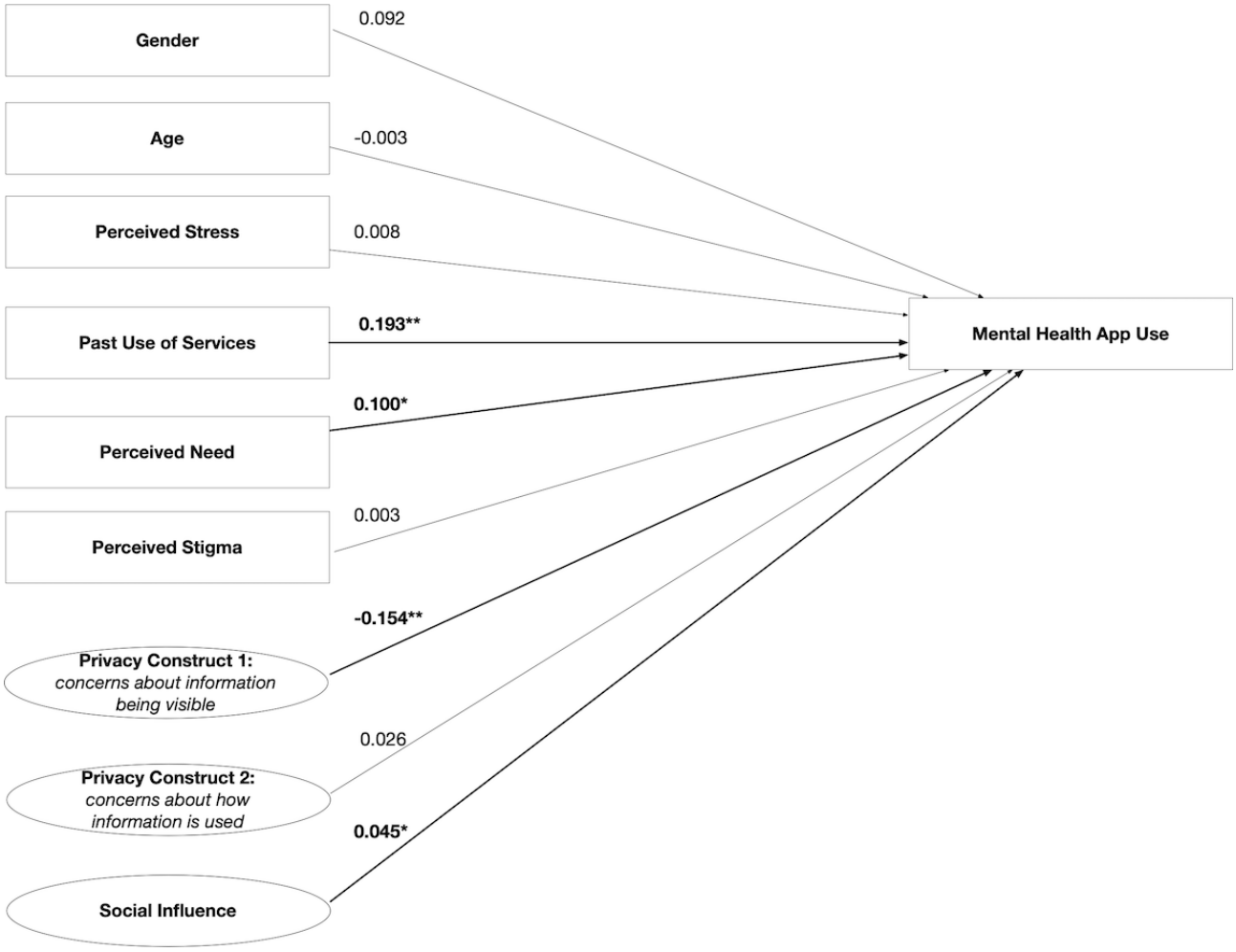
Models to Understand Factors Associated with Mental Health App Use

Not all participants answered each survey question. After imputing missing values on continuous scale variables and excluding participants with missing data on dichotomous model variables (ie, perceived need and past use of services), the structural equation model included a subsample of 449 participants.

Direct Effect Model

The standardized path coefficients of the direct effect model are shown in Figure 1. The model fit indices were χ^2_{87}=222.803, SRMR=0.089, and RMSEA=0.068. The adjusted R^2 value of the model was 0.213.

Figure 1. Direct effect model showing the path coefficients and levels of significance for relationships among variables. *Relationship is significant at $P<.05$; **relationship is significant at $P<.01$; n=449.



Full Mediation Model

The standardized path coefficients of the full mediation model are shown in Figure 2. The model fit indices showed an acceptable model fit: $\chi^2_{89}=72.804$, SRMR=0.051, and RMSEA=0.076. The adjusted R^2 value of the model was 0.326. Table 5 compares the fit indices of the direct effect model and full mediation model, and the data indicate that the full mediation model had the best fit.

The model showed that past use of services, such as a counselor, was significantly associated with mental health app use. Perceived stigma had no effect on mental health app use. Some privacy concerns (ie, Privacy Construct 1) were associated with a lower likelihood of mental health app use. Social influence, measured by the importance participants placed on other people's opinions on mental health app use, was associated with higher mental health app use. The effects of gender and age were nonsignificant.

Figure 2. Full mediation model showing the path coefficients and levels of significance for relationships among variables. *Relationship is significant at *P*<.05; **relationship is significant at *P*<.01; n=449.

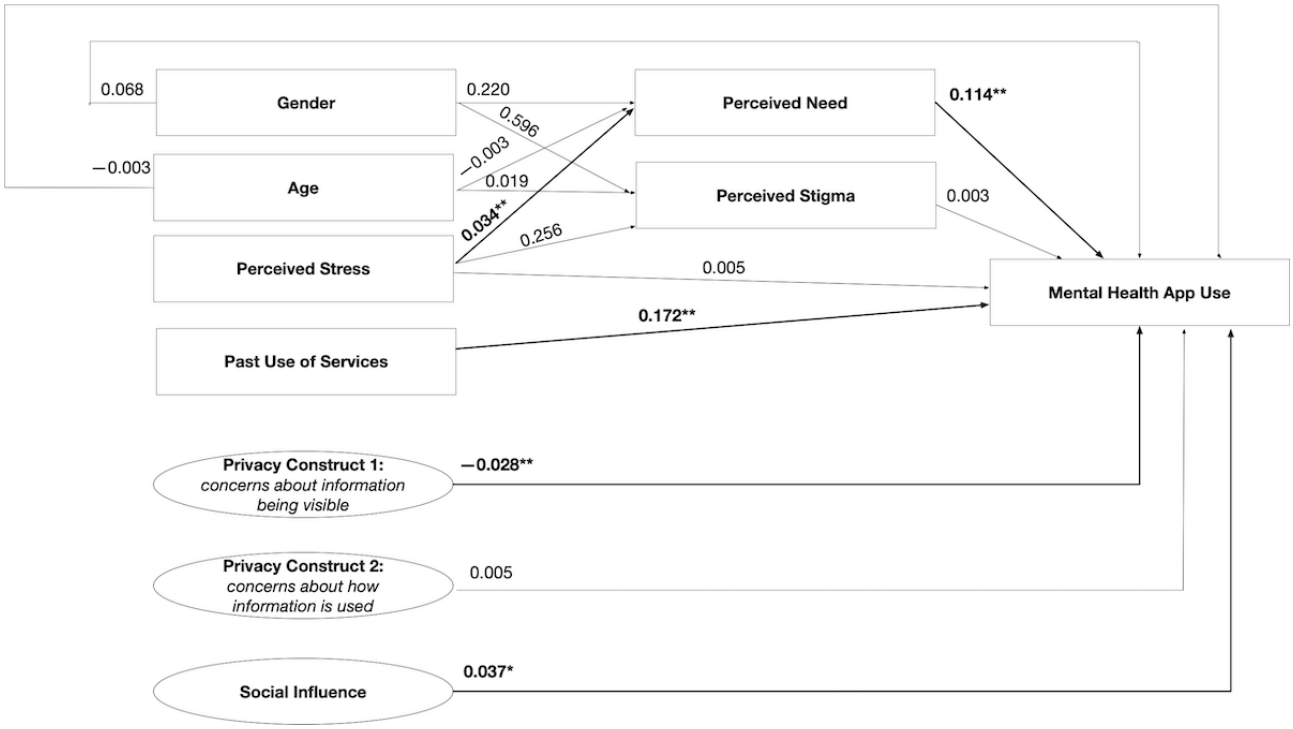


Table 5. Fit indices of the direct effects model and full mediation model.

Statistic	Direct effects model	Full mediation model
Chi square (*df*)	222.803 (87)	72.804 (89)
SRMR[a]	0.089	0.051
RMSEA[b]	0.068	0.076
R^2	.213	.326

[a]SRMR: standardized root mean square residual.

[b]RMSEA: root mean square error of approximation.

The effect of perceived stress on the likelihood of mental health app use was mediated via perceived need. As Figure 2 illustrates, the regression coefficient between perceived stress and perceived need and the regression coefficient between perceived need and mental health app use were significant. The indirect effect was (0.034)*(0.114)=0.004. We used bootstrapping to test the significance of this indirect effect. Perceived need significantly mediated the effect of perceived stress on mental health app use (B=0.004, 95% CI 0.001 to 0.007; *P*=.01). There was no significant mediating effect of perceived stigma on the effect of perceived stress (B=0.001, 95% CI –0.002 to 0.011; *P*=.47).

Discussion

Principal Findings

The aim of this paper was to identify factors associated with mental health app use among community college students. The results revealed that participants' use of mental health apps was associated with 5 factors: perceived stress, perceived need to seek help for mental health concerns, past use of professional mental health services, social influence of other people, and privacy concerns. Furthermore, the effect of stress was mediated by a perceived need to seek help. These findings support Hypotheses 1, 2, 4, and 6. Hypothesis 5 is partly supported, as only specific types of privacy concerns were negatively associated with app use. Hypothesis 3, which stated that the association of perceived stress and app use would be mediated by perceived stigma, was not supported by the results.

Perceived Stress

Stress was listed as the most common mental health concern among participants (approximately 44% of participants experienced stress), and the results of the study indicate that the level of perceived stress predicted mental health app use. This means that the more stress people experienced, the more likely they were to have used apps. App use could therefore have been motivated by a student's own perceived benefit of the apps in terms of their ability to reduce stress. Previous work has indicated a link between self-reported stress and interest in using a stress management app [21], as well as a link between perceived stress and a preference for technology over face-to-face mental health services [44]. Apolinário-Hagen et al [44] argue that people with stressful lives may find it easier to fit the use of a technology into their life than an in-person

appointment, although further work is needed to test this assumption.

Perceived Need to Seek Help

The results confirmed an association between a perceived need to seek help and app use, which extends previous work by showing that perceived need is not only a predictor of help-seeking through *nondigital* channels [3,23] but is also an important driver in using *digital* health interventions. The association indicates that students are more likely to have used mental health apps if they felt a need to seek help for concerns related to their mental health. In the current study, the effect of stress was mediated by a perceived need to seek help, which means that the effect of stress on mental health app use is even stronger if people not only experience stress, but also have a perceived need to seek help for issues related to mental health.

Past Use of Professional Mental Health Services

Of the participants in our study, 23% had used professional mental health services in the past 12 months, which is broadly proportionate to the general community college population: a recent study analyzing survey data from 23 community colleges between 2016 and 2019 found that 30% of community college students had used therapy in the past year [45]. The community college studied also provided some on-campus mental health resources, such as counseling. Although the level of on-campus mental health support can vary per college, 73% of community colleges provide mental health counseling services on campus [46]. However, a previous study found that only a small percentage (5.4%) of community college students that seek help do so on campus [45]. This suggests that although offering resources is essential, it is imperative to understand potential barriers in accessing these resources. For example, campus counseling services often cannot keep up with the demand, which can cause students to feel undersupported [47,48].

In our study, past use of services was positively associated with mental health app use. For students that are already familiar with seeking help through nondigital channels, it may be less of a step to seek out help through a mental health app. However, it will likely matter how they experienced using mental health services, as people with a positive experience are generally more open to trying new mental health technologies [26-28], but people with a negative experience are less interested in mental health technology [49-51]. Future research should clarify the association between past use of professional mental health services and subsequent app use by assessing the quality of the previous experience with in-person services.

Social Influence

The social influence construct was positively associated with mental health app use, which implies that if people close to the participants thought they should use mental health apps, they were more likely to use such apps. Technology use that is influenced by other people's expectations of using it can create a sense of belonging [52]. Furthermore, people's expectations could have been interpreted as them viewing the app as useful, as people are persuaded by messages from people close to them [35].

The effect of social influence in our model is in line with previous studies on mental health technologies, where engagement with a digital mental health program was facilitated by whether it was endorsed by friends and family [53] or the user's current health care provider [54]. A survey among 102 university students and nonstudents found that one of the main reasons that participants used mental health technologies was that someone they knew recommended their use [55].

However, people can be deterred from using mental health technologies if they feel forced to use it by others, for example, if they feel pressure to use certain technologies from their health care provider [56]. It is therefore important that social influence happens organically through endorsement and recommendations rather than being prescribed top-down. Furthermore, a previous study that used a structural equation model to explain use of physical health apps found that social influence contributes to a decision to *initially* use an app but not to the intention to *continue* to use an app [57].

Privacy

In our study, privacy played an important role in mental health app use, as supported by both the model and by privacy concerns being one of the most important aspects of and barriers to app use. Privacy of information is known to be an important aspect for students, with concerns about confidentiality of information [20] and data being leaked beyond the app [14]. A recent study reviewing 116 publicly available mobile apps for depression found that only a minority of these apps had a privacy policy and that most privacy policies were not transparent about how user data were collected, stored, and used [58]. A previous literature review of user engagement with mental health apps theorized that one reason for low user engagement with mental health apps is that many apps do not consider user privacy [59].

However, interestingly, only some aspects of privacy were significantly associated with app use in our model. Concerns related to personal information being potentially accessible by others was negatively associated with app use, meaning that people who had concerns about this aspect were less likely to use apps.

Privacy concerns about how personal information was subsequently used did not have an effect on app use. Privacy as a construct may be composed of different aspects that can impact mental health app use differently, which would be worthwhile to explore and refine further in future studies.

Stigma

Consistent with a previous study among college students [3], the most common barrier to accessing mental health resources was that students preferred to deal with issues on their own, which may reflect stigma surrounding seeking help. Although over half of the participants had moderate to severe perceived stigma, this was not associated with mental health app use. Furthermore, our study did not find that the effect of stress on app use was mediated by stigma.

These findings suggest that whereas stigma can form a barrier to mental health help seeking in general [11,25], it may not bar people from using mental health apps in particular. In a previous

study, participants who preferred digital over face-to-face mental health services even had higher levels of stigma [60]. When given the choice between digital and in-person services, digital resources may be a less stigmatizing option, although the current study did not provide support that a high level of stigma is necessarily associated with app use. Furthermore, our model included perceived stigma; however, stigma is multifaceted, and there are multiple other types of mental health stigma, such as internalized stigma and experienced stigma [61], that may affect use of mental health resources.

Barriers to Using Mental Health Resources and Important Aspects of Mental Health Apps

Previous studies have highlighted potential differences in mental health concerns and needs among university and community college students. For example, whereas high workload and academic stress is a common issue raised among university students, financial or housing stress may be more common among community college students [9,10]. A previous study found that some community college students are uninsured and use web-based health information to avoid medical costs [10]. Our results show that the most important aspect of a mental health app was that it was free (85.8% of participants listed this as an important aspect). Furthermore, though the majority of participants (80.4%) had health insurance, only 24% were sure that their plan provided coverage for mental health services. It is thus important to address costs when considering mental health apps for community colleges.

Previous focus group studies have found differences in how technology is used, or would be used, by university and community college students for supporting their health. Whereas university students primarily use digital health resources in a preventive manner to recognize onset of symptoms [14], community college students tend to use digital health resources in a reactive manner when they are already experiencing symptoms [10]. Although these prior studies used small sample sizes and caution should be used in generalizing findings, they show there are different ways in which technologies may be used by different types of students. In our study, two-thirds of participants wanted to use mental health resources to work through negative emotions and thoughts, which further suggests a need for help with symptoms they are already experiencing.

Implications

Previous work has shown that despite interest in mental health apps among students, use of these apps can be limited [20]. Our findings identify several factors associated with mental health app use among community college students. The results of this study can inform outreach and implementation of digital mental health resources on campuses. Below, we outline several implications that may be important to consider when implementing digital mental health resources for a community college student population.

Consider Factors That May Influence Engagement With Tools

Concerns about privacy were negatively associated with app use, and common barriers to accessing resources were related to costs and to students' preference to deal with issues on their own. These barriers should be considered when offering digital resources to students. It is important to be transparent to students who will have access to their information. Furthermore, it is important to consider costs and, if possible, offer resources that are either low-cost or free to students. In addition, marketing mental health apps as self-guided tools that students can use to deal with issues on their own, rather than as a help-seeking service, may further facilitate use of mental health apps.

Mental Health Needs of Community College Students

It is important to think through specific preferences for accessing and integrating mental health support. Students who experienced stress, recognized a need to seek help, and had sought out help in the past were more likely to have used mental health apps. Counseling services can promote use of mental health apps among students already seeking help. A strength of technology is its availability [44] and the ability to access it anytime. Counselors can endorse particular apps to students so they can continue to have access to resources in addition to and after finishing counseling.

Endorsement and Use by Others May Help Mental Health App Adoption

For students who do not actively seek out help, it may not be sufficient to promote an app top-down through formal mental health services, although social influence of other people was positively associated with mental health app use. If other people have positive views about apps or personal experience with them, their endorsement may encourage other students to try an app. For outreach of apps, social platforms can be used to introduce apps, as seeing others sharing their experience and promoting an app may increase interest among students. In addition, the use of apps may be introduced and piloted in classrooms or other group settings, as the effect of social influence is strengthened if a technology is used in public rather than in private [62], and use by others may motivate students to use an app.

Limitations and Future Work

This study has a number of limitations. First, the sample came from one community college, so caution must be taken in generalizing the results. We expect our findings to generalize to college student populations with similar demographics. Second, the results were collected during the COVID-19 pandemic and 1 month after the stay-at-home order took effect. Although we do not expect these circumstances to have influenced the relationships between stress and app use, it may have increased mental health concerns and increased interest in mental health apps. Given the cross-sectional nature of the survey, we do not have data available on changes in stress before and during the pandemic. Additionally, we focused on presenting and discussing a specific model to test the effects of 6 factors on mental health app use. There may be other mechanisms affecting app use, such as past experience with technology, which merit further exploration. To assess the factors of mental health app use and perceived need to seek help, single-item validated questions from prior work were used. Using short single item measures helps reduce participant burden of answering a long survey; however, it narrows the distribution

of such variables. Lastly, the response rate was 12%; although this is a typical response rate for web-based surveys among student populations [2,5,20], there is potential for response bias, and there may be differences between nonresponders and responders. For example, although not all participants reported their race or ethnicity, the proportion of Hispanic/Latinx participants in our study sample was lower than that in the California community college population. Although classes at this community college are only taught in English, the English language questionnaire could have been a barrier to participation.

Conclusions

This study focused on community college students and found 5 factors associated with mental health app use. Perceived stress, perceived need to seek help, and past use of professional services were positively associated with mental health app use. The more participants agreed that people in their social environment thought they should use apps, the more likely they were to use apps. Some privacy concerns were negatively associated with mental health app use. In addition, financial costs are an important aspect to consider for using mental health apps. The results can inform the selection and appropriate dissemination of mental health apps to meet college student needs.

Acknowledgments

This work was funded by Help@Hand (contract number 417-ITSUCI-2019), a project overseen by the California Mental Health Service Authority (CalMHSA). CalMHSA reviewed the manuscript for confidentiality. The information or content and conclusions presented here are those of the authors, and they should not be construed as official position or policy of, nor should any endorsements be inferred by, the participating Help@Hand Counties and/or CalMHSA.

Conflicts of Interest

SMS has received consulting payments from Otsuka Pharmaceuticals for work unrelated to this manuscript and is on the Scientific Advisory Board for Headspace for which he receives compensation. SMS receives funding from One Mind for the operation and management of One Mind PsyberGuide.

Multimedia Appendix 1
Survey items.
[DOCX File , 58 KB - jmir_v23i9e27745_app1.docx]

References

1. Auerbach RP, Alonso J, Axinn WG, Cuijpers P, Ebert DD, Green JG, et al. Mental disorders among college students in the World Health Organization World Mental Health Surveys. Psychol Med 2016 Oct;46(14):2955-2970 [FREE Full text] [doi: 10.1017/S0033291716001665] [Medline: 27484622]
2. Levin ME, Stocke K, Pierce B, Levin C. Do College Students Use Online Self-Help? A Survey of Intentions and Use of Mental Health Resources. Journal of College Student Psychotherapy 2017 Oct 09;32(3):181-198. [doi: 10.1080/87568225.2017.1366283]
3. Lipson SK, Kern A, Eisenberg D, Breland-Noble AM. Mental health disparities among college students of color. J Adolesc Health 2018 Sep;63(3):348-356. [doi: 10.1016/j.jadohealth.2018.04.014] [Medline: 30237000]
4. Eisenberg D, Goldrick-Rab S, Lipson S, Broton K. Too Distressed to Learn? Mental Health Among Community College Students. Wisconsin HOPE Lab. 2016. URL: https://hope4college.com/wp-content/uploads/2018/09/Wisconsin_HOPE_Lab-Too_Distressed_To_Learn.pdf [accessed 2021-09-02]
5. Goldrick-Rab S, Baker-Smith C, Coca V, Looker E. California Community Colleges #RealCollege Survey. The Hope Center. 2019 Mar. URL: https://hope4college.com/wp-content/uploads/2019/03/RealCollege-CCCCO-Report.pdf [accessed 2021-09-02]
6. Bruening M, van Woerden I, Todd M, Laska MN. Hungry to learn: the prevalence and effects of food insecurity on health behaviors and outcomes over time among a diverse sample of university freshmen. Int J Behav Nutr Phys Act 2018 Jan 18;15(1):9-10 [FREE Full text] [doi: 10.1186/s12966-018-0647-7] [Medline: 29347963]
7. Elzein A, Shelnutt K, Colby S, Olfert M, Kattelmann K, Brown O, et al. The prevalence of food insecurity and its association with health and academic outcomes among college freshmen. Adv Nutr 2017;8(1):4. [doi: 10.1093/advances/8.1.4]
8. McBride PE. Addressing the lack of mental health services for at-risk students at a two-year community college: a contemporary review. Community Coll J Res Pract 2017 Dec 08;43(2):146-148. [doi: 10.1080/10668926.2017.1409670]
9. Katz DS, Davison K. Community college student mental health. Community Coll Rev 2014 May 28;42(4):307-326. [doi: 10.1177/0091552114535466]
10. Rennis L, McNamara G, Siedel E, Shneyderman Y. Google it!: Urban community college students' use of the internet to obtain self-care and personal health information. Coll Student J 2015;49(3):414-426.

11. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
12. Bakker D, Rickard N. Engagement in mobile phone app for self-monitoring of emotional wellbeing predicts changes in mental health: MoodPrism. J Affect Disord 2018 Feb;227:432-442. [doi: 10.1016/j.jad.2017.11.016] [Medline: 29154165]
13. Kelley C, Lee B, Wilcox L. Self-tracking for mental wellness: understanding expert perspectives and student experiences. Proc SIGCHI Conf Hum Factor Comput Syst 2017 May 02;2017:629-641 [FREE Full text] [doi: 10.1145/3025453.3025750] [Medline: 28920106]
14. Rooksby J, Morrison A, Murray-Rust D. Student perspectives on digital phenotyping: the acceptability of using smartphone data to assess mental health. In: CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: The 2019 CHI Conference on Human Factors in Computing Systems; May 4, 2019; Glasgow, UK p. A. [doi: 10.1145/3290605.3300655]
15. Torous J, Firth J, Huckvale K, Larsen ME, Cosco TD, Carney R, et al. The emerging imperative for a consensus approach toward the rating and clinical recommendation of mental health apps. J Nerv Ment Dis 2018 Aug;206(8):662-666. [doi: 10.1097/NMD.0000000000000864] [Medline: 30020203]
16. Mobile fact sheet. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-09-01]
17. Montagni I, Cariou T, Feuillet T, Langlois E, Tzourio C. Exploring digital health use and opinions of university students: field survey study. JMIR mHealth uHealth 2018 Mar 15;6(3):e65-e13 [FREE Full text] [doi: 10.2196/mhealth.9131] [Medline: 29549071]
18. Rickwood DJ, Deane FP, Wilson CJ. When and how do young people seek professional help for mental health problems? Med J Aust 2007 Oct 01;187(S7):S35-S39. [doi: 10.5694/j.1326-5377.2007.tb01334.x] [Medline: 17908023]
19. IPEDS Fall 2018 Enrollment Survey. American Association of Community Colleges. 2020. URL: https://www.aacc.nche.edu/wp-content/uploads/2020/03/AACC_Fast_Facts_2020_Final.pdf [accessed 2021-09-01]
20. Kern A, Hong V, Song J, Lipson SK, Eisenberg D. Mental health apps in a college setting: openness, usage, and attitudes. Mhealth 2018 Jun;4:20 [FREE Full text] [doi: 10.21037/mhealth.2018.06.01] [Medline: 30050916]
21. Ervasti M, Kallio J, Määttänen I, Mäntyjärvi J, Jokela M. Influence of personality and differences in stress processing among Finnish students on interest to use a mobile stress management app: survey study. JMIR Ment Health 2019 May 13;6(5):e10039 [FREE Full text] [doi: 10.2196/10039] [Medline: 31094358]
22. Apolinário-Hagen J, Hennemann S, Fritsche L, Drüge M, Breil B. Determinant factors of public acceptance of stress management apps: survey study. JMIR Ment Health 2019 Nov 07;6(11):e15373 [FREE Full text] [doi: 10.2196/15373] [Medline: 31697243]
23. Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984 Sep 04;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
24. College Students Speak: A Survey Report on Mental Health. National Alliance on Mental Illness. Arlington, VA; 2012. URL: https://www.nami.org/Support-Education/Publications-Reports/Survey-Reports/College-Students-Speak_A-Survey-Report-on-Mental-H [accessed 2021-09-01]
25. DeBate RD, Gatto A, Rafal G. The effects of stigma on determinants of mental health help-seeking behaviors among male college students: an application of the information-motivation-behavioral skills model. Am J Mens Health 2018 Sep;12(5):1286-1296 [FREE Full text] [doi: 10.1177/1557988318773656] [Medline: 29749301]
26. Berry N, Lobban F, Bucci S. A qualitative exploration of service user views about using digital health interventions for self-management in severe mental health problems. BMC Psychiatry 2019 Jan 21;19(1):35 [FREE Full text] [doi: 10.1186/s12888-018-1979-1] [Medline: 30665384]
27. Gunn J, Cameron J, Densley K, Davidson S, Fletcher S, Palmer V, et al. Uptake of mental health websites in primary care: insights from an Australian longitudinal cohort study of depression. Patient Educ Couns 2018 Jan;101(1):105-112. [doi: 10.1016/j.pec.2017.07.008] [Medline: 28739180]
28. March S, Day J, Ritchie G, Rowe A, Gough J, Hall T, et al. Attitudes toward e-mental health services in a community sample of adults: online survey. J Med Internet Res 2018 Feb 19;20(2):e59 [FREE Full text] [doi: 10.2196/jmir.9109] [Medline: 29459357]
29. Meng J, Hussain SA, Mohr DC, Czerwinski M, Zhang M. Exploring user needs for a mobile behavioral-sensing technology for depression management: qualitative study. J Med Internet Res 2018 Jul 17;20(7):e10139 [FREE Full text] [doi: 10.2196/10139] [Medline: 30021710]
30. Venkatesh V, Davis F. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
31. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
32. Management Information Systems Data Mart. California Community Colleges Chancellor's Office. URL: https://datamart.cccco.edu/Default.aspx [accessed 2021-08-28]

33. CHIS 2019 adult public use files. UCLA Center for Health Policy Research. URL: https://healthpolicy.ucla.edu/chis/design/Pages/questionnairesEnglish.aspx [accessed 2021-09-01]
34. Feldt RC. Development of a brief measure of college stress: the college student stress scale. Psychol Rep 2008 Jun 01;102(3):855-860. [doi: 10.2466/pr0.102.3.855-860] [Medline: 18763455]
35. Griffiths KM, Christensen H, Jorm AF, Evans K, Groves C. Effect of web-based depression literacy and cognitive-behavioural therapy interventions on stigmatising attitudes to depression: randomised controlled trial. Br J Psychiatry 2004 Oct;185:342-349. [doi: 10.1192/bjp.185.4.342] [Medline: 15458995]
36. Venkatesh, Thong, Xu. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157-178. [doi: 10.2307/41410412]
37. Anwar N, Mohamed Shuhidan S, Zaini M. Measuring users concerns for information privacy. J Phys: Conf Ser 2020 Jun 06;1496:012009. [doi: 10.1088/1742-6596/1496/1/012009]
38. Zeng C, Li L, Hong YA, Zhang H, Babbitt AW, Liu C, et al. A structural equation model of perceived and internalized stigma, depression, and suicidal status among people living with HIV/AIDS. BMC Public Health 2018 Jan 15;18(1):138 [FREE Full text] [doi: 10.1186/s12889-018-5053-1] [Medline: 29334959]
39. Schreiber JB, Nora A, Stage FK, Barlow EA, King J. Reporting structural equation modeling and confirmatory factor analysis results: a review. J Educ Res 2010 Aug 07;99(6):323-338. [doi: 10.3200/joer.99.6.323-338]
40. Burns RP, Burns R. Business research methods and statistics using SPSS. London, UK: Sage; 2008.
41. Chin W. Commentary: issues and opinion on structural equation modeling. MIS Q 1998;22(1):7-16.
42. Cheung MW. Comparison of approaches to constructing Confidence Intervals for Mediating Effects Using Structural Equation Models. Struct Equ Model 2007 May;14(2):227-246. [doi: 10.1080/10705510709336745]
43. Rosseel Y. lavaan: an R package for structural equation modeling. J Stat Soft 2012;48(2):1-36. [doi: 10.18637/jss.v048.i02]
44. Apolinário-Hagen J, Harrer M, Kählke F, Fritsche L, Salewski C, Ebert DD. Public attitudes toward guided internet-based therapies: web-based survey study. JMIR Ment Health 2018 May 15;5(2):e10735 [FREE Full text] [doi: 10.2196/10735] [Medline: 29764797]
45. Lipson SK, Phillips MV, Winquist N, Eisenberg D, Lattie EG. Mental health conditions among community college students: a national study of prevalence and use of treatment services. Psychiatr Serv 2021 Mar 04:appips202000437. [doi: 10.1176/appi.ps.202000437] [Medline: 33657842]
46. Community College Task Force Survey of Community/2 Year College Counseling Services. American College Counseling Association. 2013. URL: https://www.sbctc.edu/resources/documents/about/task-forces-work-groups/ctc-counselors/acca-national-survey-of-community-college-counseling-services.pdf [accessed 2021-09-01]
47. Billingsley M. More than 80% of medical students with mental health issues feel under-supported, says Student BMJ survey. BMJ. 2015 Sep 01. URL: https://www.bmj.com/content/351/sbmj.h4521.full [accessed 2020-12-07]
48. Xiao H, Carney DM, Youn SJ, Janis RA, Castonguay LG, Hayes JA, et al. Are we in crisis? National mental health and treatment trends in college counseling centers. Psychol Serv 2017 Nov;14(4):407-415. [doi: 10.1037/ser0000130] [Medline: 29120199]
49. Allan S, Bradstreet S, Mcleod H, Farhall J, Lambrou M, Gleeson J, EMPOWER Group, et al. Developing a hypothetical implementation framework of expectations for monitoring early signs of psychosis relapse using a mobile app: qualitative study. J Med Internet Res 2019 Oct 24;21(10):e14366 [FREE Full text] [doi: 10.2196/14366] [Medline: 31651400]
50. Bucci S, Berry N, Morris R, Berry K, Haddock G, Lewis S, et al. "They are not hard-to-reach clients. We have just got hard-to-reach services." Staff views of digital health tools in specialist mental health services. Front Psychiatry 2019;10:344 [FREE Full text] [doi: 10.3389/fpsyt.2019.00344] [Medline: 31133906]
51. Watson H, Levine M, Zerwas S, Hamer R, Crosby R, Sprecher C, et al. Predictors of dropout in face-to-face and internet-based cognitive-behavioral therapy for bulimia nervosa in a randomized controlled trial. Int J Eat Disord 2017 May;50(5):569-577 [FREE Full text] [doi: 10.1002/eat.22644] [Medline: 27862108]
52. Schepers J, Wetzels M. A meta-analysis of the technology acceptance model: investigating subjective norm and moderation effects. Inf Manag 2007 Jan;44(1):90-103. [doi: 10.1016/j.im.2006.10.007]
53. Al-Asadi AM, Klein B, Meyer D. Posttreatment attrition and its predictors, attrition bias, and treatment efficacy of the anxiety online programs. J Med Internet Res 2014 Oct 14;16(10):e232 [FREE Full text] [doi: 10.2196/jmir.3513] [Medline: 25316533]
54. Lipschitz J, Miller CJ, Hogan TP, Burdick KE, Lippin-Foster R, Simon SR, et al. Adoption of mobile apps for depression and anxiety: cross-sectional survey study on patient interest and barriers to engagement. JMIR Ment Health 2019 Jan 25;6(1):e11334 [FREE Full text] [doi: 10.2196/11334] [Medline: 30681968]
55. Stawarz K, Preist C, Coyle D. Use of smartphone apps, social media, and web-based resources to support mental health and well-being: online survey. JMIR Ment Health 2019 Jul 12;6(7):e12546 [FREE Full text] [doi: 10.2196/12546] [Medline: 31301126]
56. Feijt MA, de Kort YA, Bongers IM, IJsselsteijn WA. Perceived drivers and barriers to the adoption of eMental health by psychologists: the construction of the levels of adoption of eMental health model. J Med Internet Res 2018 Apr 24;20(4):e153 [FREE Full text] [doi: 10.2196/jmir.9485] [Medline: 29691215]

57. Beldad AD, Hegner SM. Expanding the Technology Acceptance Model with the inclusion of trust, social influence, and health valuation to determine the predictors of German users' willingness to continue using a fitness app: a structural equation modeling approach. Int J Hum-Comput Int 2017 Nov 30;34(9):882-893. [doi: 10.1080/10447318.2017.1403220]
58. O'Loughlin K, Neary M, Adkins EC, Schueller SM. Reviewing the data security and privacy policies of mobile apps for depression. Internet Interv 2019 Mar;15:110-115 [FREE Full text] [doi: 10.1016/j.invent.2018.12.001] [Medline: 30792962]
59. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug 05;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
60. Klein B, Cook S. Preferences for e-mental health services amongst an online Australian sample. EJAP 2010 May 25;6(1):45-68. [doi: 10.7790/ejap.v6i1.184]
61. Corrigan P, Kosyluk K. Mental illness stigma: types, constructs, and vehicles for change. In: Corrigan P, editor. The Stigma of Disease and Disability: Understanding Causes and Overcoming Injustices. Washington, DC: American Psychological Association; 2014:35-56.
62. Kulviwat S, Bruner G, Al-Shuridah O. The role of social influence on adoption of high tech innovations: the moderating effect of public/private consumption. J Bus Res 2009 Jul;62(7):706-712. [doi: 10.1016/j.jbusres.2007.04.014]

Abbreviations

CalMHSA: California Mental Health Service Authority
CFI: comparative fit index
RMSEA: root mean square error of approximation
SEM: structural equation modeling
SRMR: standardized root mean square residual
TAM: Technology Acceptance Model
TLI: Tucker-Lewis index
TRA: theory of reasoned action
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by R Kukafka; submitted 04.02.21; peer-reviewed by E Lattie, CY Lin; comments to author 08.03.21; revised version received 06.05.21; accepted 27.07.21; published 14.09.21.

Please cite as:
Borghouts J, Eikey EV, Mark G, De Leon C, Schueller SM, Schneider M, Stadnick N, Zheng K, Mukamel DB, Sorkin DH
Understanding Mental Health App Use Among Community College Students: Web-Based Survey Study
J Med Internet Res 2021;23(9):e27745
URL: https://www.jmir.org/2021/9/e27745
doi:10.2196/27745
PMID:34519668

Original Paper

Mental Health Screening in General Practices as a Means for Enhancing Uptake of Digital Mental Health Interventions: Observational Cohort Study

Alexis E Whitton[1,2], PhD; Rebecca Hardy[1], PhD; Kate Cope[1], BPsychSc; Chilin Gieng[1], BDES; Leanne Gow[1], Grad Cert; Andrew MacKinnon[1], PhD; Nyree Gale[1], MPsych; Kathleen O'Moore[1], PhD; Josephine Anderson[1,2], MMed; Judith Proudfoot[1,2], PhD; Nicole Cockayne[1], MPH; Bridianne O'Dea[1,2], PhD; Helen Christensen[1,2], PhD; Jill Maree Newby[1,2], PhD

[1]Black Dog Institute, Randwick, Australia

[2]University of New South Wales, Randwick, Australia

Corresponding Author:
Alexis E Whitton, PhD
Black Dog Institute
Hospital Road
Randwick, 2031
Australia
Phone: 61 293828507
Fax: 61 293828507
Email: a.whitton@unsw.edu.au

Abstract

Background: Digital mental health interventions stand to play a critical role in managing the mental health impact of the COVID-19 pandemic. Thus, enhancing their uptake is a key priority. General practitioners (GPs) are well positioned to facilitate access to digital interventions, but tools that assist GPs in identifying suitable patients are lacking.

Objective: This study aims to evaluate the suitability of a web-based mental health screening and treatment recommendation tool (*StepCare*) for improving the identification of anxiety and depression in general practice and, subsequently, uptake of digital mental health interventions.

Methods: StepCare screens patients for symptoms of depression (9-item Patient Health Questionnaire) and anxiety (7-item Generalized Anxiety Disorder scale) in the GP waiting room. It provides GPs with stepped treatment recommendations that include digital mental health interventions for patients with mild to moderate symptoms. Patients (N=5138) from 85 general practices across Australia were invited to participate in screening.

Results: Screening identified depressive or anxious symptoms in 43.09% (1428/3314) of patients (one-quarter were previously unidentified or untreated). The majority (300/335, 89.6%) of previously unidentified or untreated patients had mild to moderate symptoms and were candidates for digital mental health interventions. Although less than half were prescribed a digital intervention by their GP, when a digital intervention was prescribed, more than two-thirds of patients reported using it.

Conclusions: Implementing web-based mental health screening in general practices can provide important opportunities for GPs to improve the identification of symptoms of mental illness and increase patient access to digital mental health interventions. Although GPs prescribed digital interventions less frequently than in-person psychotherapy or medication, the promising rates of uptake by GP-referred patients suggest that GPs can play a critical role in championing digital interventions and maximizing the associated benefits.

(J Med Internet Res 2021;23(9):e28369) doi:10.2196/28369

KEYWORDS

depression; anxiety; general practice; screening; digital mental health

Introduction

Background

Anxiety and depression are leading causes of disability worldwide, and in the wake of the COVID-19 pandemic, the personal, social, and economic costs of these conditions are expected to rise [1]. Although effective interventions exist, these conditions often go untreated [2]. This poses a significant problem for individuals and health care systems, as anxiety and depression are associated with reduced quality of life [3], increased suicide risk [4], and increased health service use [5]. Although increasing access to evidence-based mental health care seems to be an obvious solution, achieving this in the context of an already overstretched mental health system represents a major challenge.

Digital mental health interventions (eg, web-based cognitive behavioral therapy), used alone or in blended models of care, have been highlighted as crucial tools for addressing unmet mental health treatment needs [1]. Indeed, health policies in the United Kingdom [6] and Australia [7] recommend digital mental health interventions as the first-line treatment for individuals with mild or subsyndromal depression. These interventions have been shown to be as effective as in-person psychotherapy for a range of disorders [8] and can be delivered with minimal or no therapist input [9]. Furthermore, these interventions can promote improvements in symptoms that are sustained over several years [10], provide immediate support (ie, no waiting time), and are free from many barriers that prevent individuals from accessing face-to-face treatment (eg, cost, inconvenience, and stigmatization). However, the integration of digital mental health interventions into routine mental health care has been slow [11]. New strategies are needed to promote a greater uptake of digital mental health tools.

One way to increase the integration of digital mental health interventions into routine mental health care is to better support general practitioners (GPs) in prescribing these interventions to their patients. Digital mental health interventions can provide significant benefits in the general practice setting because GPs encounter high rates of common mental health conditions, with 1 in 4 patients experiencing depression [12] and 1 in 5 experiencing anxiety [13]. Furthermore, GPs are often an individual's first point of contact with mental health services and are a trusted source of information for patients, so they are well placed to facilitate greater access to, and confidence in, digital tools for mental health. However, two major challenges have prevented the adoption of digital mental health tools in general practice. First, although mental health conditions are common among general practice attendees, only half of all patients with a mental health condition are recognized by their GP as having one [14]. Accordingly, low rates of disorder detection may prevent GPs from identifying patients who are suitable candidates for digital mental health interventions. Second, given the broad array of digital apps and tools marketed for mental health, GPs report a lack of confidence in prescribing digital interventions and uncertainty about their evidence base inhibits their routine use [15].

To address these challenges, the Black Dog Institute developed StepCare, a web-based tool for use in general practice that offers digital mental health screening and treatment recommendations [16,17]. StepCare aims to facilitate the delivery of a stepped care approach to mental health care in the general practice setting. It does so by screening patients in the GP waiting area using a tablet device and then stratifying patients into groups according to symptom severity. StepCare then provides a screening report and set of treatment recommendations to GPs, which are matched in intensity to a patient's specific level of need. Importantly, evidence-based digital mental health interventions are incorporated into the treatment recommendations for patients who screen positive for mild or moderate depressive or anxious symptoms, and GPs are supported in referring their patients to these interventions via referral links embedded in the program.

Initial studies of feasibility [16] and implementation [17] demonstrated that StepCare was acceptable to GPs and feasible for delivery in general practice settings. Moreover, two recent randomized controlled trials (RCTs) of similar digital patient stratification tools further support the utility of using digital solutions to deliver stepped care in the general practice setting. Specifically, an RCT evaluating an eHealth platform (*Target-D*) that aimed to support patients in managing depression found that digitally stratifying patients and providing matched interventions resulted in greater improvements in depressive symptoms over 3 months compared with usual care [18]. Similarly, an RCT of a similar digital decision support tool (*Link-me*) demonstrated that digitally facilitated stepped care resulted in greater improvements in psychological distress in general practice patients compared with usual care [19].

Objectives

A key goal of the StepCare tool was to increase the use of digital mental health interventions in general practice by (1) helping GPs to identify patients who would be suitable candidates for digital interventions (ie, via screening for patients with mild to moderate symptoms), (2) assisting GPs in determining when a digital mental health intervention should be used alone or in combination with a higher-intensity intervention (eg, a psychologist), and (3) reducing GPs' uncertainty surrounding which digital interventions to prescribe and how to introduce them to the patient in the limited time available during a consultation. StepCare has been implemented in 85 general practices across Australia, and more than 5000 patients were screened from July 2017 to March 2020. The aim of this study is to examine whether StepCare is a suitable means for identifying patients who may be candidates for digital mental health interventions and to promote the uptake of these interventions in the general practice patient population.

Methods

Sample and Recruitment

Information about the StepCare tool was disseminated to primary health networks (PHNs) via conferences, workshops, and networks. A total of 8 PHNs from New South Wales, Australian Capital Territory, and Victoria signed up to use StepCare during the study period (July 24, 2017, and March 31,

2020), and each PHN invited expressions of interest from general practices in their region. In total, 85 general practices expressed interest, and all GPs within these practices were invited to participate. The GPs who provided informed consent to participate were then provided with information and training from either their PHN or staff from the Black Dog Institute on how to implement StepCare in their practice. Both PHNs and the Black Dog Institute staff provided ongoing support to practices throughout the implementation period.

Adult patients who were attending a GP appointment at one of the participating practices, regardless of the reason for their visit, were invited to participate in the screening. When patients presented to the reception staff, they were handed a mobile tablet that displayed an information and consent page, and patients who agreed to participate in screening indicated their consent via button click on the tablet. To assess their eligibility for screening, an initial set of questions was administered that confirmed that the patient was aged ≥18 years, had not undergone screening in the past 6 months, and could provide either a mobile phone number or email address (required for patients to receive follow-up monitoring assessments to track symptom change over time).

Ethics Statement

This study was approved by the University of New South Wales Human Research Ethics Committee (HC15827).

Design

This study used an uncontrolled, observational, prospective cohort design.

Screening Measures

Depressive symptoms were screened for using the 9-item Patient Health Questionnaire (PHQ-9) [20]. Total scores range from 0 to 27. StepCare defines severity levels as nil-minimal (0-4), mild (5-9), moderate (10-19), and severe (20-27). Item 9 was used to assess thoughts about suicide and self-harm. The presence of suicidal thoughts was flagged to the GP in a separate column on the patients' screening report, where a score of 1 on item 9 of the PHQ-9 indicated thoughts of mild severity, a score of 2 as moderate, and a score of 3 as severe. Anxiety symptoms were screened for using the 7-item Generalized Anxiety Disorder (GAD-7) scale [21]. Total scores range from 0 to 21, with severity levels defined as nil-minimal (0-4), mild (5-9), moderate (10-14), and severe (15-21).

StepCare also included 2 items that assessed whether the patient had previously discussed mental health issues with their GP and whether their current appointment was about mental health. This served as a proxy indicator of whether a patient was likely to be known to their GP as having a mental health condition. Patients were also asked sociodemographic questions, including whether they were Aboriginal or Torres Strait Islander; a carer for children, someone with a disability, someone with a chronic illness, or someone who was frail-aged; spoke a language other than English; had concerns about their accommodation or housing; and what their alcohol use habits were (Alcohol Use Disorders Identification Test) [22]. The PHQ-9 and GAD-7 scores informed the StepCare treatment recommendations; however, all responses were transmitted to GPs for consideration during the consultation.

Screening Procedure

On presentation to the practice, the reception staff provided adult patients with an internet-enabled mobile tablet featuring the StepCare information and consent page. Patients could either consent or decline to participate by clicking a button. Those who consented were asked to provide their contact details and complete the sociodemographic questions, the PHQ-9, the GAD-7, and the Alcohol Use Disorders Identification Test. The patient's screening results were sent directly to their GPs' medical inbox using a secure messaging service for review during the consultation.

Treatment Recommendations

StepCare treatment recommendations were stratified according to the patient's symptom severity. Digital mental health interventions were incorporated into Steps 1 and 2 as follows.

Step 1: Web-Based Self-Help

The self-guided digital mental health intervention myCompass [23] was recommended for patients with mild symptoms. MyCompass has been shown to be effective in improving symptoms of depression and anxiety over 7 weeks relative to waitlist and attention control conditions [23] and has more than 50,000 registered users to date.

Step 2: Guided Web-Based Therapy or Face-to-Face Psychotherapy (Consider Medication)

Guided web-based therapy via the MindSpot Clinic or face-to-face therapy with a clinical psychologist was recommended for patients with moderate symptoms. MindSpot offers therapist-guided web-based courses and has been found to produce significant improvements in depression and anxiety [24,25]. The Step 2 recommendation also suggested that the GP consider pharmacotherapy.

Step 3: Face-to-Face Psychotherapy, Antidepressant Medication (Consider a Psychiatrist)

Referral to a clinical psychologist and pharmacotherapy were recommended for patients with severe symptoms. The GP was encouraged to consider referral to a psychiatrist for pharmacotherapy management.

Managing Suicidality

For patients who scored 1 or higher on item 9 of the PHQ-9, StepCare also provided GPs with several options to assist them in supporting patients who reported suicidal thoughts, including providing prompts to help initiate a discussion with the patient regarding their responses on the screener, steps for further assessing risk, information describing how to develop a safety plan for patients who are actively suicidal, and links to relevant local and national crisis services.

GP Prescribing Patterns

During the consultation, GPs were asked to record the treatment they prescribed by filling in checkboxes accessible via the patient's screening report. Although GPs were strongly

encouraged to record prescribed treatments, to minimize the burden on GP workflow, this was not mandatory.

Patient Outcomes and Treatment Use

Individuals who screened positive for depression or anxiety were invited to complete follow-up assessments every 2 weeks for 18 weeks to monitor symptoms (those with nil-minimal symptoms were not followed up). These follow-up assessments captured key outcome measures, including depressive symptom severity (PHQ-9) and anxiety symptom severity (GAD-7). In addition to these outcome measures, patients were asked to indicate, via checkboxes, which of a series of mental health treatments they had used in the past 2 weeks. Follow-up assessments were not mandatory, and patients could opt out of receiving assessment reminders at any time.

Each time a patient completed a follow-up assessment, a report was sent to their GP's medical inbox. This report displayed the patient's PHQ-9 and GAD-7 scores, their score on item 9 of the PHQ-9 (indicating suicidal thoughts), and a line graph showing symptom improvement from baseline. Accompanying this report was also a series of alerts that notified GPs of patients who showed signs of improvement, as well as those who showed evidence of deterioration, severe symptoms that did not improve, and nonadherence to monitoring assessments. Improvement was defined as a patient whose PHQ-9 or GAD-7 scores had improved by at least one severity category compared with the previous fortnight; deterioration was flagged when a patient's PHQ-9 or GAD-7 score increased by at least one severity category relative to the previous fortnightly period; severe and unchanging symptoms were flagged where a patient with PHQ-9 or GAD-7 scores in the severe range at baseline had not improved by at least one severity category by week 4; and nonadherence to assessments was flagged when a patient who screened positive for depressive or anxious symptoms at baseline had not completed two consecutive follow-up assessments. For patients who did not improve, or who deteriorated, the GP was encouraged to consider scheduling a follow-up appointment with the patient to review their treatment plan.

For the purposes of our statistical analyses, we defined *remitters* as individuals who were nil-minimal (0-4) on the PHQ-9 and GAD-7 at their last follow-up assessment, and *responders* as individuals who showed a decrease of at least one severity category on the PHQ-9 or GAD-7 at their last assessment.

Statistical Analyses

Descriptive statistics were used to quantify patient symptoms, treatment use, and GP prescription patterns. The number needed to screen (NNS) was calculated using methods aligned with those adopted in multiarm clinical trials, as described by Rembold [26]. NNS is a statistic that is derived from the number needed to treat (NNT) statistic that is commonly reported in clinical trials assessing the effectiveness of an intervention and reflects the number of people who need to be screened to prevent one adverse event. However, it differs from NNT in that it also incorporates information about the prevalence of undetected diseases that can be potentially identified via screening. In our study, it was calculated by first computing the absolute risk (AR) of identifying undiagnosed symptoms of anxiety or depression through screening and the AR of identifying undiagnosed symptoms of anxiety or depression under a hypothetical *no screening* condition. For the purposes of our calculation, we assumed that in the hypothetical *no screening* condition, no individuals with undiagnosed symptoms of depression or anxiety would have been identified that would not have already been identified through care as usual without screening. Next, we computed the difference in AR under these two conditions (ie, the AR deduction [ARD]). Finally, we computed the NNS in the same manner as that used to compute NNT, which is the inverse of the ARD (ie, 1/ARD) [26].

This was supplemented with an intention-to-treat analysis examining symptom improvement over the follow-up period using data from all assessment points (mixed model for repeated measures analysis with random intercepts and slopes; implemented in Stata [version 13.1; StataCorp LP]). We also performed a completer analysis to examine the rates of response and remission.

Role of the Funding Source

The funder had no role in the study design, data collection, analysis, manuscript writing, or in the decision to publish the manuscript.

Results

Sample Characteristics

Of the 5138 patients who were offered the tool, 3777 (73.51%) completed screening and 3314 (64.50%) met the eligibility criteria (sample characteristics are shown in Table 1).

Table 1. Characteristics of baseline sample (N=3314).

Characteristic	Participants
Age (years), mean (SD)	43.4 (17.0)
Female, n (%)	2316 (69.89)
Aboriginal or Torres Strait Islander, n (%)	165 (4.98)
English as second language, n (%)	423 (12.76)
Cares for children, n (%)	1079 (32.56)
Cares for frail or disabled individuals, n (%)	305 (9.20)
Accommodation issues, n (%)	94 (2.84)
Seeing general practitioner for mental health reasons, n (%)	745 (22.48)
Seen general practitioner for mental health reasons previously, n (%)	1519 (45.84)

Suitability of StepCare as a Screening Tool

Prevalence and Severity of Symptoms Identified

In total, 1428 individuals with symptoms of anxiety or depression were identified through screening (detection rate: 1428/3314, 43.09%). Of the baseline sample, 13.82% (458/3314) had mild symptoms, 17.68% (586/3314) had moderate symptoms, and 11.59% (384/3314) had severe symptoms. Furthermore, just under one third (454/1428, 31.79%) of symptomatic individuals reported suicidal ideation or thoughts of self-harm. Overall, depressive symptoms (1333/3314, 40.22%) were more common than anxiety symptoms (855/3314, 25.80%). There were no differences in the proportion of males and females who screened positive for either depressive or anxious symptoms (both values of $P>.18$).

For patients who screened in the mild symptom severity range, most screened positive for depressive symptoms only (346/458, 75.5%) with fewer screening positive for anxiety symptoms only (54/458, 11.8%) or both depressive and anxious symptoms (49/458, 10.7%). In contrast, as symptom severity increased, patients were more likely to screen positive for both depressive and anxious symptoms. Specifically, in the moderate range, 59.6% (349/586) screened positive for both depressive and anxious symptoms [depressive symptoms only (202/586, 34.5%); anxious symptoms only (28/586, 4.8%)], and in the severe range, 94.3% (362/384) screened positive for both depressive and anxious symptoms [depressive symptoms only (7/384, 1.8%); anxious symptoms only (13/384, 3.4%)].

Previously Unidentified or Untreated Mental Health Symptoms

Of the 1428 patients who screened positive for depressive or anxious symptoms, 335 (23.46% of symptomatic sample or 10.11% of patients overall) had never seen their GP for mental health reasons, including at the time of screening. These previously unidentified or untreated patients were older than symptomatic patients who had previously seen their GP for mental health reasons (44.73 vs 38.58; $t_{1,426}=6.36$; $P<.001$). Most unidentified or untreated patients had mild (184/335, 54.9%) or moderate symptoms (116/335, 34.6%); however, 1 in 10 had severe symptoms (35/335, 10.4%) or reported suicidal ideation or thoughts of self-harm (49/335, 14.6%), indicating that StepCare also identified patients with significant unmet mental health treatment needs.

Results for NNS

Of the 5138 patients who underwent screening, 335 (6.52%) had anxious or depressive symptoms that were previously unidentified or untreated. Accordingly, the AR for recognizing previously unidentified symptoms via screening was $AR_{screening}=335/5138=0.0652$. Although this was not an RCT, under a hypothetical *no screening* condition, we can assume that none of these individuals would have been detected, yielding an AR of:

$$AR_{no\ screening}=0/5138=0$$

Therefore, the ARD can be calculated as $AR_{screening}–AR_{no\ screening}$:

$$ARD=335/5138–0/5138=0.0652$$

An NNS can be calculated as the inverse of the ARD:

$$NNS=1/0.0652=15.3\ (95\%\ CI\ 13.9\text{-}17.1)$$

This indicates that for every 16 patients who are offered mental health screening, 1 individual with previously unidentified or untreated depressive or anxious symptoms will be identified.

Ability of StepCare to Facilitate Uptake of Digital Mental Health Interventions

GP Prescribing Patterns

GP prescription data were available for 23.39% (334/1428) of patients. These included 30.2% (101/334) patients with mild symptoms, 41.9% (140/334) with moderate symptoms, and 27.8% (93/334) with severe symptoms. Separate independent samples 2-tailed *t* tests showed that baseline PHQ-9 and GAD-7 scores did not differ between patients whose GP did and did not provide prescribing data (both values of $P>.54$).

Although digital and high-intensity (ie, psychologists, psychiatrists, and pharmacotherapy) interventions were recommended by the StepCare tool at roughly equal rates (digital interventions were recommended for 241/334, 72.2% of patients; high-intensity interventions were recommended for 233/334, 69.8% of patients; categories not mutually exclusive), GPs prescribed high-intensity interventions nearly twice as often as low-intensity digital interventions. Specifically, a high-intensity

intervention was prescribed by the GP for 56.3% (188/334) of patients, whereas a digital mental health intervention was prescribed by the GP for 30.8% (103/334) of patients. A side-by-side comparison of the treatments recommended by StepCare, relative to the treatments prescribed by GPs, is shown in Table 2. Furthermore, GP prescription patterns among patients with different levels of symptom severity are shown in Table 3.

Table 2. Comparison of StepCare treatment recommendations for all symptomatic patients (n=1428) versus GP[a] prescribed treatments in symptomatic patients with GP prescribing data available (n=334).

Treatment[b]	Proportion of all symptomatic patients (n=1428) who were recommended a specific treatment by StepCare, n (%)	Proportion of symptomatic patients with GP prescribing data (n=334) who were prescribed a specific treatment by their GP, n (%)
Web-based self-help	458 (32.1)	60 (18)
Guided web-based therapy	586 (41)	54 (16.2)
Psychologist	970 (67.9)	135 (40.4)
Medication	970 (67.9)	109 (32.6)
Psychiatrist	384 (26.9)	31 (9.3)

[a]GP: general practitioner.

[b]Column percentage totals do not sum to 100% as patients could be prescribed multiple treatments or no treatment at all.

Table 3. General practitioner prescribing patterns as a function of patient baseline symptom severity (N=334).

Treatment prescribed by general practitioner[a]	Patients with mild symptoms (n=101), n (%)	Patients with moderate symptoms (n=140), n (%)	Patients with severe symptoms (n=93), n (%)
Prescribed web-based self-help	33 (32.7)	15 (10.7)	12 (12.9)
Prescribed guided web-based therapy	11 (10.9)	28 (20)	15 (16.1)
Prescribed psychologist	22 (21.8)	59 (42.1)	54 (58.1)
Prescribed medication	8 (7.9)	50 (35.7)	51 (54.8)
Prescribed psychiatrist	2 (2)	8 (5.7)	21 (22.6)

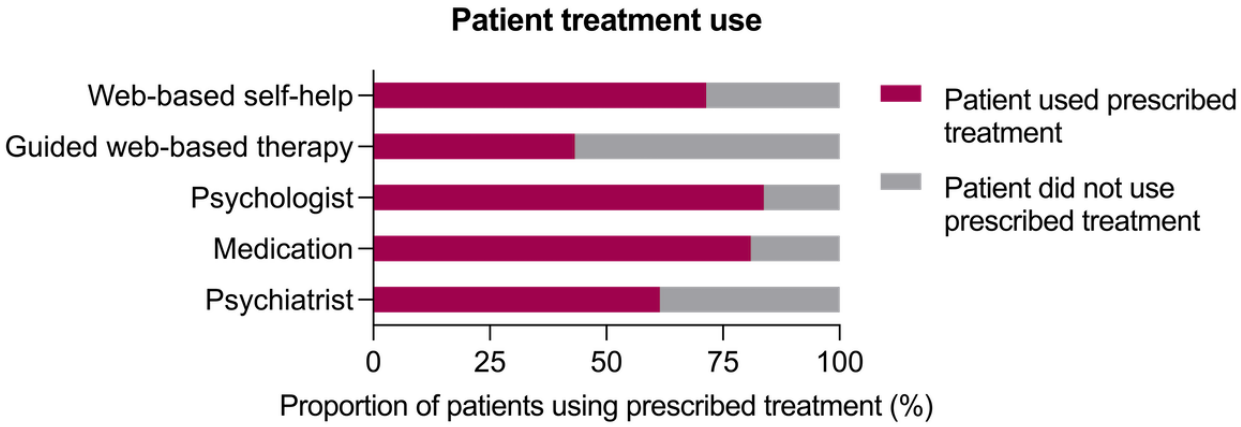
[a]Percentages reflect the percentage of patients in each symptom severity category; column percentage totals do not sum to 100% as patients in the same symptom severity category could be prescribed multiple treatments or no treatment at all.

Patient Treatment Use Patterns

Treatment use data were available for 42.44% (606/1428) of patients. Most patients (525/606, 86.6%) reported the use of at least one form of mental health treatment during the follow-up period. Patients with treatment use data had higher baseline PHQ-9 (t_{1426}=3.98; *P*<.001; mean 12.81, SD 6.16) and GAD-7 scores (t_{1408}=3.77; *P*<.001; mean 9.43, SD 6.59) than patients who did not provide treatment use data (PHQ-9 mean 11.50, SD 6.18; GAD-7 mean 8.10, SD 6.59), indicating that the following patterns may be more applicable to patients with more severe symptoms.

Of the 606 patients who provided treatment use data, 165 (27.2%) also had GP prescribing data available, which allowed us to determine whether a patient used the treatment prescribed by their GP. More than two-thirds of patients (35/52, 67%) who were prescribed a digital mental health intervention (either alone or in combination with a high-intensity intervention) reported using one of these digital interventions over the follow-up period. The rates of treatment uptake were also high for high-intensity interventions; nearly all patients (96/102, 94.1%) who were prescribed a high-intensity intervention by their GP reported that they had used a high-intensity intervention over the follow-up period. The patient use of GP-prescribed interventions is shown in detail in Figure 1.

Figure 1. Data show the rate at which patients used the treatment that was prescribed to them by their general practitioner. Values are expressed as a proportion of symptomatic patients for whom both patient adherence and general practitioner prescribing data were available (n=165).



Patient Outcomes

Symptom Improvement Over Time

The intention to treat (mixed model for repeated measures) analysis revealed a significant main effect of *time* for the PHQ-9 (*P*<.001), where scores decreased by an average of 3.62 points from baseline to week 18 (moderate to mild severity range; Cohen *d*=0.58 computed using baseline SD of 6.20). Comparisons between baseline and follow-up PHQ-9 scores (Table 4) showed that symptom severity at each follow-up point was significantly lower than the score at baseline (all *P*s<.001). Similar effects were observed for the GAD-7, where the main effect of *time* was significant (*P*<.001), with scores decreasing by an average of 1.45 points from baseline to week 18 (upper end of mild to lower end of mild severity range; Cohen *d*=0.22, baseline SD 6.62). Comparisons between baseline and follow-up GAD-7 scores (Table 5) showed that symptom severity at each follow-up point except weeks 2 and 4 was significantly lower than at baseline (all values of *P*<.001).

Table 4. Changes in depressive symptom severity (9-item Patient Health Questionnaire) over time.

Time point	Estimated marginal, mean (SE)	Modeled change from baseline, mean (95% CI[a])	*P* value[b]
Baseline	12.05 (0.16)	N/A[c]	N/A
Week 2	10.22 (0.22)	−1.84 (−2.23 to −1.44)	<.001
Week 4	9.75 (0.24)	−2.31 (−2.75 to −1.86)	<.001
Week 6	9.15 (0.25)	−2.90 (−3.37 to −2.43)	<.001
Week 8	8.86 (0.26)	−3.19 (−3.70 to −2.69)	<.001
Week 10	8.81 (0.28)	−3.24 (−3.78 to −2.69)	<.001
Week 12	8.76 (0.29)	−3.30 (−3.87 to −2.73)	<.001
Week 14	8.33 (0.31)	−3.72 (−4.31 to −3.13)	<.001
Week 16	8.42 (0.32)	−3.63 (−4.25 to −3.01)	<.001
Week 18	8.43 (0.33)	−3.62 (−4.27 to −2.97)	<.001

[a]The 95% CI refers to the change from baseline.

[b]*P* values indicate the significance level for comparisons between baseline scores and scores at each of the follow-up assessments.

[c]N/A: not applicable.

Table 5. Change in anxious symptom severity (7-item Generalized Anxiety Disorder scale) over time.

Time point	Estimated marginal, mean (SE)	Modeled change from baseline, mean (95% CI[a])	*P* value[b]
Baseline	8.67 (0.18)	N/A[c]	N/A
Week 2	8.56 (0.20)	−0.11 (−0.51 to 0.29)	.58
Week 4	8.26 (0.21)	−0.41 (−0.83 to 0.01)	.06
Week 6	7.86 (0.22)	−0.81 (−1.27 to −0.36)	<.001
Week 8	7.63 (0.24)	−1.04 (−1.54 to −0.55)	<.001
Week 10	7.33 (0.25)	−1.34 (−1.85 to −0.83)	<.001
Week 12	7.27 (0.25)	−1.40 (−1.92 to −0.88)	<.001
Week 14	7.18 (0.27)	−1.49 (−2.05 to −0.93)	<.001
Week 16	7.48 (0.30)	−1.20 (−1.81 to −0.58)	<.001
Week 18	7.22 (0.30)	−1.45 (−2.07 to −0.83)	<.001

[a]The 95% CI refers to the change from baseline.

[b]*P* values indicate the significance level for comparisons between baseline scores and scores at each of the follow-up assessments.

[c]N/A: not applicable.

Rates of Remission and Response

Completer analyses focused on symptomatic patients who completed at least one follow-up assessment (708/1428, 49.58%). Of this sample, 26.1% (185/708) scored in the asymptomatic range on the PHQ-9 and GAD-7 at follow-up and were classified as *remitters*. Furthermore, 55.4% (392/708) of cases dropped down at least one severity category on the PHQ-9 or GAD-7 and were considered *responders*. In contrast, 26.1% (185/708) increased by one or more severity categories during the follow-up period.

Discussion

Principal Findings

This study examined the performance of a digital mental health screening and treatment recommendation tool (*StepCare*) for improving the identification of depression and anxiety in general practice and for promoting the uptake of digital mental health interventions. Four key findings emerged. First, screening revealed a high prevalence of common mental health conditions among general practice patients, with 40.2% screening positive for depression and 25.8% screening positive for anxiety (cf. Australian population 12-month prevalence estimates of 6.2% for depression and 14.4% for anxiety) [27]. Nearly 1 in 4 of these patients had never consulted their GP about their mental health previously, indicating that screening identified individuals who may have otherwise been undetected. Second, the majority (89.5%) of untreated patients were in the mild to moderate range, demonstrating that screening, when followed by appropriate assessment, could assist GPs in identifying suitable candidates for digital mental health interventions. Third, although digital mental health interventions and high-intensity interventions were recommended by the StepCare tool at approximately equal rates, GPs favored high-intensity interventions over digital interventions. Fourth, although digital interventions were prescribed less often, more than two-thirds of patients who were prescribed a digital intervention by their GP reported using it. This confirms that GPs are well placed to facilitate greater uptake of digital mental health interventions.

Our findings align with prior studies showing high rates of depressive and anxious symptoms in general practice populations and the ability for mental health screening to identify patients with untreated symptoms [14]. These results extend prior research in two significant ways: (1) by estimating the NNS to detect previously untreated patients with depressive or anxious symptoms in an Australian general practice setting (NNS=16) and (2) by showing that a significant proportion of untreated patients have symptoms in the mild to moderate severity range, where digital mental health interventions may be especially useful. Although the impact of mental health screening on patient outcomes can only be determined via RCTs, our findings suggest that screening could identify important opportunities to implement low or no cost digital interventions, thereby reducing the burden on higher-intensity services.

Our finding that 1 in every 16 patients invited for screening had unidentified or untreated depressive or anxious symptoms positions mental health screening favorably when compared with other diseases routinely screened for in primary care. For example, type 2 diabetes screening for adults aged ≥40 years yields a detection rate of 1 in 32 [28], breast cancer screening for women aged 50 to 69 years yields a detection rate of approximately 1 in 94 [29], and cervical cancer screening in women aged >20 years yields a detection rate of approximately 1 in 143 [30]. Given that suicide outranks cancer as the leading cause of death for Australians aged 25 to 44 years [31] (the mean age of our sample was 43 years), mental health screening of the general practice population could significantly reduce suicide-related mortality.

We note that the rates of depressive and anxious symptoms in our study were higher than the population prevalence estimates of depression and anxiety, and there are several possible reasons for this. First, population prevalence estimates are based on individuals who meet the diagnostic criteria for depression or anxiety rather than individuals who demonstrate a positive result

on a brief screening measure of depressive or anxious symptoms. In our study, it is likely that a subset of patients who screened positive for depression or anxiety would not have met the diagnostic criteria for depression or anxiety in further follow-up assessments with their GP. Second, it is likely that some participants who screened positive were experiencing transient symptoms rather than a persistent condition, with prior studies indicating that approximately one-fifth of patients who screened positive for depression in primary care no longer met the criteria after 2 weeks [32]. Third, the PHQ-9 and GAD-7 cutoff values we used were designed to capture milder subsyndromal symptoms in addition to more severe symptomatology; hence, the symptomatic sample in our study likely encompasses patients with milder symptoms compared with individuals who are included in epidemiological studies.

A fourth factor that likely underpinned our higher prevalence rates was the fact that several practices used StepCare to selectively screen high-risk patients (as recommended by the Royal Australian College of General Practitioners) [33] as opposed to implementing universal screening. In addition to biasing screening toward more symptomatic patients, this observation raises an important question regarding the best implementation of screening programs in general practice. Although selectively screening high-risk patients (eg, postpartum women) may increase the cost-effectiveness of screening, this approach may have the unintended effect of exacerbating existing disparities in access to mental health care [34]. For example, the Royal Australian College of General Practitioners guidelines indicate that patients at increased risk for depression are those who have a family history of psychiatric illness, have chronic medical conditions, are unemployed or of low socioeconomic status, have experienced significant life events, family violence or child abuse, or are part of the lesbian, gay, bisexual, transgender, and intersex community [33]. Although these factors may be known to the GP when there is a well-established relationship with the patient, this is less likely when a patient does not have a regular GP (up to 20% of patients) [35] or visits their GP infrequently (eg, indigenous patients or patients from rural areas) [36]. Accordingly, although risk-based screening may have lower opportunity costs than universal screening, it may fail to identify patients who may warrant further assessment by their GP and who may benefit from digital mental health interventions.

Regarding GP prescribing patterns, our findings suggest that although digital mental health interventions may be well suited to a substantial portion of the symptomatic general practice population, GPs show a preference for prescribing high-intensity treatments. Why might this be? Common barriers to use reported by GPs include uncertainty about a digital intervention's evidence base and insufficient knowledge of how to refer a patient to a digital intervention [15]. However, StepCare was designed to overcome these barriers, as it only recommends empirically supported digital interventions and includes a referral pipeline that supports GPs in connecting their patients with the intervention. Additional barriers clearly exist and must be addressed. One by-product of the COVID-19 pandemic has been the rapid integration of telehealth services into general practices. This may help to break down attitudinal barriers to digital mental health tools; however, other practical changes, such as allowing GPs to formally itemize and charge for the time taken to discuss digital interventions with their patients, may also be needed to enhance their use in general practice.

Limitations and Future Directions

This study has some limitations. First, this was an uncontrolled observational study of a tool used in a clinical setting. RCTs are needed to determine whether the opportunity costs of mental health screening in general practices are offset by superior patient outcomes. Second, GP prescribing data and patient treatment use data were available for only a portion of patients because completion of these surveys was optional. Greater tracking of prescription and treatment use patterns via electronic health records will be critical for providing a comprehensive perspective on digital intervention uptake. Finally, an important question that must be addressed in future research is whether patients who are prescribed digital mental health interventions show sufficient levels of engagement with these interventions to yield therapeutic benefits. If the rates of initial uptake are high when referred by a GP, but ongoing engagement remains low, then this may warrant consideration of alternative modes of delivery. Delivery methods that combine digital mental health interventions with face-to-face services, such as blended care [37-39], may be especially useful in this regard. Blended care has been found to promote greater adherence to digital programs, and preliminary evidence has demonstrated its efficacy in the general practice setting (eg, when guidance is provided by a nurse practitioner [40]).

Conclusion

In conclusion, our findings indicate that a digital mental health screening and treatment recommendation tool may increase the opportunities to use digital mental health interventions in general practice. Leveraging these opportunities will be critical in addressing increased mental health treatment needs arising from the COVID-19 pandemic and in reducing existing disparities in access to affordable, evidence-based mental health care.

Acknowledgments

The authors wish to thank Mariam Faraj and other staff of the Central and Eastern Sydney PHN for their involvement in the co-design of StepCare, as well as all the GPs, staff, and patients who participated in the research. This research was funded by the Paul Ramsay Foundation. Associate Professor JMN was supported by an Australian Medical Research Future Fund Fellowship.

Conflicts of Interest

StepCare was designed and developed by Proudfoot, Anderson, Cockayne, O'Moore, Gale, and Christensen. PHNs in Australia pay a small fee to the Black Dog Institute to use StepCare in their general practices, which contributes to the operational, information technology infrastructure, and personnel costs associated with implementing StepCare, hosting the data, and providing technical support. The authors of this study derive no personal financial benefits from StepCare.

References

1. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
2. Thornicroft G, Chatterji S, Evans-Lacko S, Gruber M, Sampson N, Aguilar-Gaxiola S, et al. Undertreatment of people with major depressive disorder in 21 countries. Br J Psychiatry 2017 Feb;210(2):119-124 [FREE Full text] [doi: 10.1192/bjp.bp.116.188078] [Medline: 27908899]
3. Voros V, Fekete S, Tenyi T, Rihmer Z, Szili I, Osvath P. Untreated depressive symptoms significantly worsen quality of life in old age and may lead to the misdiagnosis of dementia: a cross-sectional study. Ann Gen Psychiatry 2020 Sep 15;19:52 [FREE Full text] [doi: 10.1186/s12991-020-00302-6] [Medline: 32944058]
4. Ribeiro JD, Huang X, Fox KR, Franklin JC. Depression and hopelessness as risk factors for suicide ideation, attempts and death: meta-analysis of longitudinal studies. Br J Psychiatry 2018 May;212(5):279-286. [doi: 10.1192/bjp.2018.27] [Medline: 29587888]
5. Pálinkás A, Sándor J, Papp M, Kőrösi L, Falusi Z, Pál L, et al. Associations between untreated depression and secondary health care utilization in patients with hypertension and/or diabetes. Soc Psychiatry Psychiatr Epidemiol 2019 Feb;54(2):255-276. [doi: 10.1007/s00127-018-1545-7] [Medline: 29947863]
6. Depression in adults: recognition and management. Clinical guideline CG90. National institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/cg90 [accessed 2021-08-26]
7. Australian government response to contributing lives, thriving communities - Review of mental health programmes and services. Commonwealth of Australia Canberra, Australia. URL: https://www.health.gov.au/sites/default/files/documents/2020/11/australian-government-response-to-contributing-lives-thriving-communities-review-of-mental-health-programmes-and-services.pdf [accessed 2021-08-26]
8. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
9. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
10. Hedman E, Furmark T, Carlbring P, Ljótsson B, Rück C, Lindefors N, et al. A 5-Year follow-up of internet-based cognitive behavior therapy for social anxiety disorder. J Med Internet Res 2011 Jun 15;13(2):e39 [FREE Full text] [doi: 10.2196/jmir.1776] [Medline: 21676694]
11. Ebert DD, Berking M, Cuijpers P, Lehr D, Pörtner M, Baumeister H. Increasing the acceptance of internet-based mental health interventions in primary care patients with depressive symptoms. A randomized controlled trial. J Affect Disord 2015 May 01;176:9-17. [doi: 10.1016/j.jad.2015.01.056] [Medline: 25682378]
12. Tamburrino MB, Lynch DJ, Nagel RW, Smith MK. Primary care evaluation of mental disorders (PRIME-MD) screening for minor depressive disorder in primary care. Prim Care Companion J Clin Psychiatry 2009;11(6):339-343 [FREE Full text] [doi: 10.4088/PCC.08.m00711] [Medline: 20098526]
13. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007 Mar 06;146(5):317-325. [doi: 10.7326/0003-4819-146-5-200703060-00004] [Medline: 17339617]
14. Carey M, Jones K, Meadows G, Sanson-Fisher R, D'Este C, Inder K, et al. Accuracy of general practitioner unassisted detection of depression. Aust N Z J Psychiatry 2014 Jun;48(6):571-578 [FREE Full text] [doi: 10.1177/0004867413520047] [Medline: 24413807]
15. Breedvelt JJ, Zamperoni V, Kessler D, Riper H, Kleiboer AM, Elliott I, et al. GPs' attitudes towards digital technologies for depression: an online survey in primary care. Br J Gen Pract 2019 Mar;69(680):e164-e170 [FREE Full text] [doi: 10.3399/bjgp18X700721] [Medline: 30559111]
16. Anderson J, O Moore K, Faraj M, Proudfoot J. Stepped care mental health service in Australian primary care: codesign and feasibility study. Aust Health Rev 2020 Dec;44(6):873-879. [doi: 10.1071/AH19078] [Medline: 31662177]
17. Proudfoot J, Gale N, O'Moore K, Faraj M, Gieng C, Anderson J. Implementation of an augmented stepped mental health care service in Australian primary care: a mixed method study. Adv Ment Health 2019 Sep 12;19(1):75-93. [doi: 10.1080/18387357.2019.1662318]

18. Fletcher S, Chondros P, Densley K, Murray E, Dowrick C, Coe A, et al. Matching depression management to severity prognosis in primary care: results of the Target-D randomised controlled trial. Br J Gen Pract 2021 Jan 28;71(703):e85-e94 [FREE Full text] [doi: 10.3399/BJGP.2020.0783] [Medline: 33431380]
19. Fletcher S, Spittal MJ, Chondros P, Palmer VJ, Chatterton ML, Densley K, et al. Clinical efficacy of a Decision Support Tool (Link-me) to guide intensity of mental health care in primary practice: a pragmatic stratified randomised controlled trial. Lancet Psychiatry 2021 Mar;8(3):202-214. [doi: 10.1016/S2215-0366(20)30517-4] [Medline: 33571453]
20. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
21. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
22. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. Arch Intern Med 1998 Sep 14;158(16):1789-1795. [doi: 10.1001/archinte.158.16.1789] [Medline: 9738608]
23. Proudfoot J, Clarke J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013 Nov 18;13:312 [FREE Full text] [doi: 10.1186/1471-244X-13-312] [Medline: 24237617]
24. Titov N, Dear BF, Staples LG, Bennett-Levy J, Klein B, Rapee RM, et al. The first 30 months of the MindSpot Clinic: evaluation of a national e-mental health service against project objectives. Aust N Z J Psychiatry 2017 Dec;51(12):1227-1239. [doi: 10.1177/0004867416671598] [Medline: 27733709]
25. Titov N, Dear BF, Staples LG, Bennett-Levy J, Klein B, Rapee RM, et al. MindSpot clinic: an accessible, efficient, and effective online treatment service for anxiety and depression. Psychiatr Serv 2015 Oct;66(10):1043-1050. [doi: 10.1176/appi.ps.201400477] [Medline: 26130001]
26. Rembold CM. Number needed to screen: development of a statistic for disease screening. BMJ 1998 Aug 01;317(7154):307-312 [FREE Full text] [doi: 10.1136/bmj.317.7154.307] [Medline: 9685274]
27. National survey of mental health and wellbeing: summary of results. Australian Bureau of Statistics. URL: https://www.abs.gov.au/statistics/health/mental-health/national-survey-mental-health-and-wellbeing-summary-results/latest-release#summary-of-findings [accessed 2021-08-26]
28. Lee CM, Versace VL, Malo JA, Shaw JE, Dunbar JA, Colagiuri S. Screening for diabetes prevention with diabetes risk scores - a balancing act. Diabetes Res Clin Pract 2018 Jan;135:120-127. [doi: 10.1016/j.diabres.2017.11.009] [Medline: 29155122]
29. BreastScreen Australia monitoring report 2020. Australian Institute of Health and Welfare. URL: https://www.aihw.gov.au/reports/cancer-screening/breastscreen-australia-monitoring-report-2020/contents/summary [accessed 2021-08-26]
30. Cervical screening in Australia. Australian Institute of Health and Welfare. URL: https://www.aihw.gov.au/reports/cancer-screening/cervical-screening-in-australia-2018/contents/table-of-contents [accessed 2021-08-26]
31. Deaths in Australia. Australian Institute of Health and Welfare. 2021. URL: https://www.aihw.gov.au/reports/life-expectancy-death/deaths-in-australia/contents/leading-causes-of-death [accessed 2021-08-26]
32. Gunn JM, Gilchrist GP, Chondros P, Ramp M, Hegarty KL, Blashki GA, et al. Who is identified when screening for depression is undertaken in general practice? Baseline findings from the Diagnosis, Management and Outcomes of Depression in Primary Care (diamond) longitudinal study. Med J Aust 2008 Jun 16;188(S12):S119-S125. [doi: 10.5694/j.1326-5377.2008.tb01874.x] [Medline: 18558911]
33. Guidelines for Preventive Activities in General Practice 9th edition. Australia: The Royal Australian College of General Practitioners; 2016.
34. Akincigil A, Matthews EB. National rates and patterns of depression screening in primary care: results from 2012 and 2013. Psychiatr Serv 2017 Jul 01;68(7):660-666. [doi: 10.1176/appi.ps.201600096] [Medline: 28196461]
35. Wright M, Hall J, van Gool K, Haas M. How common is multiple general practice attendance in Australia? Aust J Gen Pract 2018 May;47(5):289-296 [FREE Full text] [doi: 10.31128/AJGP-11-17-4413] [Medline: 29779298]
36. Harrison C, Henderson J, Miller G, Britt H. Predicting patient use of general practice services in Australia: models developed using national cross-sectional survey data. BMC Fam Pract 2019 Feb 14;20(1):28 [FREE Full text] [doi: 10.1186/s12875-019-0914-y] [Medline: 30764778]
37. Erbe D, Eichert H, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017 Sep 15;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]
38. Massoudi B, Blanker MH, van Valen E, Wouters H, Bockting CL, Burger H. Blended care vs. usual care in the treatment of depressive symptoms and disorders in general practice [BLENDING]: study protocol of a non-inferiority randomized trial. BMC Psychiatry 2017 Jun 13;17(1):218 [FREE Full text] [doi: 10.1186/s12888-017-1376-1] [Medline: 28610561]

39. Mathiasen K, Andersen TE, Riper H, Kleiboer AA, Roessler KK. Blended CBT versus face-to-face CBT: a randomised non-inferiority trial. BMC Psychiatry 2016 Dec 05;16(1):432 [FREE Full text] [doi: 10.1186/s12888-016-1140-y] [Medline: 27919234]
40. Van der Zweerde T, Lancee J, Slottje P, Bosmans JE, Van Someren EJ, van Straten A. Nurse-guided internet-delivered cognitive behavioral therapy for insomnia in general practice: results from a pragmatic randomized clinical trial. Psychother Psychosom 2020;89(3):174-184 [FREE Full text] [doi: 10.1159/000505600] [Medline: 32069463]

Abbreviations

AR: absolute risk
ARD: absolute risk detection
GAD-7: 7-item Generalized Anxiety Disorder scale
GP: general practitioner
NNS: number needed to screen
NNT: number needed to treat
PHN: primary health network
PHQ-9: 9-item Patient Health Questionnaire
RCT: randomized controlled trial

Edited by R Kukafka; submitted 03.03.21; peer-reviewed by L Martinengo, C Kaylor-Hughes; comments to author 19.06.21; revised version received 07.07.21; accepted 27.07.21; published 16.09.21.

Please cite as:
Whitton AE, Hardy R, Cope K, Gieng C, Gow L, MacKinnon A, Gale N, O'Moore K, Anderson J, Proudfoot J, Cockayne N, O'Dea B, Christensen H, Newby JM
Mental Health Screening in General Practices as a Means for Enhancing Uptake of Digital Mental Health Interventions: Observational Cohort Study
J Med Internet Res 2021;23(9):e28369
URL: https://www.jmir.org/2021/9/e28369
doi:10.2196/28369
PMID:34528896

Original Paper

The Associations Between Mental Health Problems and Attitudes Toward Web-Based Health and Social Care Services: Evidence From a Finnish Population-Based Study

Teemu Rantanen[1], PhD; Kia Gluschkoff[2,3], PhD; Piia Silvennoinen[1], PhD; Tarja Heponiemi[2], PhD

[1]Unit of Digital Education and Master Programmes, Laurea University of Applied Sciences, Vantaa, Finland

[2]Welfare State Research and Reform, Finnish Institute for Health and Welfare, Helsinki, Finland

[3]Department of Psychology and Logopedics, University of Helsinki, Helsinki, Finland

Corresponding Author:
Teemu Rantanen, PhD
Unit of Digital Education and Master Programmes
Laurea University of Applied Sciences
Ratatie 22
Vantaa, 01300
Finland
Phone: 358 408306149
Email: teemu.rantanen@laurea.fi

Abstract

Background: The significance of web-based health and social care services has been highlighted in recent years. There is a risk that the digitalization of public services will reinforce the digital and social exclusion of vulnerable groups, such as individuals with mental health problems.

Objective: This study aims to examine the associations between mental health problems and attitudes toward web-based health and social care services in the general population. The attitudes measured include lack of interest, perceived need for face-to-face encounters, and concern for safety. The study also evaluates whether sociodemographic characteristics (age, gender, education level, and poverty) modify these associations.

Methods: Cross-sectional population-based data were collected from 4495 Finnish adults in 2017. Linear regression was used to examine the main effects and interactions of poor mental health and sociodemographic characteristics on attitudes toward web-based health and social care services.

Results: The results show that mental health was associated with attitudes toward web-based health and social care services. Individuals with mental health problems were especially concerned about the safety of web-based services. Poor mental health was independently associated with negative attitudes toward web-based services over the effects of sociodemographic factors. Some of the associations between poor mental health and negative attitudes toward web-based services were stronger among older people and men. With regard to sociodemographic characteristics, particularly higher age, low education, and poverty were associated with negative attitudes toward web-based health and social care services.

Conclusions: Poor mental health is associated with negative attitudes toward web-based health and social care services and thus indirectly with exclusion. It seems that being older and being male both reinforce the link between poor mental health and exclusion. In supporting the digital inclusion of people with mental health problems, attention should be paid to guidance and counseling, reliability, and the user-friendliness of web-based services as well as to the prevention of poverty. In addition, it is essential to see web-based services as complementary to, and not a substitute for, face-to-face services.

(J Med Internet Res 2021;23(9):e28066) doi:10.2196/28066

KEYWORDS

digital inclusion; digital exclusion; digital divide; mental health; attitudes

Introduction

Background

The push for web-based health and social care services is increasing at an accelerating pace in developed countries. This development is based on the assumption that patients and clients are able to independently engage with services and experience them as beneficial [1-4]. However, there is an increasing number of vulnerable groups whose capability to engage with services is poor. In addition, some individuals either do not engage at all or choose to disengage from web-based services [2-4]. Thus, the digitalization of health and social care services might lead to a greater digital divide and more exclusion than anticipated because of the perceived burden associated with web-based services among vulnerable groups [2,3].

Various vulnerable groups of people, such as those with mental health disorders, are at risk of digital exclusion. According to Greer et al [5], mental health difficulties, specifically psychosis, impact digital exclusion. For example, a lack of knowledge and deficiencies in financial resources can be significant risk factors for digital exclusion among mental health service users [5]. Farooq et al [6] argued that digital inclusion cannot be considered separately from economic and social inclusion. According to them, it is likely that socially disadvantaged and mentally ill people will lag behind the rest of the population if digital inclusion is not addressed as a priority issue. In addition, other factors such as education level, age, and social deprivation play an important role in estimating who are willing and able to use digital services [7,8].

Attitudes can be seen as one component of digital inclusion [9,10]. According to the Helsper [11] theory, social exclusion increases the risk of digital exclusion via a lack of digital skills, poor access to the internet and digital services, and negative attitudes toward technology. The central role of attitudes has also been highlighted in research on the adoption of new technologies [12]. According to the technology acceptance model [13] and the theory of planned behavior [14], attitudes are one of the key factors in the adoption of and behavioral intention to use technology. Therefore, people's attitudes affect how willing they are to use new technology.

The concept of attitude is multidimensional and difficult to define; however, attitude is typically understood as the valuation of a given target [15]. As people with mental health problems have been found to be critical of face-to-face services being replaced by digital services [16,17], this study focuses first on the ways in which people view digital services compared with face-to-face services. As previous studies have shown that a user's belief in a specific service posing no security or privacy threats is an important factor in the adoption of a web-based service [18,19], this study also examines the perceived trustworthiness of web-based health and social care services. According to Gong et al [20], trust in the provider also affects the perceived benefits and risks of web-based services. Furthermore, the technology acceptance model posits that perceived usefulness is an important factor in the acceptance of new technology [13]. Therefore, as an additional dimension, this study examines attitudes regarding the perceived usefulness of web-based services and the general interest in using them.

In summary, this study approaches the issues of digital inclusion and exclusion from the perspective of attitudes. Specifically, we studied the combined effects of poor mental health and sociodemographic characteristics on attitudes toward web-based health and social care services.

Mental Health, Sociodemographics, and Web-Based Services

Although aggregate digital inclusion has improved overall, the inclusion gap has widened in recent years [10]. Moreover, it has been suggested that the digital exclusion of vulnerable groups will increase rather than decrease in the future [2,3]. The degree of digital literacy and people's ability to use digital services vary within age groups and align with socioeconomic status [2,3]. The greater an individual's offline resources, the more information and communication technology leads to beneficial economic, social, and educational outcomes [2,3,21]. There are several key factors, such as use, income, and education, that discriminate between web-based and offline health information seekers. Indeed, individuals who seek health materials on the web, for example, are likely to be younger, have a better socioeconomic position, and themselves feel happier than those who report only seeking health information offline [22].

A study by Chen and Zhu [8] concluded that youth and high socioeconomic positions are significant predictors of internet access and use. Previous findings also show that older individuals with mental health disorders tend to have fewer skills and financial resources and less familiarity, confidence, and access in engaging with and using digital mental health services than their younger counterparts [23,24]. Moreover, it has been shown that older individuals with mental health disorders do not experience these services as beneficial and do not wish to engage with technology [23,24].

In a Finnish study [25], 89% (40/45) of adolescents with symptoms of depression or anxiety who used a web-based depression support system experienced the support system as reliable and safe. Specifically, the support system was experienced as safe and beneficial when users were able to trust the system's content and felt safe using it. In addition, the criteria for reliability and safety were met when web-based mental health treatment was carried out in a way that protected the confidentiality of adolescent users. A further study on people living with serious mental health illnesses highlights the importance of social media in conducting symptom assessment, as well as in peer support [26].

People belonging to vulnerable groups are more likely to be cynical toward web-based health and social care services and to prefer face-to-face interaction over web-based interactions [16,17,27-29]. The communication technology involved is considered to not convey warmth, empathy, and nonverbal cues at the same depth as direct human contact [16,17,30]. In addition, vulnerable groups have been found to be suspicious of the confidentiality, safety, and privacy of web-based health and social care services [16,17,27-29]. Furthermore, a systematic

review of the acceptability of web-based- and mobile phone–delivered interventions indicated that people with severe mental health problems are generally reluctant to engage with these interventions [28].

A study exploring the reception and use of computerized cognitive behavioral therapy in the treatment of depression among university students concluded that a significant proportion of the participants preferred face-to-face contact alongside computer-assisted therapy [27]. In general, participants also experienced computerized treatment as less credible than face-to-face counseling, and attempting to use a computerized cognitive behavioral therapy program alone might be too laborious a task to undertake when depressed [27]. Similarly, a further qualitative study involving those with severe mental health problems highlighted that participants experienced digital health interventions as merely a complementary method to supplement face-to-face care [17]. Digital health interventions were not perceived as helpful and effective as face-to-face interactions, and some individuals were concerned about the confidentiality of digital services. A Finnish study [16] found that marginalized young people prefer face-to-face interaction over web-based interaction when they want to discuss their problems with a social worker, despite being quite fluent in the use of digital devices for personal purposes. However, older adults have been found to lack confidence and trust in sharing personal information and tend to be suspicious of the ways in which the collected data are processed and their privacy maintained [29].

In addition, previous studies have identified a digital gender gap, for example, in digital skills, and access to and the use of digital services and technologies [31-33]. However, according to Thomas et al [9], gender is not as significant a risk factor for low digital inclusion as, for example, low income, a low level of education, or older age. Attitudes, on the other hand, are the dimensions of digital inclusion in which the gender gap is greatest. Therefore, it is justified to also look at the gender gap when examining digital attitudes.

In general, women have been found to be more likely to search for health-related information on the web than men, but men are more open to engaging in a virtual relationship [34]. Furthermore, previous studies have found significant differences in the help-seeking behaviors of men and women with mental health problems. In fact, males, young people, and people living in affluent areas, for example, were the least likely to seek help [35]. Moreover, according to Smail-Crevier et al [36], women were found to be more likely to use the internet for medical or health-related information and guidance on reducing symptoms of stress and depression, as well as an interactive self-help software program, whereas men preferred to receive information in a video game format.

Women have been found to perceive cognitive behavioral therapy delivered on the web as more acceptable and easier to adhere to than men [37]. However, according to Zhang et al [38], men have a higher level of intention toward mobile health adoption than women. Accordingly, Ellis et al [39] pointed out that the key challenge for web-based mental health services is to design interventions specifically for young men that are action based, focus on shifting behavior and stigma, and are not simply about increasing mental health knowledge. However, not all studies have found gender differences. For example, in the Australian study of Klein and Cook [40], the majority of respondents stated a preference for using traditional mental health services, and no differences were found in relation to any demographic variable or previous use of mental health services between groups of e-preferers and non-e-preferers [41].

Aims

First, previous studies have shown that attitudes are significant factors of influence in the acceptance and use of new technologies [12-14] and in digital exclusion [9,11]. Second, studies have also shown a link between mental health problems and negative attitudes toward web-based services [16,17]. Finally, previous studies suggest that sociodemographic factors influence the use of web-based services [7,8]. However, there is a lack of research on the common effects of mental health problems and sociodemographic factors on attitudes toward web-based services. Moreover, although previous research has examined the moderating effects of socioeconomic characteristics on the association between mental health and health services use [42], web-based services have not received attention in the relevant empirical literature.

This study examined the attitudes toward web-based health and social care services. The measures of attitudes used included *lack of interest*, *perceived need for face-to-face encounters*, and *concern for safety*. In particular, the study analyzed the effects of poor mental health and sociodemographic characteristics (age, gender, education level, and poverty) on these attitudes. Moreover, the moderating effect of sociodemographics on the association between mental health problems and attitudes was examined.

Methods

Sample and Procedure

This cross-sectional study was based on a random sample representative of the Finnish population. In Finland, although almost all people aged <45 years use the internet every day, this use decreases with age. For example, 90% of people aged 45-54 years use the internet every day, compared with 83% of people aged 55-64 years, and only 57% of people aged 65-74 years use the internet every day [43]. The majority of Finns (80%) use mobile phones to access the internet [43]. Although the use of the internet and the skills required to do so are generally common in Finland, there are gaps in digital skills and an obvious need for support among some socially marginalized groups [44].

A sample of 10,000 people was collected from the Population Register Center of Finland. For those aged ≥75 years, a double-picking probability was used to guarantee a sufficient group size. A questionnaire was sent by mail to the participants in 2017. The paper questionnaire form also provided instructions for answering the questionnaire on the web, if preferred. Reminders were sent three times to those who did not respond.

The material is collected by post using a paper form or via a web-based form. Only a very limited number of people have

access to a logistics system in which forms are managed. A secure connection is used for the web-based form, and the answers are stored encrypted in the Finnish Institute for Health and Welfare (THL) database. In the analysis phase, the data obtained from paper forms and electronic forms were combined in a research database. The analyses used pseudonymized research data from which direct personal data were removed. The identification information is stored on the THL's network drive, which can only be accessed by THL employees responsible for managing the material.

The total number of respondents was 4495, yielding a response rate of 44.95% (4495/10,000). The respondents were slightly older, more often women, and had a higher education level than the eligible general population [45]. Inverse probability weighting was applied to address possible bias based on age, gender, marital status, education level, living region, and the degree of urbanization of the residential municipality. This method has previously performed well in correcting the possible effects of nonresponse on the representativeness of the results [46]. Owing to missing values in some of the study variables (Table 1), the analytic sample size ranged from 3200 to 3459, depending on the analysis being carried out.

Table 1. Descriptive statistics (N=4495).

Variable	Value
Age (years), mean (SD)	51.34 (18.38)
Gender, n (%)	4495 (100)
Female	2301 (51.19)
Male	2194 (48.81)
Education, n (%)	4259 (94.75)
High	1328 (31.18)
Average	1440 (33.81)
Low	1491 (35.01)
Poverty, n (%)	4344 (96.64)
Yes	807 (18.58)
No	3537 (81.42)
Poor mental health, n (%)	3853 (85.72)
Yes	569 (14.76)
No	3284 (85.24)
Attitudes (scale 1-5), mean (SD)	3643 (81.04)
Lack of interest	2.13 (1.27)
Need for face-to-face encounters	3.16 (1.13)
Concern for safety	2.62 (1.25)

The study complied with the instructions provided by the Finnish National Board on Research Integrity [47] and the ethical principles of the World Medical Association Declaration of Helsinki [48]. Ethical approval for the study was obtained from the Research Ethics Committee of THL (THL/637/6.02.01/2017).

Measures

Attitudes Toward Web-Based Health and Social Care Services

Although the attitude questions on the form were not directly based on any particular theoretical approach, the questions concerned different aspects of the attitudes toward web-based services identified in previous studies [12,16-20]. When preparing the form, attitude questions were considered from the perspective of learning and quality and, for example, in relation to the information systems success model [49], which emphasizes the importance of information quality, system quality, service quality, user satisfaction, and net benefits for use. In addition, the questions used in the citizens' surveys of other countries were taken into account to allow comparisons between countries.

The respondents were asked, "What factors make it impossible or difficult for you to use online social and health care services?" The response options (rated on a scale from 1=completely disagree to 5=completely agree) measured three types of negative attitudes toward web-based services: *lack of interest* ("Electronic services do not interest me;" Cronbach α=.92), *perceived need for face-to-face encounters* ("Face-to-face encounters cannot be replaced by electronic contacts;" Cronbach α=.81), and *concern for safety* ("I am concerned about data security when it comes to my personal details;" Cronbach α=.91). These three dimensions of limiting attitudes were derived using factor analysis (see Table S1 of Multimedia Appendix 1 [50,51]).

Poor Mental Health

Poor mental health was operationalized using both (1) the Mental Health Inventory (MHI)-5 and (2) responses to a single question enquiring about recent treatment for or diagnosis of depression ("Have you been diagnosed with depression or been treated for depression by a doctor over the past 12 months?"). The MHI-5 [52] is a short 5-item scale that measures feelings of depression and anxiety. More specifically, the MHI-5 measures mental health in the domains of anxiety, general positive affect, depression, and behavioral or emotional control. Respondents who scored ≤52 points on the MHI-5 or indicated that they had been recently diagnosed or treated for depression were classified as having poor mental health.

Sociodemographic Characteristics

Sociodemographic characteristics included age (continuous variable), gender, level of education, and poverty. Education level was measured with the years of education (top coded at 30 years and divided into age-group specific [<26, 26-35, 36-45, 46-55, ≥56 years] tertiles; 1=high, 2=average, and 3=low). Poverty was measured with three items enquiring about not having money for food, medication, or medical treatment. ("Have you within the past 12 months ever feared that you will run out of food before you can get money to buy more"). Responding *yes* to any of the three items was used as an indicator for poverty.

Statistical Analyses

Linear regression analysis was used to examine the associations. Poor mental health and sociodemographic factors (age, gender, education, and poverty) were the independent variables, and the factor scores of the three factors for limiting attitudes were the dependent variables. Each attitude was examined using a separate regression model. The analyses were conducted in three steps for each of the three outcome variables. First, unadjusted analyses were performed to examine the associations between poor mental health and negative attitudes (step 1). Second, we adjusted the analyses for the effects of age, gender, education, and poverty (step 2). Quadratic age was additionally included in the analyses to test for potential nonlinear associations between age and attitudes. Finally, the interactions between poor mental health and sociodemographic characteristics were tested (step 3). All analyses were conducted using sampling weights in R version 3.6.1 (R Foundation for Statistical Computing) [53].

Results

Main Effect Models

Descriptive statistics for the sample are shown in Table 1, and bivariate associations are shown in Table S2 and Figure S1 of Multimedia Appendix 1. On average, the sample was 51.3 (SD 18.38) years, and approximately half were female (2301/4495, 51.19%). The results of the regression analysis (steps 1 and 2) are presented in Table 2. Poor mental health was associated with negative attitudes toward web-based health and social care services, with regard to all three types of attitudes (step 1). The association between poor mental health and negative attitudes was strongest in terms of *concern for safety*. The associations were attenuated but remained significant after adjusting for sociodemographic factors (step 2). In terms of sociodemographic factors, higher age and lower education were associated with all types of negative attitudes. Some of the associations of age were nonlinear, with stronger associations among older respondents. Being a woman and experiencing poverty were (weakly) associated with the *need for face-to-face encounters* and *concerns for safety*.

Table 2. Results of regression analysis.

Predictor	Attitudes					
	Lack of interest		Need for face-to-face encounters		Concern for safety	
	Value	*P* value	Value	*P* value	Value	*P* value
Step 1 (n=3459)						
Poor mental health, β	.24	*<.001* [a]	.34	*<.001*	.38	*<.001*
R^2 (%)	1	N/A[b]	2	N/A	2	N/A
Step 2 (n=3200), β						
Poor mental health	.20	*<.001*	.28	*<.001*	.31	*<.001*
Age	.33	*<.001*	.31	*<.001*	.29	*<.001*
Age^2	.13	*<.001*	.06	*<.001*	.03	.10
Gender (female)	.05	.13	.09	*.008*	.07	*.047*
Education, β						
High	Reference	N/A	Reference	N/A	Reference	N/A
Average	.17	*<.001*	0.11	*<.001*	.17	*<.001*
Low	.42	*<.001*	0.25	*.008*	.30	*<.001*
Poverty (yes), β	.06	.15	0.19	*<.001*	.19	*<.001*
R^2 (%)	23	N/A	14	N/A	13	N/A

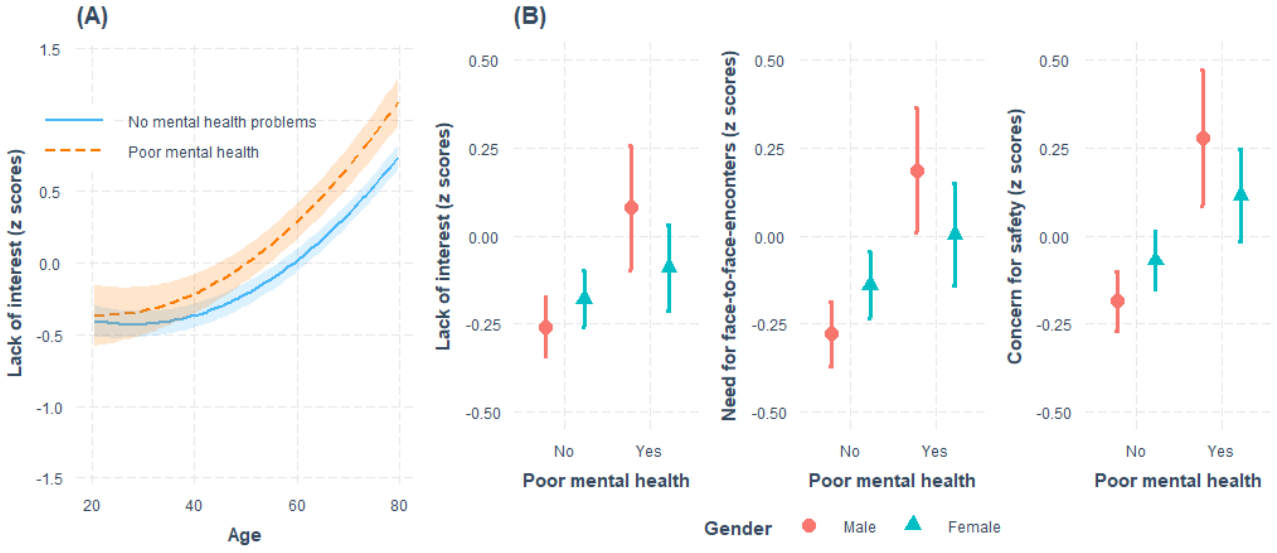
[a]Significant *P* values (*P*<.05) are shown in italics.

[b]N/A: not applicable.

Interactions

Regarding the interactions between poor mental health and sociodemographic factors (step 3), age moderated the association of poor mental health with a lack of interest in web-based services (β=.10; *P*=.03). The association between poor mental health and a lack of interest was stronger among older adults. Gender moderated the association between poor mental health and *lack of interest in web-based services* (β=−.25; *P*=.02), *need for face-to-face encounters* (β=−.32; *P*<.001), and *concerns for safety* (β=−.28; *P*=.01). Compared with women, the association of poor mental health with all types of negative attitudes was stronger among male respondents. These interactions are shown in Figure 1.

Figure 1. The interaction (A) between poor mental health and age and (B) between poor mental health and gender in predicting negative attitudes toward web-based health and social care services.



Discussion

Principal Findings

This study found a significant difference in attitudes toward web-based health and social care services between people with mental problems and others. People with mental health problems were more likely to be concerned about using web-based services and have a lack of interest in them compared with other people. Moreover, they were more likely to consider web-based services as not being able to replace face-to-face encounters. The associations between poor mental health and negative attitudes remained significant, although other factors, such as age, gender, low education level, and poverty, were included in the model. Thus, according to our results, poor mental health is an independent risk factor for negative attitudes toward web-based services. In addition, we found that older age, low education level, and poverty increased the criticality of attitudes. The significance of gender was particularly reflected in the fact that women emphasized the importance of face-to-face encounters more than men.

We found that the effects of low levels of education and poverty were similar among people with mental health problems and among other people. In addition, our study shows that age has a significant effect on attitudes toward web-based health and social care services, both among people with mental health problems and among other people. Moreover, age intensifies the association between poor mental health and attitudes. Among young people, interest in using web-based health and social care services did not depend on mental health conditions. In contrast, older people had more negative attitudes toward web-based services, with poor mental health further increasing this negativity. Furthermore, gender moderated the association of poor mental health with attitudes, and the association of poor mental health with a *lack of interest in web-based services*, the *need for face-to-face encounters*, and *concerns for safety* was stronger among men than among women.

Reflection on Results

Mental health services are increasingly digital, similar to other social and health services. Various web-based therapies have recently been developed, and public mental health services have also begun to be provided on the internet. In addition, free-to-use websites have been built that provide information, tests, and guidelines related to mental health issues. On the basis of the results of this study, it is reasonable to assume that individuals with poor mental health are at risk of digital exclusion and, thus, also social exclusion. However, they were not cohesive groups. Previous studies have identified three major themes that perpetuate digital exclusion among mental health service users: gaps in knowledge, personal circumstances, such as financial barriers, and poor mental health [5]. Consistent with this, this study also highlights the effects of economic factors and mental health on attitudes toward web-based health and social care services and, thus, indirectly also on digital inclusion. However, personal support has been seen as a key factor in improving the digital inclusion of vulnerable people [5,44].

On the basis of this study, it appears that challenges are particularly acute for individuals with mental health problems who have a lower level of education, are older, or are poorer. Therefore, personal support related to the use of services should be targeted specifically for these groups of people. Overcoming the digital exclusion of people with poor mental health also requires addressing wider societal issues related to finance and living circumstances [5].

There are distinctions between age groups because the younger generations have more positive attitudes toward and trust in web-based health services, for example, in the treatment of mental health disorders [25,54]. In fact, the age-related digital divide is not yet narrowing [10]. It can be assumed that new digital tools will especially reach young people with mental health problems. However, for older people and service users with other risk factors, it is also important that the opportunity for face-to-face service continues in the future and that adequate support for the use of digital services is provided. Otherwise, there is a risk that a significant proportion of clients will drop out of mental health services.

On the other hand, the level of education and socioeconomic status play an important role in the adoption of digital services and use of the internet, even among young people who are labeled as technologically more savvy than older generations [7,55]. According to Robards et al [55], marginalized young people particularly need assistance and guidance to use digitalized health care services.

Previous studies have shown that the use of web-based services is related to gender. Although men's attitudes toward internet use and technology are generally seen to be more positive than women's [9], women are more willing to seek help using eHealth services [36,39]. The analyses of this study show that the gender gap is different among people with mental health problems, as well as among other people. This finding is new but not entirely surprising. The general interest in using the internet may also increase men's interest in using general eHealth services. On the other hand, if a person needs to seek help for a mental health problem, for example, the situation is likely to be different because of the stigma associated with the service, among other things. The greater willingness of women to seek professional help [35] can also be a significant factor in the use of eHealth services. However, our findings highlight the need to pay special attention to men with mental health problems when developing eHealth services.

The results show that mental health is related not only to general attitudes but also to more specific attitudes. These results are consistent with those of previous studies. According to previous studies, people with severe mental health problems are reluctant to engage in web-based and mobile phone–delivered interventions [28]. They experience digital health interventions merely as complementary methods to face-to-face care [8]. Correspondingly, a study by Mitchell and Gordon [27] showed that a large proportion of people with depression prefer face-to-face contact alongside computerized cognitive behavior therapy. Granholm [16] has suggested that marginalized young people prefer face-to-face interaction over web-based interaction when they want to discuss their problems with a social worker, even if they are otherwise competent in the use of digital devices. In addition, people with mental health problems have

been shown to be concerned about the confidentiality of digital services, having doubts about issues of data protection and handling [17]. This study found critical attitudes toward replacing face-to-face interactions with web-based interactions among people with mental health problems. Similarly, this study supports the view that vulnerable groups are suspicious of the confidentiality, safety, and privacy of web-based health and social care services. In addition to providing face-to-face support and guidance, issues related to information security and the reliability of web-based services are seen as key to encouraging people with mental health problems to use web-based services.

It is important to improve the usability of these services to promote positive attitudes toward web-based services. This would be especially important for vulnerable groups, such as those with mental health problems and older people. Focusing on better technical solutions, usability, and user needs when designing solutions has been found to lead to more positive user experiences [56]. Moreover, improving the usability of the systems may increase people's confidence in managing their health on the web [57].

To attain an inclusive digital society, web-based health services should meet the particular needs of vulnerable groups. In particular, there should be a wider understanding of specific personal vulnerabilities and how they affect web-based health services [3]. Facilitating digital inclusion among vulnerable groups means helping them develop skills and confidence in using technology [24]. Thus, a conscious, considered, and flexible choice of digital tools and operating environments based on the client's needs is essential in developing an inclusive social and health care service culture [16].

Strengths and Limitations

This study is based on an extensive national survey, which represents the Finnish population quite well. The respondents of this study were slightly older, more often women, and had a higher education than those seen in the eligible population [45]. The analysis was based on Finnish data, and therefore, the results cannot be readily generalized directly to other countries. Finland is also seen as one of the forerunners in the field of digitalization, with the national digital *Kanta* services performing well and tax-financed universal health care for all residents. However, this study did not examine any national specificities in the Finnish social and health care service system, but rather digital services in general, which, in turn, helps to improve the transferability of the results.

As a final consideration, this study relied on self-reported measures, and consequently, the problems associated with an inflation of the strengths of relationships and common method variance may be an issue here. This problem is decreased because the reliability of the measures we used was quite high. However, we cannot rule out the possibility of residual confounding, despite controlling for many factors, including age, gender, education level, and poverty.

Conclusions

With the digitalization of public services, the importance of digital inclusion and avoidance of the digital divide is emphasized. This issue concerns not only digital skills and access to web-based services but also people's attitudes. Thus, supporting the use of web-based services requires both training and influencing people's ways of thinking. From the perspective of digitalization, individuals with mental health problems are not a cohesive group. Low levels of education, older age, and poverty pose risks of digital exclusion. In terms of general attitudes toward web-based health and social care services, people with mental health problems are no different from those of young people. However, people with mental health problems combined with other risks, such as older age, need special support in using web-based services. According to our findings, women's digital attitudes were more negative than men's, especially with regard to their attitude about the importance of face-to-face encounters. On the other hand, among people with poor mental health, men's digital attitudes are more negative than women´s. Thus, it seems that among people with mental health problems, men seem to be more at risk of digital exclusion.

In addition to social and digital support, attention must be paid to social and economic living conditions. In addition, the development of more accessible, user-friendly, and reliable web-based health care and social welfare services can influence people's attitudes toward web-based services, thereby preventing digital and social exclusion. Services that combine digital and face-to-face interactions also play a key role in creating challenges for the competence of workers in various areas of health care and social welfare services. In particular, the results of this study emphasize the importance of guidance and counseling.

Acknowledgments

This research was conducted as part of the *Toward a Socially Inclusive Digital Society: Transforming Service Culture* project and was funded by the Strategic Research Council at the Academy of Finland (grant numbers 327169 and 327145).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Factor analysis to determine the dimensionality of attitudes toward web-based health and social care services.
[DOCX File , 139 KB - jmir_v23i9e28066_app1.docx]

References

1. Lupton D. The digitally engaged patient: Self-monitoring and self-care in the digital health era. Soc Theory Health 2013 Jun 19;11(3):256-270. [doi: 10.1057/sth.2013.10]
2. van Deursen A, Helsper E. The third-level digital divide: who benefits most from being online? In: Communication and Information Technologies Annual (Studies in Media and Communications). Bingley, United Kingdom: Emerald Group Publishing Limited; Dec 12, 2015:29-52.
3. Keeling DI, de Ruyter K, Mousavi S, Laing A. Technology push without a patient pull: examining digital unengagement (DU) with online health services. Eur J Mark 2019 Sep 09;53(9):1701-1732. [doi: 10.1108/ejm-10-2017-0692]
4. Heponiemi T, Jormanainen V, Leemann L, Manderbacka K, Aalto A, Hyppönen H. Digital divide in perceived benefits of online health care and social welfare services: national cross-sectional survey study. J Med Internet Res 2020 Jul 07;22(7):e17616 [FREE Full text] [doi: 10.2196/17616] [Medline: 32673218]
5. Greer B, Robotham D, Simblett S, Curtis H, Griffiths H, Wykes T. Digital exclusion among mental health service users: qualitative investigation. J Med Internet Res 2019 Jan 09;21(1):e11696 [FREE Full text] [doi: 10.2196/11696] [Medline: 30626564]
6. Farooq S, Taylor CD, Gire N, Riley M, Caton N, Husain N. Digital inclusion: the concept and strategies for people with mental health difficulties. Aust N Z J Psychiatry 2015 Sep;49(9):772-773. [doi: 10.1177/0004867415591827] [Medline: 26113744]
7. Hargittai E, Hinnant A. Digital Inequality. Commun Res 2008 Aug 04;35(5):602-621. [doi: 10.1177/0093650208321782]
8. Chen J, Zhu S. Online information searches and help seeking for mental health problems in urban China. Adm Policy Ment Health 2016 Jul;43(4):535-545. [doi: 10.1007/s10488-015-0657-6] [Medline: 25981055]
9. Thomas J, Barraket J, Wilson C, Cook K, Louie Y, Holcombe-James I, et al. Measuring Australia's digital divide: the Australian digital inclusion index 2018. Anal Policy Obs. URL: https://apo.org.au/node/184091 [accessed 2021-08-09]
10. Wilson CK, Thomas J, Barraket J. Measuring digital inequality in Australia: the Australian digital inclusion index. J Telecomm Digit Econ 2019 Jun 30;7(2):102-120. [doi: 10.18080/jtde.v7n2.187]
11. Helsper E. Digital Inclusion: An Analysis of Social Disadvantage the Information Society. London, UK: Department for Communities Local Government; 2008:1-69.
12. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
13. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
14. Ajzen I. Nature and operation of attitudes. Annu Rev Psychol 2001;52:27-58. [doi: 10.1146/annurev.psych.52.1.27] [Medline: 11148298]
15. The psychology of attitudes. A.H. Eagly & S. Chaiken. Fort Worth, TX: Harcourt, Brace, & Janovich, 1993, 794 pp. Reviewed by Christopher Leone, University of North Florida. Psychol Mark 1995;12(5):459-466. [doi: 10.1002/mar.4220120509]
16. Granholm C. Social Work in Digital Transfer : Blending Services for the Next Generation. Helsinki, Finland: Finlandssvenska kompetenscentret inom det sociala området; Apr 15, 2016:1-169.
17. Berry N, Lobban F, Bucci S. A qualitative exploration of service user views about using digital health interventions for self-management in severe mental health problems. BMC Psychiatry 2019 Jan 21;19(1):35 [FREE Full text] [doi: 10.1186/s12888-018-1979-1] [Medline: 30665384]
18. Gao S, Krogstie J, Siau K. Developing an instrument to measure the adoption of mobile services. Mob Inf Syst 2011;7(1):45-67. [doi: 10.3233/MIS-2011-0110]
19. Zhao Y, Ni Q, Zhou R. What factors influence the mobile health service adoption? A meta-analysis and the moderating role of age. Int J Inf Manag 2018 Dec;43:342-350. [doi: 10.1016/j.ijinfomgt.2017.08.006]
20. Gong Z, Han Z, Li X, Yu C, Reinhardt JD. Factors influencing the adoption of online health consultation services: the role of subjective norm, trust, perceived benefit, and offline habit. Front Public Health 2019 Oct 4;7:286 [FREE Full text] [doi: 10.3389/fpubh.2019.00286] [Medline: 31637229]
21. Donkin A. Social gradient. In: The Wiley Blackwell Encyclopedia of Health, Illness, Behavior, and Society. Oxford: Wiley-Blackwell; 2014:2172-2178.
22. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. Soc Sci Med 2004 Nov;59(9):1795-1806. [doi: 10.1016/j.socscimed.2004.02.020] [Medline: 15312915]
23. Ennis L, Rose D, Denis M, Pandit N, Wykes T. Can't surf, won't surf: the digital divide in mental health. J Ment Health 2012 Aug;21(4):395-403 [FREE Full text] [doi: 10.3109/09638237.2012.689437] [Medline: 22712756]
24. Robotham D, Satkunanathan S, Doughty L, Wykes T. Do we still have a digital divide in mental health? A five-year survey follow-up. J Med Internet Res 2016 Nov 22;18(11):e309 [FREE Full text] [doi: 10.2196/jmir.6511] [Medline: 27876684]
25. Anttila K, Anttila M, Välimäki M. A web-based adolescent depression support system: feedback and implications for the future. Inform Health Soc Care 2020;45(2):111-129. [doi: 10.1080/17538157.2019.1656206] [Medline: 31684788]
26. Naslund JA, Aschbrenner KA, Bartels SJ. How people with serious mental illness use smartphones, mobile apps, and social media. Psychiatr Rehabil J 2016 Dec;39(4):364-367 [FREE Full text] [doi: 10.1037/prj0000207] [Medline: 27845533]

27. Mitchell N, Gordon PK. Attitudes towards computerized CBT for depression amongst a student population. Behav Cogn Ther 2007 May;35(4):421-430. [doi: 10.1017/s1352465807003700]
28. Berry N, Lobban F, Emsley R, Bucci S. Acceptability of interventions delivered online and through mobile phones for people who experience severe mental health problems: a systematic review. J Med Internet Res 2016 May 31;18(5):e121 [FREE Full text] [doi: 10.2196/jmir.5250] [Medline: 27245693]
29. Kim TK, Choi M. Older adults' willingness to share their personal and health information when adopting healthcare technology and services. Int J Med Inform 2019 Jun;126:86-94. [doi: 10.1016/j.ijmedinf.2019.03.010] [Medline: 31029268]
30. Kohonen-Aho L. Video-based Observations on the Construction of Social Presence in Virtual Worlds. Espoo, Finland: Alto University; 2017:1-71.
31. Martínez-Cantos JL. Digital skills gaps: a pending subject for gender digital inclusion in the European Union. Eur J Commun 2017 Jul 20;32(5):419-438. [doi: 10.1177/0267323117718464]
32. Davaki K. The underlying causes of the digital gender gap and possible solutions for enhanced digital inclusion of women and girls. European Parliament. URL: https://www.europarl.europa.eu/RegData/etudes/STUD/2018/604940/IPOL_STU(2018)604940_EN.pdf [accessed 2021-02-14]
33. Fatehkia M, Kashyap R, Weber I. Using Facebook ad data to track the global digital gender gap. World Dev 2018 Jul;107:189-209. [doi: 10.1016/j.worlddev.2018.03.007]
34. Bidmon S, Terlutter R. Gender differences in searching for health information on the internet and the virtual patient-physician relationship in Germany: exploratory results on how men and women differ and why. J Med Internet Res 2015 Jun 22;17(6):e156 [FREE Full text] [doi: 10.2196/jmir.4127] [Medline: 26099325]
35. Oliver MI, Pearson N, Coe N, Gunnell D. Help-seeking behaviour in men and women with common mental health problems: cross-sectional study. Br J Psychiatry 2005 Apr;186:297-301. [doi: 10.1192/bjp.186.4.297] [Medline: 15802685]
36. Smail-Crevier R, Powers G, Noel C, Wang J. Health-related internet usage and design feature preference for eMental health programs among men and women. J Med Internet Res 2019 Mar 18;21(3):e11224 [FREE Full text] [doi: 10.2196/11224] [Medline: 30882361]
37. Peynenburg VA, Mehta S, Hadjistavropoulos HD. Postsecondary student perceptions and preferences for the treatment of depression and anxiety: comparison of internet-delivered cognitive behaviour therapy to face-to-face cognitive behaviour therapy and medication. Can J Behav Sci 2020;52(3):220-230. [doi: 10.1037/cbs0000165]
38. Zhang X, Guo X, Lai K, Guo F, Li C. Understanding gender differences in mHealth adoption: a modified theory of reasoned action model. Telemed J E Health 2014 Jan;20(1):39-46. [doi: 10.1089/tmj.2013.0092] [Medline: 24161004]
39. Ellis LA, Collin P, Davenport TA, Hurley PJ, Burns JM, Hickie IB. Young men, mental health, and technology: implications for service design and delivery in the digital age. J Med Internet Res 2012 Nov 22;14(6):e160 [FREE Full text] [doi: 10.2196/jmir.2291] [Medline: 23171827]
40. Klein B, Cook S. Preferences for e-mental health services amongst an online Australian sample? E J Appl Psychol 2010 May 25;6(1):28-39. [doi: 10.7790/ejap.v6i1.184]
41. Musiat P, Goldstone P, Tarrier N. Understanding the acceptability of e-mental health--attitudes and expectations towards computerised self-help treatments for mental health problems. BMC Psychiatry 2014 Apr 11;14:109 [FREE Full text] [doi: 10.1186/1471-244X-14-109] [Medline: 24725765]
42. Goodman E, Huang B. Socioeconomic status, depression, and health service utilization among adolescent women. Womens Health Issues 2001;11(5):416-426. [doi: 10.1016/s1049-3867(01)00077-9] [Medline: 11566284]
43. [Use of the Internet by Finns 2019]. Tilastokeskus. 2019. URL: http://www.stat.fi/til/sutivi/2019/sutivi_2019_2019-11-07_kat_001_fi [accessed 2021-06-21]
44. Järveläinen E, Rantanen T. Incarcerated people's challenges for digital inclusion in Finnish prisons. Nord J Criminol 2020 Sep 14:1-20 (forthcoming). [doi: 10.1080/2578983x.2020.1819092]
45. Hyppönen H, Aalto A. Citizen experiences of e-health and e-welfare services. In: Vehko T, Ruotsalainen S, Hyppönen H, editors. E-health and E-welfare of Finland Check Point 2018. Helsinki: National Institute for Health and Welfare (THL); Jul 2019:148-167.
46. Härkänen T, Kaikkonen R, Virtala E, Koskinen S. Inverse probability weighting and doubly robust methods in correcting the effects of non-response in the reimbursed medication and self-reported turnout estimates in the ATH survey. BMC Public Health 2014 Nov 06;14:1150 [FREE Full text] [doi: 10.1186/1471-2458-14-1150] [Medline: 25373328]
47. Responsible conduct of research and procedures for handling allegations of misconduct in Finland. Finnish National Board of Research Integrity Tenk. 2012. URL: http://www.tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf [accessed 2021-02-14]
48. WMA declaration of Helsinki – ethical principles for medical research involving human subjects. World Medical Association. URL: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ [accessed 2021-02-14]
49. Delone W, McLean E. The DeLone and McLean Model of information systems success: a ten-year update. J Manag Inf Syst 2014 Dec 23;19(4):9-30. [doi: 10.1080/07421222.2003.11045748]
50. Horn JL. A rationale and test for the number of factors in factor analysis. Psychometrika 1965 Jun;30(2):179-185. [doi: 10.1007/BF02289447] [Medline: 14306381]

51. Muthén LK, Muthén BO. Mplus Statistical Analysis with Latent Variables: User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén; 2015:1-876.
52. Stewart AL, Hays RD, Ware JE. The MOS short-form general health survey. Reliability and validity in a patient population. Med Care 1988 Jul;26(7):724-735. [doi: 10.1097/00005650-198807000-00007] [Medline: 3393032]
53. R: a language and environment for statistical computing. R Foundation for Statistical Computing. URL: http://softlibre.unizar.es/manuales/aplicaciones/r/fullrefman.pdf [accessed 2021-08-09]
54. Metsäranta K, Kurki M, Valimaki M, Anttila M. How do adolescents use electronic diaries? A mixed-methods study among adolescents with depressive symptoms. J Med Internet Res 2019 Feb 20;21(2):e11711 [FREE Full text] [doi: 10.2196/11711] [Medline: 30785408]
55. Robards F, Kang M, Usherwood T, Sanci L. How marginalized young people access, engage with, and navigate health-care systems in the digital age: systematic review. J Adolesc Health 2018 Apr;62(4):365-381. [doi: 10.1016/j.jadohealth.2017.10.018] [Medline: 29429819]
56. Kujala S, Ammenwerth E, Kolanen H, Ervast M. Applying and extending the FITT framework to identify the challenges and opportunities of successful ehealth services for patient self-management: qualitative interview study. J Med Internet Res 2020 Aug 12;22(8):e17696 [FREE Full text] [doi: 10.2196/17696] [Medline: 32784175]
57. Lee M, Kang D, Yoon J, Shim S, Kim I, Oh D, et al. The difference in knowledge and attitudes of using mobile health applications between actual user and non-user among adults aged 50 and older. PLoS One 2020 Oct 27;15(10):e0241350 [FREE Full text] [doi: 10.1371/journal.pone.0241350] [Medline: 33108792]

Abbreviations

MHI: Mental Health Inventory
THL: Finnish Institute for Health and Welfare

Edited by R Kukafka; submitted 19.02.21; peer-reviewed by W Buente, C Jacob; comments to author 18.05.21; revised version received 22.06.21; accepted 05.07.21; published 21.09.21.

Please cite as:
Rantanen T, Gluschkoff K, Silvennoinen P, Heponiemi T
The Associations Between Mental Health Problems and Attitudes Toward Web-Based Health and Social Care Services: Evidence From a Finnish Population-Based Study
J Med Internet Res 2021;23(9):e28066
URL: https://www.jmir.org/2021/9/e28066
doi:10.2196/28066
PMID:34546184

Original Paper

Just-in-Time Adaptive Mechanisms of Popular Mobile Apps for Individuals With Depression: Systematic App Search and Literature Review

Gisbert W Teepe[1], MSc; Ashish Da Fonseca[2], MA; Birgit Kleim[3,4], ProfDr; Nicholas C Jacobson[5], PhD; Alicia Salamanca Sanabria[6], PhD; Lorainne Tudor Car[7,8], PhD; Elgar Fleisch[1,2], ProfDr; Tobias Kowatsch[1,2,6], PhD

[1]Centre for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

[2]Centre for Digital Health Interventions, Institute of Technology Management, University of St. Gallen, St. Gallen, Switzerland

[3]Experimental Psychopathology and Psychotherapy, Department of Psychology, University of Zurich, Zurich, Switzerland

[4]Department of Psychiatry, Psychotherapy and Psychosomatics, University of Zurich, Zurich, Switzerland

[5]Center for Technology and Behavioral Health, Departments of Biomedical Data Science and Psychiatry, Geisel School of Medicine, Dartmouth College, Hanover, NH, United States

[6]Future Health Technologies, Singapore-ETH Centre, Campus for Research Excellence And Technological Enterprise, Singapore, Singapore

[7]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[8]Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom

Corresponding Author:
Gisbert W Teepe, MSc
Centre for Digital Health Interventions
Department of Management, Technology, and Economics
ETH Zurich
Weinbergstrasse 56/58
Zurich, 8092
Switzerland
Phone: 41 76 419 09 91
Email: gteepe@ethz.ch

Abstract

Background: The number of smartphone apps that focus on the prevention, diagnosis, and treatment of depression is increasing. A promising approach to increase the effectiveness of the apps while reducing the individual's burden is the use of just-in-time adaptive intervention (JITAI) mechanisms. JITAIs are designed to improve the effectiveness of the intervention and reduce the burden on the person using the intervention by providing the right type of support at the right time. The right type of support and the right time are determined by measuring the state of vulnerability and the state of receptivity, respectively.

Objective: The aim of this study is to systematically assess the use of JITAI mechanisms in popular apps for individuals with depression.

Methods: We systematically searched for apps addressing depression in the Apple App Store and Google Play Store, as well as in curated lists from the Anxiety and Depression Association of America, the United Kingdom National Health Service, and the American Psychological Association in August 2020. The relevant apps were ranked according to the number of reviews (Apple App Store) or downloads (Google Play Store). For each app, 2 authors separately reviewed all publications concerning the app found within scientific databases (PubMed, Cochrane Register of Controlled Trials, PsycINFO, Google Scholar, IEEE Xplore, Web of Science, ACM Portal, and Science Direct), publications cited on the app's website, information on the app's website, and the app itself. All types of measurements (eg, open questions, closed questions, and device analytics) found in the apps were recorded and reviewed.

Results: None of the 28 reviewed apps used JITAI mechanisms to tailor content to situations, states, or individuals. Of the 28 apps, 3 (11%) did not use any measurements, 20 (71%) exclusively used self-reports that were insufficient to leverage the full potential of the JITAIs, and the 5 (18%) apps using self-reports and passive measurements used them as progress or task indicators only. Although 34% (23/68) of the reviewed publications investigated the effectiveness of the apps and 21% (14/68) investigated their efficacy, no publication mentioned or evaluated JITAI mechanisms.

Conclusions: Promising JITAI mechanisms have not yet been translated into mainstream depression apps. Although the wide range of passive measurements available from smartphones were rarely used, self-reported outcomes were used by 71% (20/28) of the apps. However, in both cases, the measured outcomes were not used to tailor content and timing along a state of vulnerability or receptivity. Owing to this lack of tailoring to individual, state, or situation, we argue that the apps cannot be considered JITAIs. The lack of publications investigating whether JITAI mechanisms lead to an increase in the effectiveness or efficacy of the apps highlights the need for further research, especially in real-world apps.

(J Med Internet Res 2021;23(9):e29412) doi:10.2196/29412

KEYWORDS

depression; digital mental health; smartphone applications; just-in-time adaptive interventions; effectiveness; mobile phone

Introduction

Background

Globally, each year, at least 246 million people are affected by depression [1], and depression is the leading cause of years lived with disability [2]. Although effective treatments for depression exist [3-5], most individuals in need still do not receive effective treatment [6], or those obtaining treatment do not benefit. People seeking help often face barriers such as the high costs of treatment, a shortage of trained clinicians, the stigma associated with seeking help, and accessibility difficulties [7-10].

Mobile apps may have the potential to address the rising prevalence of depression and the insufficient resources available for treatment [11,12]. Apps are already an integral part of most people's everyday lives [13], and the threshold for engagement with apps is assumed to be low, resulting in prompt, flexible, portable, and anonymous treatment for people with depression [14]. Individuals otherwise not reachable could receive treatment [15], and the interventions could be delivered in economies that have limited resources for mental health treatment [16]. Small to large effect sizes have been reported by several systematic reviews that show that apps and other digital interventions reduce symptoms of mental health problems, including depression [17-20]. Finally, apps can be used in real-life situations, where behavior change is most desirable and clinicians are unable to intervene [14].

Objective

This study aims to complement the existing assessment of apps that address depression by focusing on the use of just-in-time adaptive intervention (JITAI) mechanisms [21,22]. JITAIs aim to deliver an adaptive treatment (ie, personalized or tailored) at a time of vulnerability (ie, "person's transient tendency to experience adverse health outcomes or to engage in maladaptive behaviors" [21]) and receptivity (ie, "the person's transient tendency to receive, process, and use the support provided" [21]). To be considered a JITAI, an intervention needs to deliver the content when a state of vulnerability and state of receptivity are detected. Vulnerability refers to *the right time,* that is, a moment of heightened exposure or receptiveness for a negative health outcome [23]. For example, a JITAI that addresses depression should deliver an intervention component when changes in mood are detected. Ideally, the component delivered should aim to improve the symptom that has been detected. The treatment must not only be delivered at *the right time* (*vulnerability)* but the recipient also needs to be *receptive* (*receptivity)*. Receptivity refers to *the availability*, that is, the detection of the time when an individual can receive, process, and use the support provided [23]. For a JITAI that addresses depression, this could be the time when the individual is at home because delivery of a behavior activation component needs a safe space.

The measured outcomes used to tailor content and timing along a state of vulnerability and state of receptivity are referred to as tailoring variables [21,22]. Although ecological momentary assessments may facilitate the detection of these tailoring variables, passive measurements (eg, using the location derived from a smartphone's GPS data) are regarded as the gold standard of tailoring variables for JITAIs. These passive measurements have the advantage of enabling unobtrusive, continuous observation [23]. Although tailoring content and timing to the state of receptivity and state of vulnerability may be possible by asking ecological momentary assessments (eg, asking through message, "Would now be a good time to start an activity?"), the benefit of passive measurements may be that the content and timing are tailored without deliberate interaction with the intervention itself (eg, detecting a state of vulnerability and state of receptivity and sending a notification, "It seems that you are at home and in a bad mood. Would you like to start an exercise?"). Therefore, it has been proposed that JITAIs that tailor the content to the person, situation, and time by using these passive measurements reduce the burden on users and increase the effectiveness of the intervention [21,22].

Evidence of the higher effectiveness of JITAIs compared with that of non-JITAI treatments and wait-list control groups was investigated in a recent meta-analysis [24], which found moderate to large effect sizes (Hedges g=1.65 when compared with wait-list control and Hedges g=0.89 when compared with non-JITAI treatments) of the primary outcomes produced by 33 empirical studies. Owing to the potential of JITAIs to increase effectiveness while reducing the burden on users and the prominence of the JITAI framework in the scientific community, the aim of this study is to determine the degree to which popular apps that address depression use JITAI mechanisms by reviewing which relevant symptoms of depression (eg, mood) are measured as well as how they are measured. We are also interested in learning whether peer-reviewed studies can be found that investigated the increased effectiveness or efficacy of these apps attributable to the use of JITAI mechanisms.

To this end, we systematically assessed popular apps that target depression, that is, apps that have the most reviews on the Apple

App Store and most downloads on the Google Play Store. We argue that the investigation of JITAI mechanisms is necessary because of their potential to increase effectiveness while simultaneously decreasing the burden on users. The focus on popular apps is important because they are listed at the top of search results and thus are very likely to be downloaded and used [25]. Moreover, a high number of downloads implies that they have been found to be useful by users [26] and may indicate that people continue to use or recommend them. Recent evidence also indicates that the 2 most popular apps for depression and anxiety were responsible for 90% of the 13.3 million active monthly users (Headspace with 6,959,000 users equivalent to 52% of the active users and Calm with 5,000,000 users equivalent to 38% of the active users) [27]. Although the review by Wasil and Gillespie [27] focused on depression and anxiety apps, similar user statistics should be observed when focusing only on apps for depression.

Methods

Search Strategy and Selection Criteria

We conducted this systematic app search and literature review following the methods used in existing reviews [26,28-30] of popular apps that address mental health problems. We systematically identified and reviewed apps that were publicly available in the US and UK app stores because number of venture capital backed digital health deals in the first half of 2020 in these countries ranked first (United States: 475) and second (United Kingdom: 41) of all English-speaking countries [31]. The Apple App Store and Google Play Store were used because they have a combined market share of approximately 99.4% [32]. We searched the two stores by entering the term *depression* in the search fields of the stores and included all apps found in both stores of both countries. We also reviewed curated lists of health apps from prominent organizations, namely the Anxiety and Depression Association of America [33], the National Health Service [34], and the American Psychological Association [35], to ensure that we did not miss any app recommended by important institutions and experts for mental health. The apps found in these lists addressed several different mental health problems. We selected only those apps that addressed depression for further assessment. The searches were carried out in August 2020. All apps found regardless of the category in which they were found (eg, lifestyle), download category, or number of reviews were included in the initial list of apps. This list of all the apps found was screened according to the inclusion and exclusion criteria outlined below.

For further assessment, we included the apps that targeted the treatment of depression or reduction of the symptoms of depression by delivering at least one active ingredient and were available in English. We defined an active ingredient, on the lines of the study by Michie et al [36], as a function supporting users in their management of depression that is designed to reliably and causally change processes that govern behavior [37]. An example of an active ingredient for depression could be a goal-setting task, a breathing exercise, or a recording of daily mood. Apps targeting other mental health illnesses such as anxiety or posttraumatic stress disorder were not excluded if depression was addressed as well. We included both free and paid apps. Browser-based treatments were not included. We excluded apps that only targeted professionals (eg, Depression Psychopharmacology), only offered a diagnostic service (eg, Patient Health Questionnaire-9 Depression Test Questionnaire), only provided quotes or inspirational text (eg, Depression Quote Wallpapers), or only conveyed information without the goal of eliciting behavior change or engaging with individuals (eg, Psychology Book: 1000+ Amazing Psychology Facts).

Keeping the inclusion and exclusion criteria in mind, 2 authors (GWT and ADF) separately reviewed each app identified from the initial search for depression apps in the two stores and the curated lists. The interrater agreement was excellent (Cohen κ=0.91). In case of disagreements, a consensus was reached through discussion. After this initial assessment, we ranked all included apps from the Apple App Store separately by their number of reviews and all included apps from the Google Play Store by their download category (eg, 1,000,000+ and 500,000+ downloads). We decided to screen the apps before ranking them to ensure that the most popular apps for depression—and not the most popular apps from our initial lists—were included in the analysis.

Next, we separately identified the most popular apps available only in the Apple App Store, available only in the Google Play Store, or available in both stores. For apps only available in the Apple App Store, we selected the 5 most reviewed apps because users rarely scroll past the first 5 apps on the list [25]. For the Google Play Store, we used the download category of the app ranked fifth on the list (eg, 500,000+ downloads). All apps in the 500,000+ download category were included. For apps available in both app stores, we used the Google Play Store's 500,000+ download category to determine inclusion, regardless of the number of reviews on the Apple App Store. Regardless of the number of downloads or reviews, we included all apps from the curated lists that met the inclusion criteria and did not violate the exclusion criteria.

Data Analysis

Our evaluation covered the following areas: general information about the app, potential mechanisms for delivery of the JITAI, and peer-reviewed evidence. We developed our evaluation framework before reviewing the apps and used Covidence Systematic Review software (Veritas Health Innovation Ltd; version accessed in August 2020) to review the apps. All the questions are listed in the codebook in Tables S1-S6 of Multimedia Appendix 1 [21,23,38-48], along with the sources from which we derived them. Each included app was evaluated separately by 2 raters (GWT and ADF) as follows:

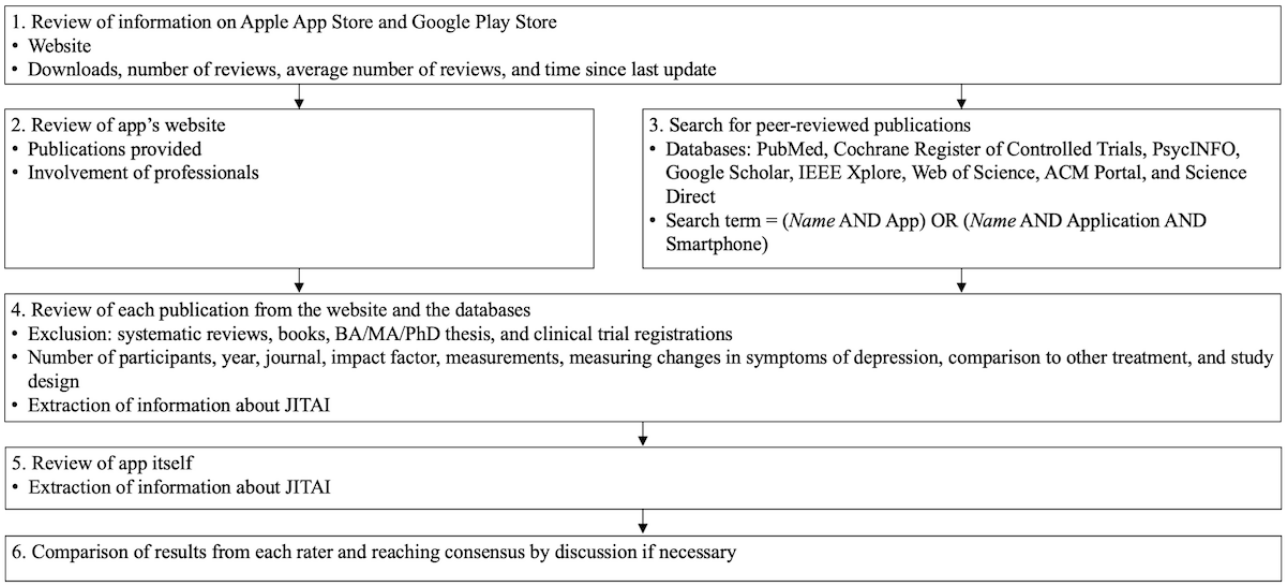
First, we gathered general information about the apps, including the name of the provider, additional affiliated organizations (eg, other companies, universities, governments, or nongovernmental organizations), and time since last update. Second, we reviewed the app's website and recorded all the publications provided as well as information about JITAIs. Third, we searched for peer-reviewed publications in PubMed, Cochrane Register of Controlled Trials, PsycINFO, Google Scholar, IEEE Xplore, Web of Science, ACM Portal, and Science Direct using the

search term *[(Name AND App) OR (Name AND Application AND Smartphone)]*.

Fourth, we reviewed the full text of each study found on the website and in the different databases. We excluded books, theses, systematic reviews evaluating several different apps, and clinical trial registrations. Subsequently, each study was evaluated in line with prior work [38], including the year of publication, journal name, journal impact factor, the number of subjects, study purpose, and study design (ie, randomized controlled trial or open trial). We also extracted the information available about *the JITAI mechanism*. We determined the degree to which the apps could be considered *JITAIs* by reviewing whether, and to what degree, the relevant features derived from the JITAI concept described in the study by Nahum-Shani et al [22] had been implemented. In the first step, we reviewed how potential tailoring variables were measured. In the second step, we reviewed whether these measurements were used to determine a state of vulnerability (ie, a moment of heightened exposure or receptiveness for a negative health outcome) or state of receptivity (ie, the time when the input could be perceived, processed, and used). To leverage the potential of JITAIs, the state of vulnerability and state of receptivity not only need to be detected through measurements but the treatment content and timing also need to be tailored along these measurements. We assessed how the tailoring variables were measured by reviewing the *symptoms of depression* that were measured (derived from the *International Classification of Diseases, 10th Revision* and *Diagnostic and Statistical Manual of Mental Disorders, Edition 5*) and the *self-report* data or *sensor and device analytics* (derived from related work [23] from the Android Developers Guide [39] and the iPhone operating system Security Guide [40]) that were used. We also reviewed whether tailoring to traits (ie, "tailoring-to-people" [21]) was used by checking for questions about demographics and socioeconomic status. Furthermore, we checked whether receptivity was measured by GPS location, self-reports, or by measuring symptoms. As it has been proposed that JITAI mechanisms increase the effectiveness or efficacy of such apps [22], we reviewed whether the publications addressed effectiveness or efficacy and whether JITAI mechanisms were investigated in these publications.

Finally, we reviewed the app itself and extracted the information available about JITAI mechanisms. The results from each rater were compared, and consensus was reached by discussion, if necessary. We reviewed each app in September 2020, and the process is illustrated in Figure 1.

Figure 1. Review process for each app. BA: bachelor; JITAI: just-in-time adaptive intervention; MA: master; PhD: doctor of philosophy.

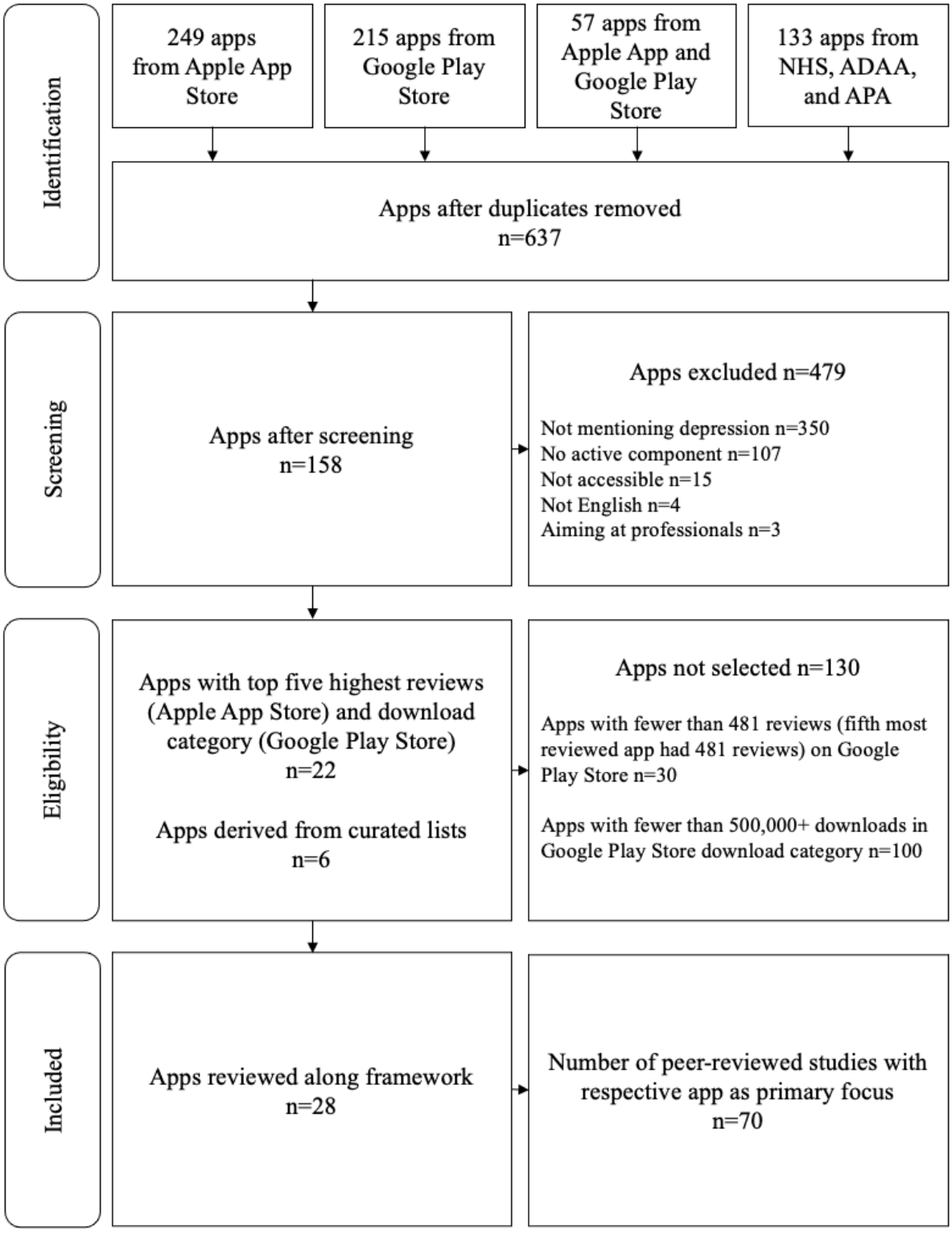


Results

Apps

We found 249 apps in the Apple App Store, 217 apps in the Google Play Store, 57 apps in both stores, and 135 apps in the curated list, yielding a total of 658 apps. Of the 658 apps, we removed 17 (2.6%) duplicates, 349 (53%) apps that did not mention depression, 123 (18.7%) apps with no active component, 8 (1.2%) apps that were not accessible, 1 (0.2%) app not available in English, and 1 (0.2%) app that targeted professionals. The reasons for excluding each app are presented in Table S7 of Multimedia Appendix 1. The reasons for excluding apps because of missing active ingredients were also summarized and are presented in Table S8 of Multimedia Appendix 1 with further details. We ranked the apps found only in the Apple App Store based on their number of reviews and included the 5 most reviewed apps. We ranked the remaining apps found in the Google Play Store and in both stores according to their number of downloads category. The fifth most downloaded app in the Google Play Store fell in the 500,000+ download category. Therefore, we included all apps found in the Google Play Store and all apps available in both Apple App Store and Google Play Store with more than 500,000+ downloads, yielding 17 apps. We included 6 apps from the curated lists that met the inclusion criteria of mentioning depression and did not violate the exclusion criteria, yielding a total of 28 apps. A flowchart of the results of the review process is presented in Figure 2.

Figure 2. Inclusion and exclusion process of reviewed apps. ADAA: Anxiety and Depression Association of America; APA: American Psychological Association; NHS: National Health Service.



Using methods similar to those used in the study by Baumel et al [49], we calculated the number of users and reviews covered by our selection of the top 5 reviewed apps from the Apple App Store and apps with more than 500,000 downloads found in the Google Play Store and apps found in both the Apple App Store and Google Play Store. The calculations are presented in Tables S9 and S10 of Multimedia Appendix 1 (Table S9 for apps ranked by their download category and Table S10 for apps ranked according to their number of reviews). Using the cutoff point of the 5 most reviewed apps, we included 91% of the reviews of the 35 apps found only in the Apple App Store that met the inclusion criteria. Including the sixth most reviewed app would have resulted in an increase of 1.97% of the reviews in the sample. By including apps with more than 500,000 downloads (ie, the 500,000+ download category), we included 96.7% of the users of the 118 apps found in the Google Play Store and

both Google Play Store and Apple App Store that remained after screening. Including the download category of more than 100,000 downloads would have resulted in an increase of 3% in the downloads in the sample.

Publications

We found 70 peer-reviewed publications for the 28 reviewed apps (Multimedia Appendix 1 TableS12) [50-119]. We found at least one publication for 57% (16/28) of the apps, at least one peer-reviewed publication investigating the effectiveness of 32% (9/28) of the apps, and at least one peer-reviewed publication investigating the efficacy of 18% (5/28) of the apps. Of the 70 publications, although 23 (33%) investigated the effectiveness of the apps, and 14 (20%) investigated their efficacy, no publication evaluated an increase in effectiveness or efficacy achieved by using JITAI mechanisms. The extracted information from all the reviewed publications is presented in Tables S11 and S12 of Multimedia Appendix 1.

Included Apps

The 28 apps included were rated 2,808,465 times, with each app being rated on average 100,302.32 times (SD 279,723.5; range 127-1,229,623; median 8,109, IQR 28,310). The average rating for all apps was 4.44 stars (SD 0.34; range 3.40-4.95; median 4.50, IQR 0.37) out of a possible five stars. The number of reviews and average reviews were calculated using weighted means with information from the Apple App Store and Google Play Store, if available. Table 1 lists the general information and JITAI mechanisms of each reviewed app and Table 2 summarizes the 28 apps that were included in the review and evaluated.

Table 1. General information and JITAI[a] mechanisms of the reviewed apps (N=28).

App	Operating system	Number of reviews	AVR[b]	NODC[c]	Symptoms measured and measurements used	NDS[d] used	NSR[e] used	NSDA[f] used
Calm	AND[g], iOS[h]	1,229,623	4.71	10,000,000+	Activity, mood, unhelpful beliefs, sleep, open questions, multiple-choice questions	4	2	—[i]
Headspace	AND, iOS	872,025	4.86	10,000,000+	—	—	—	—
Daylio	AND, iOS	328,048	4.62	10,000,000+	Activity, mood, appetite, sleep, ratings, multiple-choice questions	4	2	—
Youper	AND, iOS	61,478	4.67	1,000,000+	Activity, mood, interest or pleasure, sleep, vital signs, location, open questions, ratings, multiple-choice questions	4	3	2
Moodpath	AND, iOS	57,011	4.64	1,000,000+	Activity, mood, interest or pleasure, appetite, unhelpful beliefs, concentration, sleep, vital signs, open questions, ratings, multiple-choice questions	7	3	1
Wysa	AND, iOS	56,520	4.70	1,000,000+	Activity, mood, interest or pleasure, unhelpful beliefs, concentration, ratings, multiple-choice questions	5	2	—
Friend Shoulder	AND	32,847	4.30	1,000,000+	Mood, unhelpful beliefs, suicidal thoughts, open questions	3	1	—
BetterHelp	AND, iOS	30,592	4.63	500,000+	Mood, unhelpful beliefs, suicidal thoughts, open questions	3	1	—
Sanvello	AND, iOS	27,536	4.67	1,000,000+	Activity, mood, interest or pleasure, unhelpful beliefs, suicidal thoughts, sleep, microphone, open questions, ratings	6	2	1
7 Cups	AND, iOS	20,709	4.29	1,000,000+	Mood, interest or pleasure, cognition, unhelpful beliefs, distorted perception, open questions, ratings	5	2	—
Control and Monitor	AND	16,149	4.50	1,000,000+	Activity, mood, appetite, unhelpful beliefs, suicidal thoughts, sleep, ratings, multiple-choice questions	6	2	—
#SelfCare	AND, iOS	14,670	4.57	500,000+	—	—	—	—
Remente	AND, iOS	11,494	4.34	1,000,000+	Activity, mood, open questions, ratings, multiple-choice questions	2	3	—
Reflexio	AND	8118	4.30	1,000,000+	Mood, open questions, ratings	1	2	—
Moodnotes	iOS	8100	4.70	—	Mood, illogical thinking, camera, open questions, ratings, multiple-choice questions	2	3	1
Online Therapy—Mental Help	AND	5979	4.40	500,000+	Mood, unhelpful beliefs, suicidal thoughts, open questions	3	1	—
InnerHour	AND, iOS	5402	4.50	500,000+	Activity, mood, interest or pleasure, unhelpful beliefs, concentration, suicidal thoughts, sleep, ratings, multiple-choice questions	8	2	—

App	Operating system	Number of reviews	AVR[b]	NODC[c]	Symptoms measured and measurements used	NDS[d] used	NSR[e] used	NSDA[f] used
Happify	AND, iOS	5164	4.21	500,000+	Activity, mood, interest or pleasure, cognition, unhelpful beliefs, sleep, illogical thinking, distorted perception, vital signs, camera, open questions, ratings, multiple-choice questions	8	3	2
What's Up?—A Mental Health App	AND, iOS	3446	4.22	500,000+	Activity, mood, interest or pleasure, open questions, ratings, multiple-choice questions	3	3	—
MoodTools—Depression Aid	AND, iOS	3167	4.31	100,000+	Activity, mood, interest or pleasure, appetite, cognition, unhelpful beliefs, concentration, suicidal thoughts, sleep, distorted perception, ratings, multiple-choice questions, Patient Health Questionnaire	10	3	—
DBT[j] Coach	AND, iOS	3067	4.95	10,000+	Activity, mood, interest or pleasure, appetite, unhelpful beliefs, concentration, suicidal thoughts, sleep, ratings, multiple-choice questions, Patient Health Questionnaire	8	3	—
CBT[k] Thought Diary	AND, iOS	2182	4.58	100,000+	Mood, distorted perception, open questions, ratings, multiple-choice questions	2	3	—
T2 Mood Tracker	AND, iOS	1873	3.40	100,000+	Activity, mood, cognition, unhelpful beliefs, concentration, sleep, open questions, ratings	6	2	—
Joyable	AND, iOS	1522	4.48	5000+	Activity, mood, interest or pleasure, appetite, unhelpful beliefs, suicidal thoughts, sleep, open questions, ratings, multiple-choice questions	7	3	—
Breeze	iOS	1313	4.70	—	Activity, mood, ratings, multiple-choice questions	2	2	—
Moodkit	iOS	159	4.40	—	Activity, mood, appetite, sleep, open questions, ratings, multiple-choice questions	4	3	—
Catch It	AND, iOS	144	3.66	50,000+	Mood, open questions, ratings, multiple-choice questions	1	3	—
Feeling Good	AND, iOS	127	3.91	10,000+	—	—	—	—

[a]JITAI: just-in-time adaptive intervention.

[b]AVR: average rating out of five possible stars.

[c]NODC: number of downloads category.

[d]NDS: number of depression symptoms.

[e]NSR: number of self-reports.

[f]NSDA: number of sensors and device analytics.

[g]AND: Android operating system.

[h]iOS: Apple operating system.

[i]App did not measure a symptom, did not use a self-report, or did not use sensors and device analytics.

[j]DBT: dialectical behavior therapy.

[k]CBT: cognitive behavioral therapy.

Table 2. Summary of the number of reviews, average rating, download category, and JITAI[a] mechanisms used by the reviewed apps (N=28).

	Sum	Count	Value, mean (SD; range)	Value, median (IQR)
Reviews	2,808,465	28	100,302.32 (279,723.50; 127-1,229,623)	8109 (28,310)
Rating[b]	—[c]	28	4.44 (0.34; 3.40-4.95)	4.50 (0.37)
Download category	42,375,000	25	1,695,000 (3,153,748.43; 5000-10,000,000)	500,000 (900,000)
Depression symptoms	114	25	4.56 (2.47; 1-10)	4 (3)
Self-reports	59	25	2.36 (0.7; 1-3)	2 (1)
Sensors and device analytics	7	5	1.4 (0.55; 1-2)	1 (1)

[a]JITAI: just-in-time adaptive intervention.

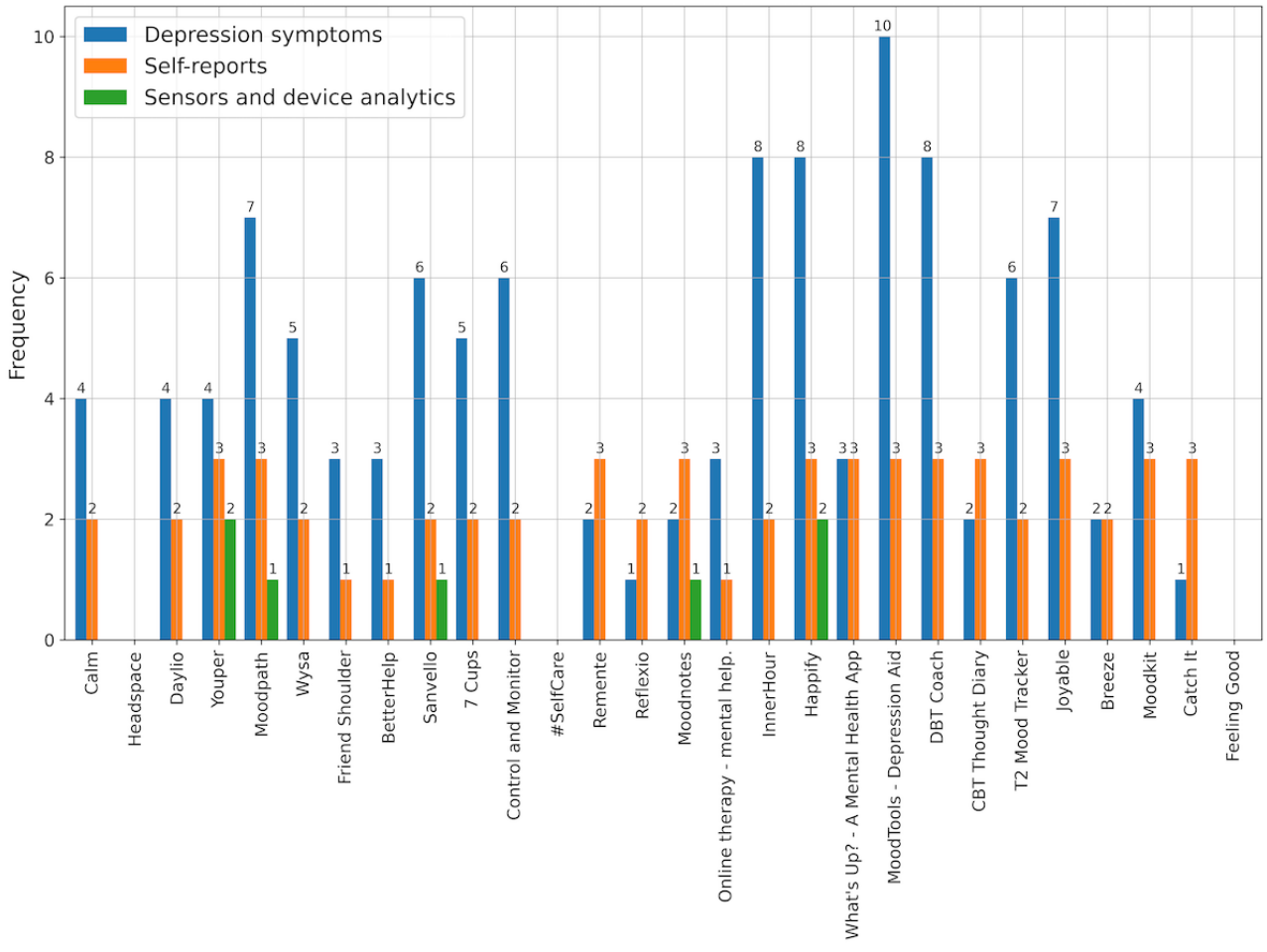
[b]Out of five possible stars.

[c]Not available.

JITAI Mechanisms

We found that, of the 28 reviewed apps, 25 (89%) measured some kind of *depression symptom* when users interacted with the app (eg, initial assessment when starting the app), whereas 3 (11%) did not use any measurements, 20 (71%) used at least one *self-report* (eg, daily report of mood changes through a rating), and 5 (18%) used *self-reports* and *sensors and device analytics* (eg, tracking activity through location data). Figure 3 illustrates the number of *depression symptoms* measured by different *self-reports* or *sensors and device analytics* for each of the reviewed apps. *Mood Tools—Depression Aid* measured the most *depression symptoms* (10 different symptoms measured) while not using any *sensors and device analytics*. *Happify* and *Youper* measured fewer *depression symptoms* (eight and four, respectively) but used two different *sensors and device analytics*.

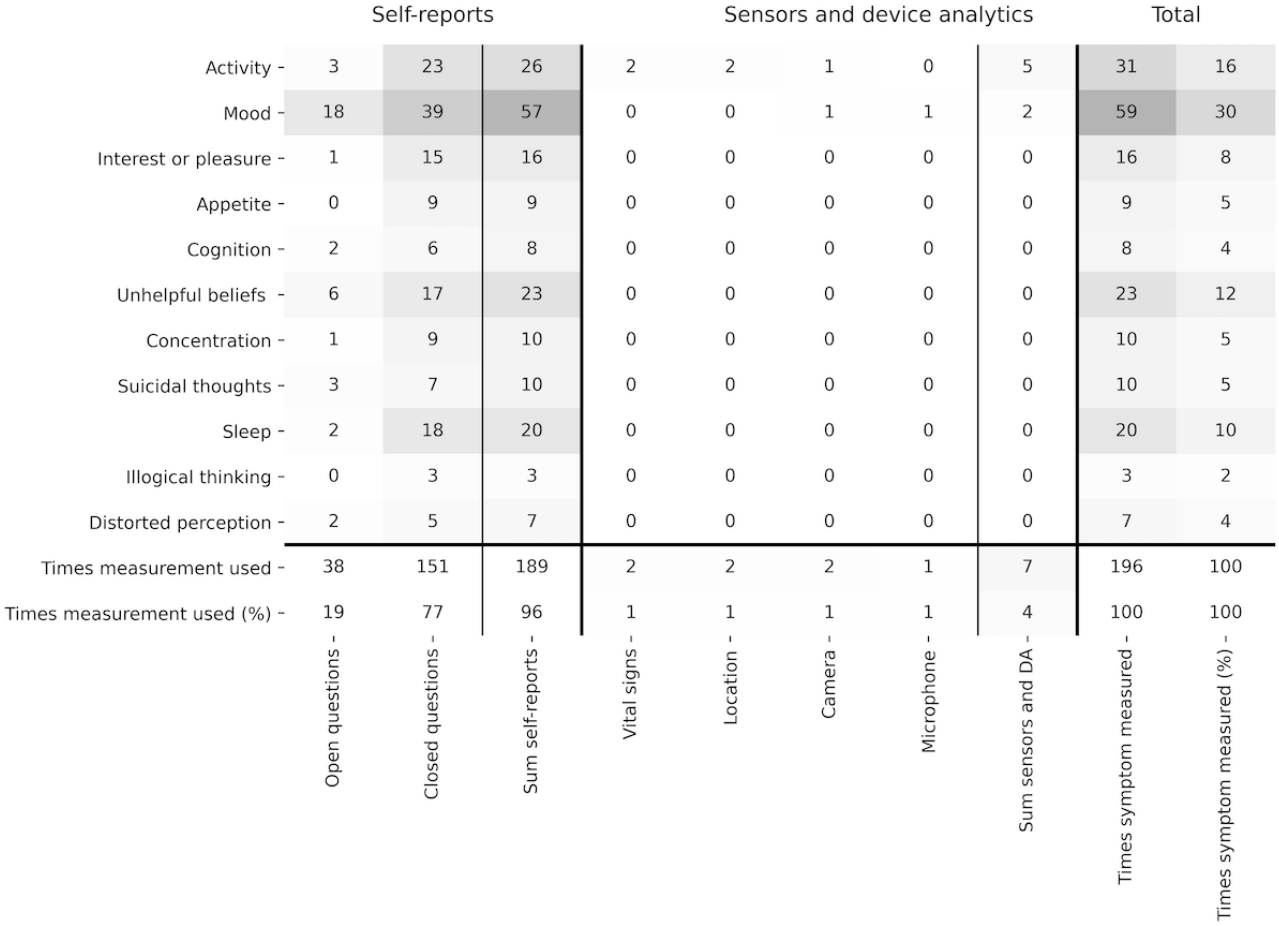
Figure 3. Depressive symptoms measured and frequency of measurements used for each of the 28 reviewed apps. CBT: cognitive behavioral therapy; DBT: dialectical behavior therapy.



Our findings regarding the use of *self-reports* and *sensors and device analytics* are summarized in Figure 4. We found that a symptom was measured by a *self-report* or *sensors and device analytics* 196 times. For example, the app *Calm* measured four

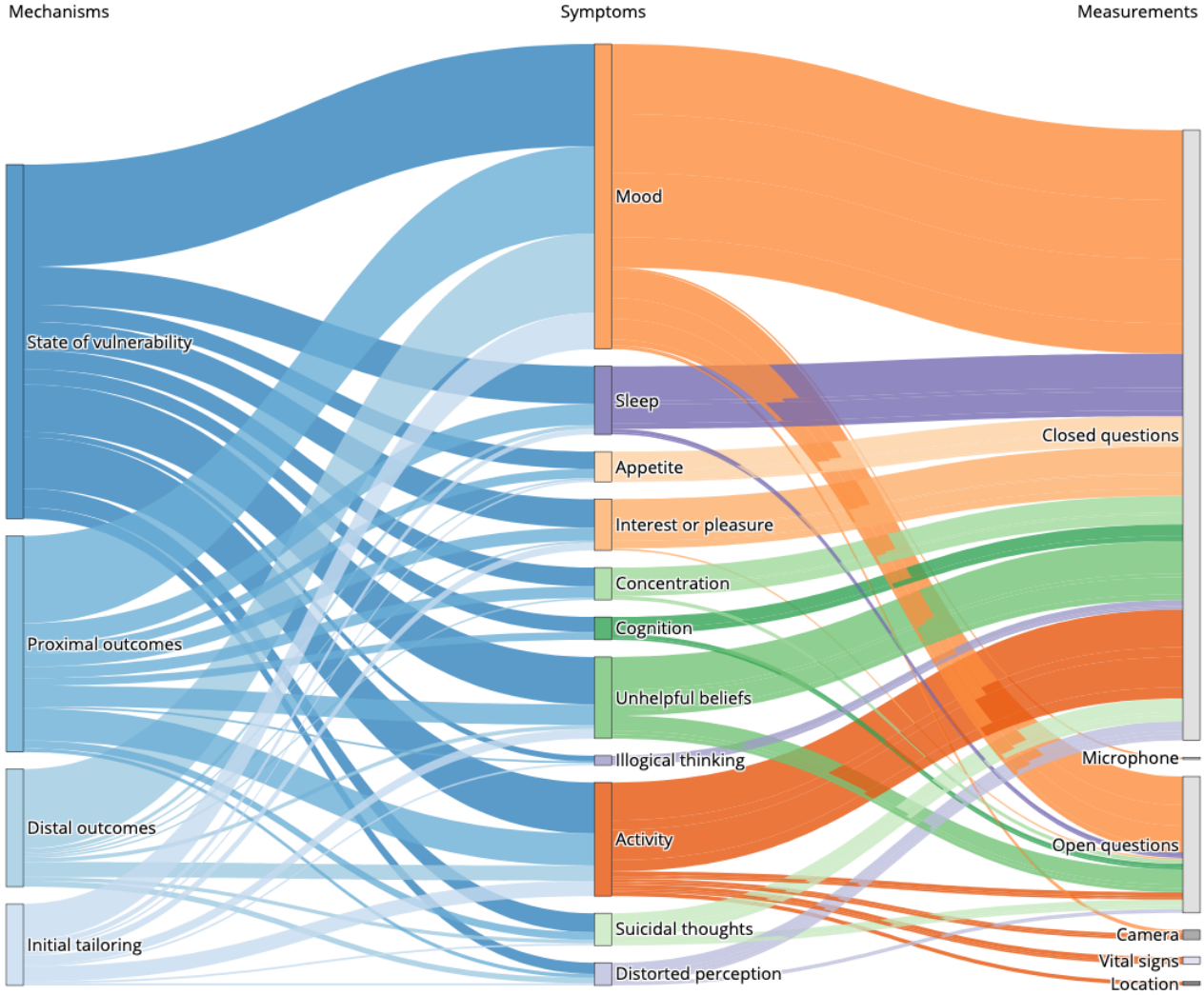
symptoms (activity, mood, unhelpful beliefs, and sleep) by using two different self-reports for each symptom (open questions and multiple-choice questions), resulting in a symptom measurement rate of 4.1% (8/196). To measure different *depressive symptoms*, *self-reports* were used almost exclusively (189/196, 96.4%), and *sensors and device analytics* were used rarely (7/196, 3.6%). The s*elf-reports* used most frequently to measure different *depressive symptoms* were *closed questions* consisting of ratings (eg, rate your mood from 1 to 10), Likert scales (eg, indicate how often you felt sad last week, with the following options: never, almost never, sometimes, most of the time, and always), and multiple-choice questions (eg, select the activities that made you happy in the last week) for a measurement rate of 77% (151/196). The measurement rate for *open questions* (ie, questions without a fixed response or text field with open input) was 19.4% (38/196). Of the 3.6% (7/196) using *sensors and device analytics*, the most frequently used were *vital signs* (mostly heart rate), *location* and camera with each 0.5% (1/196) measurement rate. The measurement rate for the symptom *mood* (most frequently measured symptom) was 30.1% (59/196), for *activity* 15.8% (31/196), for *unhelpful beliefs* 11.7% (23/196), and for s*leep* 10.2% (20/196).

Figure 4. Heatmap of measurements used to measure symptoms. The heatmap illustrates the number of times that symptoms of depression were measured by self-reports or sensors and device analytics, summarized over the 28 reviewed apps. A darker color illustrates a higher number of occurrences, also indicated by the annotation in the cells. DA: device analytics.



When possible, we tried to match the measurement of the depressive symptom to a mechanism relevant to the JITAI concept. We were able to do so for s*tate of vulnerability*, *proximal outcomes*, *distal outcomes*, and *initial tailoring*. Some of the measurements could have been used as tailoring variables for two or three *JITAI* mechanisms. Therefore, it is possible that double counting of the symptoms and measurements for each mechanism occurred. Figure 5 illustrates our findings, including which measurements were used to measure which symptom and for which *JITAI* feature. The figure shows that some s*ensors and device analytics* were not used as passive measurements; rather, they were used to actively capture changes. For example, the *camera* was used as a measurement for *activity*, with users being asked to take pictures of locations that they had visited or a picture of something that made them sad to describe their *mood*.

Figure 5. Connection among just-in-time adaptive intervention (JITAI) mechanisms, symptoms, and measurements. Sankey diagram illustrating the different JITAI mechanisms (state of vulnerability, proximal outcomes, distal outcomes, and tailoring variables) for which we were able to match a depressive symptom (eg, mood) and the measurements used to capture the changes (eg, closed question). The JITAI mechanisms are displayed in blue; depressive symptoms in orange, green, and purple; and measurements in gray. The size of the rectangle indicates the number of times that the mechanism, symptom, or measurement was found. The thickness of the connection indicates the number of times that a measurement or symptom was used. Some measurements have been assigned to two or three JITAI mechanisms; therefore, double counting is possible.



Discussion

Principal Findings

We reviewed the 28 most popular or recommended apps for depression found in the Apple App Store, Google Play Store, and in curated lists compiled by respected mental health authorities. Regarding our main aim to investigate JITAI mechanisms, we found that none of the reviewed apps specifically mentioned the use of JITAI mechanisms in the app, on their websites, or in the identified peer-reviewed publications. Of the 28 apps, we found that 3 (11%) did not use any measurements, and 20 (71%) only used *self-reports* (189/196, 96.4% of all measurements were self-reported). Although such *self-reports* can be used as *in-the-moment assessments* (ie, ecological momentary assessments) that are closely related to the JITAI concept [24], we argue that they are insufficient for leveraging the full potential of JITAIs. Of the 28 apps, we found that 5 apps (18%) also used *sensors anddevice analytics* (7/196, 3.6% of all measurements involved *sensoranddevice analytics*). However, we found that most *sensors* such as the *camera* were used as *in-the-moment* assessments or as part of an app's features; they were not used to tailor the content or timing. Some of the apps measured *depressive symptoms* by using *self-reports* when the app was first opened to determine what kind of content should be presented (eg, measuring the need to focus on sleep or mood). However, these measurements were not used to tailor the content that was presented afterward. Furthermore, static tailoring, which has been observed to be less effective than dynamic tailoring [120], is, in our view, not sufficient for an app to be considered as a JITAI. Relying on this initial assessment may also be insufficient because symptoms of depression seem to change frequently through the day and may be highly specific to an individual [121].

Interestingly, we found that besides *mood* (symptom measurement rate of 59/196, 30.1%) and *decreased activity* (symptom measurement rate of 31/196, 15.8%), other symptoms of depression were measured less frequently. Given the broad variety and severity of depression [122] and the high comorbidity with other mental health problems such as anxiety [123], this focus on a subset of symptoms may not be sufficient

to detect changes that might indicate a need for support. In addition, a focus on the improvement of the main symptoms (eg, *mood* and *activity*, derived from the *Diagnostic and Statistical Manual of Mental Disorders, Edition 5*) may not be sufficient to contribute to the understanding of the complex processes involved in depression. Accurate and continuous measurements of psychophysiological changes enabled by passive measurements of various physiological features (eg, changes in breathing patterns or vital signs) may, however, improve the understanding of depression in general. Such an understanding could, in turn, enable an even more successful implementation of JITAIs.

Related work has shown that user adherence to various mental health apps dropped substantially after interacting with the app for only a couple of days [49]. Although details of user engagement with the apps reviewed in the study by Baumel et al [49] were not available, several apps that were included in our study were also included in this review (eg, *Headspace*, *Calm*, *Daylio*, *7Cups*, *Happify*, *Moodpath*, and *T2 Mood Tracker*). These reviewed apps showed a drastic decline in user numbers after a couple of days, highlighting the importance of finding ways to increase user adherence. A proposed key advantage of a JITAI is to address this challenge of declining adherence by offering support in times of need and reducing the burden on the individual. An app that solely relies on self-reports entered by the user while interacting with the app cannot achieve this goal. Individuals would need to open the app and engage with it to enter the responses before tailoring to the right time could be carried out. Continuously requiring such self-reports may even be counterproductive to adherence because of the increase in the burden. Therefore, future work should address the extent to which JITAIs that tailor content along passive measurements are capable of increasing adherence. A comparison among apps that do not use any measurements, that use static tailoring to measurements, that use tailoring to self-reports only, or that use tailoring to both self-reports and passive measurements would be necessary to determine the benefit of dynamic tailoring with regard to increasing user adherence.

Considering self-reported questions further by focusing on open questions, the question arises as to what degree the answers to these questions could be used to tailor content without any further use of analytics. Although self-reported closed questions already suffer from some biases (eg, recall bias and biases due to mental illness), this shortcoming is addressed by using different questions to infer an underlying construct and to validate questionnaires in different samples. In contrast, open questions may still be limited to measuring symptoms of depression and need to rely on methods derived from natural language processing to cope consistently and accurately with the large amount of data entered. Contrary to closed-question questionnaires, these methods have not been validated as rigorously as existing questionnaires for symptoms and symptom severity. Therefore, it is unclear to what degree open questions can be used to measure symptoms of depression. Existing work using free text has focused on using data from other sources to measure symptom changes. In a recent review by Chancellor and De Choudhury [124], 75 such studies were investigated. In these studies, social media data were used to infer various mental health problems. The authors share our concern regarding consistency and accuracy by concluding that construct validity and lack of reflection may limit the ability to use text to measure health status. In contrast, passive measurements that could be derived when users answer open questions, such as keystroke frequency or pressure on display, have been used to measure cognitive functions associated with mental health [125].

As stated in the Introduction section, the JITAI concept aims to tailor content and timing based on the measured state of vulnerability and state of receptivity. Our results indicate that the measurements used in the reviewed apps may only serve for tailoring to the state of vulnerability (eg, mood and activity). We did not find any measurement that may serve to tailor to the state of receptivity. Plausible examples of such receptivity tailoring are still conceivable for both self-reports (eg, asking whether an individual is alert enough to read a text or complete a demanding exercise) and passive measurements (eg, detecting the user's arrival at home through GPS data and time of day to trigger a journaling exercise). In general, studies investigating possible markers to detect receptivity and the effectiveness of such receptivity tailoring are sparse. This research gap becomes even more evident when one considers the various studies that aim to develop passive measurements that could be used to determine vulnerability [121,126-132] together with the finding that *tailoring to what* had a significant impact on the effectiveness of JITAIs [24].

Other concepts related to receptivity, such as interruptibility [133], have been investigated in greater detail [134]. In accordance with the study by Künzler et al [135], we would still argue that receptivity extends these concepts by combining concepts such as the willingness to receive an intervention (interruptibility), engage with the intervention, and process the provided information. To the best of our knowledge, only the following studies have specifically focused on investigating receptivity. The study by Künzler et al [135] developed machine learning models to predict receptivity in a physical activity intervention. In the study by Choi et al [136], the authors describe the results of a 3-week study that they conducted to investigate the context and cognitive or physical state to understand the response to the components of sedentary behavior. The study by Sarker et al [137] used the response to ecological momentary assessments—but not to an intervention component—to explore discriminative features for building machine learning models to detect receptivity and reported 77.9% accuracy. Although these studies used post hoc analysis to determine receptivity, the study by Mishra et al [134] used data from the study by Künzler et al [135] to test the two different models in a real-world study. The authors found that receptivity increased by 40% compared with the control when a static machine learning model was used and that a dynamic machine learning model led to an increase in receptivity over time, whereas receptivity in the control conditions declined. These results are promising because they show that with relatively easily derived information such as time of day or Wi-Fi connection, receptivity and engagement can be maintained, or engagement can even be increased over time,

which is something that mental health interventions have been observed to struggle with [49].

The finding that most measurements could have been used to tailor to the state of vulnerability but not to the state of receptivity may be explained by three considerations. First, as shown in the study by Wang and Miller [24], the *tailoring to what* (ie, both previous behavioral patterns of people and their current need states) showed significant improvement. However, results regarding the *tailoring to when* have not been reported. Second, most related work aimed at developing markers for detecting changes in symptom severity or depressive severity [121,126-132]. Although some of these outcomes may also be relevant for the state of receptivity, the primary use of these measurements would be to determine the state of vulnerability (eg, mood changes detected by voice), distal outcome (eg, depression severity score calculated by combining different symptom outcomes), and proximal outcome (eg, sleep quality measured through vital signs). Finally, as mentioned above, it seems that receptivity as a concept has been investigated less often, with related work focusing on concepts such as interruptibility [133] or detection of boredom [138]. Future work should focus on transferring the findings on contextual factors such as interruptibility or boredom by using the day of the week [138], time of day [138], location [137], or different proxies for social context (eg, number of calls and messages sent or received) [138] to investigate tailoring to the state of receptivity.

Taken together, these findings highlight that although the JITAI concept seems to be widely known in the scientific digital health community [24] and different studies outline the possibility of detecting changes in depression or *depressive symptoms* such as mood by using different passive measurements [121,126-132], these mechanisms have not been implemented in the real world, aside from baseline or progress assessments. Although tailoring to both the state of vulnerability and state of receptivity may be challenging, it is still surprising that no JITAI mechanisms have been implemented by the reviewed apps. This finding is especially unexpected given the existing work on passive sensing and the popularity of the JITAI concept.

Related to these findings, we were interested in the degree to which the effectiveness and efficacy of the included apps were investigated in peer-reviewed publications because it has been proposed that JITAIs increase the effectiveness or efficacy of such apps. None of the 70 reviewed publications investigated JITAI mechanisms. Therefore, our findings highlight that the proposed increase in effectiveness or efficacy achieved by using JITAIs is not evaluated in real-world app settings. In addition, we found great variability in the scientific evidence provided for the reviewed apps—despite an increased interest in digital health—in several publications addressing this topic, especially within the last 5 years [26,139,140].

Limitations

The strengths of this study are the large number of apps initially screened; the analyses using a framework developed from existing work; and the rigorous methodology used to review all identified studies addressing the apps, the apps' websites, and the apps themselves. However, this study includes several limitations. We reviewed the apps at a single point, which is a shortcoming found in related studies as well. We are aware that the app stores are dynamic with constant changes [25], but a long-term review of the apps would not have been feasible. We may address this issue in our future work. Besides the lists we reviewed from the Anxiety and Depression Association of America, National Health Service, and American Psychological Association, rating systems or lists of reviewed apps are offered by other organizations, too. These include, but are not limited to, the American Psychiatric Association, PsyberGuide, and iMedicalApps. We did not review these lists because we expected a high number of overlaps and because not all the apps found in the lists had been reviewed (eg, Dartmouth PATH had not been reviewed on PsyberGuide when last checked [January 27, 2021]), and not all the apps mentioned in the lists had been recommended (eg, Mood Watch Review, with low credibility, user experience, and transparency ratings on PsyberGuide when last checked [January 27, 2021]). However, we see the value in a central platform for reviews of mental health apps and suggest incorporating the findings regarding the use of JITAI mechanisms into the existing review criteria.

We did not search for apps developed by research groups as was done in a recent meta-analytic review by Wang and Miller [24] that evaluated studies addressing the effectiveness of JITAIs. Their search included, but was not limited to, apps. Although this work is highly relevant for the field, we argue that a review investigating the degree to which tailoring is used in popular apps was necessary. This notion was derived from findings highlighting that academia has not always kept up with the high speed of digital health developments in the real world [41]. In addition, apps developed for research purposes may not be available to the public, limiting the impact of these apps to a small number of individuals participating in the review. Our aim was to investigate whether popular apps that reach many individuals use these promising mechanisms. Finally, the review of the apps initially included other aspects such as the use of evidence-based treatment, conversational agents, and the revenue model. Reporting these findings would have exceeded the scope of this review.

Comparison With Prior Work

We found 11 reviews investigating different aspects of the apps that address depression [14,20,26,27,29,141-145]. Of these 11 reviews, 6 (45%) assessed the content or features of the apps [14,20,26,29,141,145]; of these 6 reviews, 1 (17%) adjusted its analysis to the number of users [27]. A meta-analysis focused on the efficacy of apps for depression, indicating a moderate positive effect when the apps were compared with inactive control and a small effect when comparing the apps with active control [20]. The remaining studies investigated usability [143], adherence to clinical guidelines [144], claims [142], or data sharing and privacy practice [142]. A meta-analytic review investigated the effect sizes of JITAIs compared with control groups or other interventions, but it did not focus on apps or mental health [24]. This meta-analytic review indicated that two aspects of tailoring—*tailoring to what* and the tailoring *approach*—were significantly associated with greater JITAI efficacy. We found no study investigating the use of JITAI mechanisms or reviewing the measurements used to capture

changes when discussing the relevant features of such apps. Furthermore, we did not find any studies reviewing whether real-world apps provide evidence of an improvement in their effectiveness or efficacy achieved by using JITAI mechanisms.

Conclusions

In conclusion, our findings indicate that because of the limited use of measurements for depressive symptoms, except for self-reports as indicators of progress or initial tailoring, the 28 most popular or recommended apps that address depression cannot be considered JITAIs. Although self-reports and, occasionally, passive measurements were used to capture changes in depressive symptoms, these changes were not used to dynamically tailor the content to a state of vulnerability or state of receptivity. An increase in app effectiveness or efficacy achieved by using JITAI mechanisms was also not evaluated in any of the reviewed publications. Therefore, future work should focus on evaluating the effectiveness of real-world JITAIs that tailor their content and timing to the state of vulnerability and state of receptivity. On the basis of these findings, we argue that the reviewed apps do not yet leverage the full potential of digital health interventions by providing tailored support when it is most needed and in the most helpful way.

Acknowledgments

The authors would like to thank Jacqueline Louise Mair for reviewing the manuscript and providing feedback. This work was partially supported by the National Institute of Mental Health (R01 MH123482) and CSS Insurance, Switzerland. This research was also partially supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise program and was in part conducted at the Future Health Technologies program which was established collaboratively between ETH Zurich and the National Research Foundation Singapore. This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise program. The National Institute of Mental Health, the National Research Foundation, and CSS Insurance had no role in the study design, data collection, data analysis and interpretation, writing the manuscript, or reviewing and approving the manuscript for publication. The reviewed apps and reviewed studies are publicly available. The extracted data used in this review will be shared with full open access beginning 3 months and ending 24 months after the publication of the study. For further questions or material requests, please contact GWT.

Authors' Contributions

GWT designed the evaluation framework using inputs from TK and ADF. GWT and TK designed and implemented a search strategy. GWT and ADF screened and coded the apps, websites, and studies, and extracted the data. GWT analyzed the data and drafted the initial manuscript, supervised by TK. BK, NCJ, ASS, LTC, and EF provided methodological guidance and feedback on the manuscript. All authors reviewed and approved the final manuscript.

Conflicts of Interest

GWT, EF, and TK are affiliated with the Centre for Digital Health Interventions, a joint initiative of the Department of Management, Technology, and Economics at ETH Zurich and the Institute of Technology Management at the University of St. Gallen, which is funded in part by the Swiss health insurer CSS. EF and TK are also cofounders of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, Pathmate Technologies was not involved in this study. NCJ and Dartmouth College are the owners of a depression and anxiety app named Mood Triggers. However, owning Mood Triggers is not a financial conflict of interest, given that Mood Triggers is not intended to be revenue generating; rather, it is used to deliver and evaluate no-cost scalable treatment treatments using just-in-time adaptive interventions (JITAIs).

Multimedia Appendix 1
Codebook for app review, calculations of downloads and reviews covered, number of studies and type of studies found for each app, and list of all reviewed publications.
[DOCX File , 168 KB - jmir_v23i9e29412_app1.docx]

References

1. Depression and other common mental disorders: global health estimates. World Health Organization. 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf [accessed 2021-08-07]
2. James S, Abate D, Abate K, Abay S, Abbafati C, Abbasi N, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet 2018 Nov;392(10159):1789-1858. [doi: 10.1016/S0140-6736(18)32279-7]
3. Cuijpers P, Stringaris A, Wolpert M. Treatment outcomes for depression: challenges and opportunities. Lancet Psychiatry 2020 Nov;7(11):925-927. [doi: 10.1016/S2215-0366(20)30036-5] [Medline: 32078823]

4. Delgadillo J, Duhne P. Targeted prescription of cognitive-behavioral therapy versus person-centered counseling for depression using a machine learning approach. J Consult Clin Psychol 2020 Jan;88(1):14-24. [doi: 10.1037/ccp0000476] [Medline: 31841021]
5. Delgadillo J, Rubel J, Barkham M. Towards personalized allocation of patients to therapists. J Consult Clin Psychol 2020 Sep;88(9):799-808. [doi: 10.1037/ccp0000507] [Medline: 32378910]
6. Kazdin AE, Blase SL. Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspect Psychol Sci 2011 Jan;6(1):21-37. [doi: 10.1177/1745691610393527] [Medline: 26162113]
7. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychol Med 2013 Aug 09;44(6):1303-1317. [doi: 10.1017/s0033291713001943]
8. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
9. Global health workforce, finances remain low for mental health. World Health Organization. 2015. URL: https://www.who.int/news/item/14-07-2015-global-health-workforce-finances-remain-low-for-mental-health [accessed 2021-08-07]
10. Wilhelm S, Weingarden H, Ladis I, Braddick V, Shin J, Jacobson NC. Cognitive-behavioral therapy in the digital age: presidential address. Behav Ther 2020 Jan;51(1):1-14 [FREE Full text] [doi: 10.1016/j.beth.2019.08.001] [Medline: 32005328]
11. Fairburn CG, Patel V. The impact of digital technology on psychological treatments and their dissemination. Behav Res Ther 2017 Jan;88:19-25 [FREE Full text] [doi: 10.1016/j.brat.2016.08.012] [Medline: 28110672]
12. Kazdin AE. Addressing the treatment gap: A key challenge for extending evidence-based psychosocial interventions. Behav Res Ther 2017 Jan;88:7-18. [doi: 10.1016/j.brat.2016.06.004] [Medline: 28110678]
13. The Lancet Child and Adolecent Health. Growing up in a digital world: benefits and risks. Lancet Child Adol Health 2018 Feb;2(2):79. [doi: 10.1016/s2352-4642(18)30002-6]
14. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019;2:118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
15. Ebert DD, Van Daele T, Nordgreen T, Karekla M, Compare A, Zarbo C, et al. Internet- and mobile-based psychological interventions: applications, efficacy, and potential for improving mental health efficacy, and potential for improving mental health. Eur Psychol 2018 May;23(2):167-187. [doi: 10.1027/1016-9040/a000318]
16. Chisholm D, Sweeny K, Sheehan P, Rasmussen B, Smit F, Cuijpers P, et al. Scaling-up treatment of depression and anxiety: a global return on investment analysis. Lancet Psychiatry 2016 May;3(5):415-424 [FREE Full text] [doi: 10.1016/S2215-0366(16)30024-4] [Medline: 27083119]
17. Andrews G, Basu A, Cuijpers P, Craske M, McEvoy P, English C, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
18. Deady M, Choi I, Calvo RA, Glozier N, Christensen H, Harvey SB. eHealth interventions for the prevention of depression and anxiety in the general population: a systematic review and meta-analysis. BMC Psychiatry 2017 Aug 29;17(1):310 [FREE Full text] [doi: 10.1186/s12888-017-1473-1] [Medline: 28851342]
19. Ebert DD, Zarski A, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. PLoS One 2015;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
20. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
21. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. Health Psychol 2015 Dec;34 Suppl:1209-1219. [doi: 10.1037/hea0000306] [Medline: 26651462]
22. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. Ann Behav Med 2016 Sep 23:446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
23. Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 2018 Jan;77:120-132. [doi: 10.1016/j.jbi.2017.12.008] [Medline: 29248628]
24. Wang L, Miller LC. Just-in-the-Moment Adaptive Interventions (JITAI): A meta-analytical review. Health Commun 2019 Sep 05:1-14. [doi: 10.1080/10410236.2019.1652388] [Medline: 31488002]
25. Dogruel L, Joeckel S, Bowman ND. Choosing the right app: An exploratory perspective on heuristic decision processes for smartphone app selection. Mobile Media Commun 2014 Dec 29;3(1):125-144. [doi: 10.1177/2050157914557509]
26. Baumel A, Torous J, Edan S, Kane JM. There is a non-evidence-based app for that: A systematic review and mixed methods analysis of depression- and anxiety-related apps that incorporate unrecognized techniques. J Affect Disord 2020 Aug 01;273:410-421. [doi: 10.1016/j.jad.2020.05.011] [Medline: 32560936]

27. Wasil AR, Gillespie S, Patel R, Petre A, Venturo-Conerly KE, Shingleton RM, et al. Reassessing evidence-based content in popular smartphone apps for depression and anxiety: Developing and applying user-adjusted analyses. J Consult Clin Psychol 2020 Nov;88(11):983-993. [doi: 10.1037/ccp0000604] [Medline: 32881542]
28. Shen N, Levitan M, Johnson A, Bender JL, Hamilton-Page M, Jadad AA, et al. Finding a depression app: a review and content analysis of the depression app marketplace. JMIR Mhealth Uhealth 2015;3(1):e16 [FREE Full text] [doi: 10.2196/mhealth.3713] [Medline: 25689790]
29. Qu C, Sas C, Roquet C, Doherty G. Functionality of top-rated mobile apps for depression: systematic search and evaluation. JMIR Ment Health 2020 Jan 24;7(1):e15321. [doi: 10.2196/15321]
30. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: A comprehensive quality and therapeutic potential evaluation tool for mobile and web-based ehealth interventions. J Med Internet Res 2017 Mar 21;19(3):e82 [FREE Full text] [doi: 10.2196/jmir.7270] [Medline: 28325712]
31. Number of venture capital backed digital health deals in select countries in HY1 2020. Statista. 2020. URL: https://www.statista.com/statistics/1063913/vc-backed-digital-health-deals-in-select-countries/ [accessed 2021-08-07]
32. Mobile operating system market share worldwide. statcounter Global Stats. 2021. URL: https://gs.statcounter.com/os-market-share/mobile/worldwide [accessed 2020-04-04]
33. ADAA Reviewed mental health apps. Anxiety and Depression Association of America (ADAA). 2020. URL: https://adaa.org/finding-help/mobile-apps [accessed 2020-07-04]
34. NHS Apps Library. National Health Service. 2020. URL: https://www.nhs.uk/apps-library/ [accessed 2020-07-04]
35. Mobile apps and websites. American Psychological Associatin. 2020. URL: https://www.apa.org/careers/early-career/useful-apps?tab=3 [accessed 2020-07-04]
36. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: The CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664]
37. Carey RN, Connell LE, Johnston M, Rothman AJ, de Bruin M, Kelly MP, et al. Behavior change techniques and their mechanisms of action: a synthesis of links described in published intervention literature. Ann Behav Med 2019 Jul 17;53(8):693-707 [FREE Full text] [doi: 10.1093/abm/kay078] [Medline: 30304386]
38. Safavi K, Mathews SC, Bates DW, Dorsey ER, Cohen AB. Top-funded digital health companies and their impact on high-burden, high-cost conditions. Health Aff (Millwood) 2019 Jan;38(1):115-123. [doi: 10.1377/hlthaff.2018.05081] [Medline: 30615535]
39. Sensors overview. Developers Android. 2020. URL: https://developer.android.com/guide/topics/sensors/sensors_overview [accessed 2020-07-07]
40. iOS security guide. Apple Platform Security. URL: https://support.apple.com/en-gb/guide/security/welcome/web [accessed 2020-07-07]
41. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. NPJ Digit Med 2019;2:38 [FREE Full text] [doi: 10.1038/s41746-019-0111-3] [Medline: 31304384]
42. Barlow DH, Farchione TJ, Sauer-Zavala S, Murray-Latin H, Bullis JR, Ellard K, et al. Unified Protocol for Transdiagnostic Treatment of Emotional Disorders: Therapist Guide. Oxford, United Kingdom: Oxford University Press; 2017:1-208.
43. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
44. Hamilton M. Hamilton rating scale for depression. In: Sartorius N, Ban TA, editors. Assessment of Depression. Berlin: Springer; 1986:143-152.
45. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-34. [Medline: 9881538]
46. Beck AT, Steer RA, Brown G. Beck Depression Inventory–II. Psychol Assess 1996:A. [doi: 10.1037/t00742-000]
47. Lambert MJ, Gregersen AT, Burlingame GM. The Outcome Questionnaire-45. In: Maruish ME, editor. The use of psychological testing for treatment planning and outcomes assessment: Instruments for adults. Oxfordshire, United Kingdom: Routledge; 2004:1-528.
48. Evans C, Mellor-Clark J, Margison F, Barkham M, Audin K, Connell J, et al. CORE: Clinical outcomes in routine evaluation. J Mental Health 2009 Jul 06;9(3):247-255. [doi: 10.1080/jmh.9.3.247.255]
49. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
50. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the mindfulness meditation mobile app "Calm" to reduce stress among college students: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jun 25;7(6):e14273 [FREE Full text] [doi: 10.2196/14273] [Medline: 31237569]
51. Huberty J, Eckert R, Larkey L, Kurka J, De Jesús SA, Yoo W, et al. Smartphone-based meditation for myeloproliferative neoplasm patients: feasibility study to inform future trials. JMIR Form Res 2019 Apr 29;3(2):e12662 [FREE Full text] [doi: 10.2196/12662] [Medline: 31033443]

52. Huberty J, Eckert R, Larkey L, Joeman L, Mesa R. Experiences of using a consumer-based mobile meditation app to improve fatigue in myeloproliferative patients: qualitative study. JMIR Cancer 2019 Jul 22;5(2):e14292 [FREE Full text] [doi: 10.2196/14292] [Medline: 31333197]
53. Huberty J, Vranceanu A, Carney C, Breus M, Gordon M, Puzia ME. Characteristics and usage patterns among 12,151 paid subscribers of the calm meditation app: cross-sectional survey. JMIR Mhealth Uhealth 2019 Nov 03;7(11):e15648 [FREE Full text] [doi: 10.2196/15648] [Medline: 31682582]
54. Huberty J, Puzia M, Eckert R, Larkey L. Cancer patients' and survivors' perceptions of the calm app: cross-sectional descriptive study. JMIR Cancer 2020 Jan 25;6(1):e16926 [FREE Full text] [doi: 10.2196/16926] [Medline: 32039812]
55. Puzia ME, Huberty J, Eckert R, Larkey L, Mesa R. Associations between global mental health and response to an app-based meditation intervention in myeloproliferative neoplasm patients. Integr Cancer Ther 2020 Jun 20;19:1534735420927780 [FREE Full text] [doi: 10.1177/1534735420927780] [Medline: 32564631]
56. Huberty J, Puzia ME, Larkey L, Irwin MR, Vranceanu A. Use of the consumer-based meditation app calm for sleep disturbances: cross-sectional survey study. JMIR Form Res 2020 Nov 13;4(11):e19508 [FREE Full text] [doi: 10.2196/19508] [Medline: 33185552]
57. Clarke J, Draper S. Intermittent mindfulness practice can be beneficial, and daily practice can be harmful. An in depth, mixed methods study of the "Calm" app's (mostly positive) effects. Internet Interv 2020 Mar;19:100293 [FREE Full text] [doi: 10.1016/j.invent.2019.100293] [Medline: 31890639]
58. Callender KA, Trustey CE, Alton L, Hao Y. Single case evaluation of a mindfulness-based mobile application with a substance abuse counselor. Counsel Outcome Res Eval 2019 Nov 22;12(1):16-29. [doi: 10.1080/21501378.2019.1686353]
59. Puzia M, Laird B, Green J, Huberty J. Parents' perceptions of their children's engagement in a consumer-based meditation mobile app: cross-sectional survey study. JMIR Pediatr Parent 2020 Nov 13;3(2):e24536 [FREE Full text] [doi: 10.2196/24536] [Medline: 33185564]
60. Howells A, Ivtzan I, Eiroa-Orosa FJ. Putting the 'app' in Happiness: A randomised controlled trial of a smartphone-based mindfulness intervention to enhance wellbeing. J Happiness Stud 2014 Oct 29;17(1):163-185. [doi: 10.1007/s10902-014-9589-1]
61. Lim D, Condon P, DeSteno D. Mindfulness and compassion: an examination of mechanism and scalability. PLoS One 2015 Feb 17;10(2):e0118221 [FREE Full text] [doi: 10.1371/journal.pone.0118221] [Medline: 25689827]
62. Taylor M, Hageman JR, Brown M. A mindfulness intervention for residents: relevance for pediatricians. Pediatr Ann 2016 Oct 01;45(10):373-376. [doi: 10.3928/19382359-20160912-01] [Medline: 27735974]
63. Laurie J, Blandford A. Making time for mindfulness. Int J Med Inform 2016 Dec;96:38-50. [doi: 10.1016/j.ijmedinf.2016.02.010] [Medline: 26965526]
64. Wen L, Sweeney TE, Welton L, Trockel M, Katznelson L. Encouraging mindfulness in medical house staff via smartphone app: a pilot study. Acad Psychiatry 2017 Oct 9;41(5):646-650. [doi: 10.1007/s40596-017-0768-3] [Medline: 28795335]
65. Wylde CM, Mahrer NE, Meyer RM, Gold JI. Mindfulness for novice pediatric nurses: smartphone application versus traditional intervention. J Pediatr Nurs 2017 Sep;36:205-212. [doi: 10.1016/j.pedn.2017.06.008] [Medline: 28888505]
66. Mistler LA, Ben-Zeev D, Carpenter-Song E, Brunette MF, Friedman MJ. Mobile mindfulness intervention on an acute psychiatric unit: feasibility and acceptability study. JMIR Ment Health 2017 Aug 21;4(3):e34 [FREE Full text] [doi: 10.2196/mental.7717] [Medline: 28827214]
67. Bennike IH, Wieghorst A, Kirk U. Online-based mindfulness training reduces behavioral markers of mind wandering. J Cogn Enhanc 2017 Apr 25;1(2):172-181. [doi: 10.1007/s41465-017-0020-9]
68. Kubo A, Altschuler A, Kurtovich E, Hendlish S, Laurent CA, Kolevska T, et al. A pilot mobile-based mindfulness intervention for cancer patients and their informal caregivers. Mindfulness (N Y) 2018 Dec;9(6):1885-1894 [FREE Full text] [doi: 10.1007/s12671-018-0931-2] [Medline: 30740187]
69. Noone C, Hogan MJ. A randomised active-controlled trial to examine the effects of an online mindfulness intervention on executive control, critical thinking and key thinking dispositions in a university student sample. BMC Psychol 2018 Apr 05;6(1):13 [FREE Full text] [doi: 10.1186/s40359-018-0226-3] [Medline: 29622047]
70. Yang E, Schamber E, Meyer RM, Gold JI. Happier healers: randomized controlled trial of mobile mindfulness for stress management. J Altern Complement Med 2018 May;24(5):505-513. [doi: 10.1089/acm.2015.0301] [Medline: 29420050]
71. Economides M, Martman J, Bell MJ, Sanderson B. Improvements in stress, affect, and irritability following brief use of a mindfulness-based smartphone app: a randomized controlled trial. Mindfulness (N Y) 2018;9(5):1584-1593 [FREE Full text] [doi: 10.1007/s12671-018-0905-4] [Medline: 30294390]
72. DeSteno D, Lim D, Duong F, Condon P. Meditation inhibits aggressive responses to provocations. Mindfulness 2017 Nov 2;9(4):1117-1122. [doi: 10.1007/s12671-017-0847-2]
73. Rosen KD, Paniagua SM, Kazanis W, Jones S, Potter JS. Quality of life among women diagnosed with breast Cancer: A randomized waitlist controlled trial of commercially available mobile app-delivered mindfulness training. Psychooncology 2018 Aug 01;27(8):2023-2030. [doi: 10.1002/pon.4764] [Medline: 29766596]
74. Champion L, Economides M, Chandler C. The efficacy of a brief app-based mindfulness intervention on psychosocial outcomes in healthy adults: A pilot randomised controlled trial. PLoS One 2018;13(12):e0209482 [FREE Full text] [doi: 10.1371/journal.pone.0209482] [Medline: 30596696]

75. Kubo A, Kurtovich E, McGinnis M, Aghaee S, Altschuler A, Quesenberry C, et al. A randomized controlled trial of mhealth mindfulness intervention for cancer patients and informal cancer caregivers: a feasibility study within an integrated health care delivery system. Integr Cancer Ther 2019 May 16;18:1534735419850634 [FREE Full text] [doi: 10.1177/1534735419850634] [Medline: 31092044]
76. Flett JA, Conner TS, Riordan BC, Patterson T, Hayne H. App-based mindfulness meditation for psychological distress and adjustment to college in incoming university students: a pragmatic, randomised, waitlist-controlled trial. Psychol Health 2020 Sep;35(9):1049-1074. [doi: 10.1080/08870446.2019.1711089] [Medline: 32046499]
77. Collins E, Cox A, Wilcock C, Sethu-Jones G. Digital games and mindfulness apps: comparison of effects on post work recovery. JMIR Ment Health 2019;6(7):e12853. [doi: 10.2196/12853] [Medline: 31322125]
78. Bostock S, Crosswell AD, Prather AA, Steptoe A. Mindfulness on-the-go: Effects of a mindfulness meditation app on work stress and well-being. J Occup Health Psychol 2018 May 03:127-138. [doi: 10.1037/ocp0000118] [Medline: 29723001]
79. Kirk U, Wieghorst A, Nielsen CM, Staiano W. On-the-spot binaural beats and mindfulness reduces behavioral markers of mind wandering. J Cogn Enhanc 2018 Nov 26;3(2):186-192. [doi: 10.1007/s41465-018-0114-z]
80. Quinones C, Griffiths MD. Reducing compulsive Internet use and anxiety symptoms via two brief interventions: A comparison between mindfulness and gradual muscle relaxation. J Behav Addict 2019 Sep 01;8(3):530-536 [FREE Full text] [doi: 10.1556/2006.8.2019.45] [Medline: 31505967]
81. Björkstrand J, Schiller D, Li J, Davidson P, Rosén J, Mårtensson J, et al. The effect of mindfulness training on extinction retention. Sci Rep 2019 Dec 27;9(1):19896 [FREE Full text] [doi: 10.1038/s41598-019-56167-7] [Medline: 31882606]
82. Flett JA, Fletcher BD, Riordan BC, Patterson T, Hayne H, Conner TS. The peril of self-reported adherence in digital interventions: A brief example. Internet Interv 2019 Dec;18:100267 [FREE Full text] [doi: 10.1016/j.invent.2019.100267] [Medline: 31890620]
83. Avalos LA, Aghaee S, Kurtovich E, Quesenberry C, Nkemere L, McGinnis MK, et al. A mobile health mindfulness intervention for women with moderate to moderately severe postpartum depressive symptoms: feasibility study. JMIR Ment Health 2020 Nov 12;7(11):e17405 [FREE Full text] [doi: 10.2196/17405] [Medline: 33180028]
84. Nübold A, Van Quaquebeke N, Hülsheger UR. Be(com)ing Real: A multi-source and an intervention study on mindfulness and authentic leadership. J Bus Psychol 2019 Jun 21;35(4):469-488. [doi: 10.1007/s10869-019-09633-y]
85. Rung AL, Oral E, Berghammer L, Peters ES. Feasibility and acceptability of a mobile mindfulness meditation intervention among women: intervention study. JMIR Mhealth Uhealth 2020 Jun 02;8(6):e15943 [FREE Full text] [doi: 10.2196/15943] [Medline: 32442147]
86. Ball E, Newton S, Rohricht F, Steed L, Birch J, Dodds J, et al. mHealth: providing a mindfulness app for women with chronic pelvic pain in gynaecology outpatient clinics: qualitative data analysis of user experience and lessons learnt. BMJ Open 2020 Mar 12;10(3):e030711 [FREE Full text] [doi: 10.1136/bmjopen-2019-030711] [Medline: 32165550]
87. Hussain A, Mkpojiogu E, Hussein I, Muhi O, Yosri M. The effectiveness, efficiency and reliability-in-use of daylio mobile app. Int J Adv Sci Technol 2020;29(8):180-187 [FREE Full text]
88. Cristol S. Patient's perspective on using mobile technology as an aid to psychotherapy. JMIR Ment Health 2018 Oct 24;5(4):e10015 [FREE Full text] [doi: 10.2196/10015] [Medline: 30355552]
89. Chaudhry BM. Daylio: mood-quantification for a less stressful you. Mhealth 2016;2:34 [FREE Full text] [doi: 10.21037/mhealth.2016.08.04] [Medline: 28293607]
90. Scherr S, Goering M. Is a self-monitoring app for depression a good place for additional mental health information? Ecological momentary assessment of mental help information seeking among smartphone users. Health Commun 2020 Jul 26;35(8):1004-1012. [doi: 10.1080/10410236.2019.1606135] [Medline: 31025888]
91. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
92. Kretzschmar K, Tyroll H, Pavarini G, Manzini A, Singh I, NeurOx Young People's Advisory Group. Can your phone be your therapist? Young people's ethical perspectives on the use of fully automated conversational agents (Chatbots) in mental health support. Biomed Inform Insights 2019 Mar 05;11:1178222619829083 [FREE Full text] [doi: 10.1177/1178222619829083] [Medline: 30858710]
93. Inkster B, O'Brien R, Selby E, Joshi S, Subramanian V, Kadaba M, et al. Digital health management during and beyond the COVID-19 pandemic: Opportunities, barriers, and recommendations. JMIR Ment Health 2020 Jun 01:- [FREE Full text] [doi: 10.2196/19246] [Medline: 32484783]
94. Marcelle ET, Nolting L, Hinshaw SP, Aguilera A. Effectiveness of a Multimodal Digital Psychotherapy Platform for Adult Depression: A Naturalistic Feasibility Study. JMIR Mhealth Uhealth 2019 Jan 23;7(1):e10948 [FREE Full text] [doi: 10.2196/10948] [Medline: 30674448]
95. Moberg C, Niles A, Beermann D. Guided Self-Help Works: Randomized Waitlist Controlled Trial of Pacifica, a Mobile App Integrating Cognitive Behavioral Therapy and Mindfulness for Stress, Anxiety, and Depression. J Med Internet Res 2019 Jun 08;21(6):e12556 [FREE Full text] [doi: 10.2196/12556] [Medline: 31199319]

96. Broglia E, Millings A, Barkham M. Counseling With Guided Use of a Mobile Well-Being App for Students Experiencing Anxiety or Depression: Clinical Outcomes of a Feasibility Trial Embedded in a Student Counseling Service. JMIR Mhealth Uhealth 2019 Aug 15;7(8):e14318 [FREE Full text] [doi: 10.2196/14318] [Medline: 31418424]
97. Baumel A. Online emotional support delivered by trained volunteers: users' satisfaction and their perception of the service compared to psychotherapy. J Ment Health 2015 Oct 20;24(5):313-320. [doi: 10.3109/09638237.2015.1079308] [Medline: 26485198]
98. Baumel A, Schueller SM. Adjusting an Available Online Peer Support Platform in a Program to Supplement the Treatment of Perinatal Depression and Anxiety. JMIR Ment Health 2016 Mar 21;3(1):e11 [FREE Full text] [doi: 10.2196/mental.5335] [Medline: 27001373]
99. Baumel A, Tinkelman A, Mathur N, Kane JM. Digital Peer-Support Platform (7Cups) as an Adjunct Treatment for Women With Postpartum Depression: Feasibility, Acceptability, and Preliminary Efficacy Study. JMIR Mhealth Uhealth 2018 Feb 13;6(2):e38 [FREE Full text] [doi: 10.2196/mhealth.9482] [Medline: 29439944]
100. Baumel A, Correll CU, Birnbaum M. Adaptation of a peer based online emotional support program as an adjunct to treatment for people with schizophrenia-spectrum disorders. Internet Interv 2016 May;4:35-42 [FREE Full text] [doi: 10.1016/j.invent.2016.03.003] [Medline: 30135788]
101. Malik A, Kalidhar A, Hanon C. InnerHour: Solving the Treatment Gap in Mental Health. Using technology to disrupt the mental health landscape. L'information psychiatrique 2020;96(5):348-354. [doi: 10.1684/ipe.2020.2113]
102. Carpenter J, Crutchley P, Zilca RD, Schwartz HA, Smith LK, Cobb AM, et al. Seeing the “Big” Picture: Big Data Methods for Exploring Relationships Between Usage, Language, and Outcome in Internet Intervention Data. J Med Internet Res 2016 Aug 31;18(8):e241 [FREE Full text] [doi: 10.2196/jmir.5725]
103. Parks AC, Williams AL, Tugade MM, Hokes KE, Honomichl RD, Zilca RD. Testing a scalable web and smartphone based intervention to improve depression, anxiety, and resilience: A randomized controlled trial. Intnl. J. Wellbeing 2018 Dec 08;8(2):22-67. [doi: 10.5502/ijw.v8i2.745]
104. Williams AL, Parks AC, Cormier G, Stafford J, Whillans A. Improving resilience among employees high in depression, anxiety, and workplace distress. Int J Manag Res 2018;9(1-2):4-22 [FREE Full text]
105. Hunter JF, Olah MS, Williams AL, Parks AC, Pressman SD. Effect of Brief Biofeedback via a Smartphone App on Stress Recovery: Randomized Experimental Study. JMIR Serious Games 2019 Nov 26;7(4):e15974 [FREE Full text] [doi: 10.2196/15974] [Medline: 31769761]
106. Martínez-Guzmán A, Lara A. Affective modulation in positive psychology’s regime of happiness. Theory Psychol 2019 May 30;29(3):336-357. [doi: 10.1177/0959354319845138]
107. Parks AC, Williams AL, Kackloudis GM, Stafford JL, Boucher EM, Honomichl RD. The effects of a digital well-being intervention on patients with chronic conditions: observational study. J Med Internet Res 2020 Jan 10;22(1):e16211 [FREE Full text] [doi: 10.2196/16211] [Medline: 31922491]
108. Sarkar U, Gourley GI, Lyles CR, Tieu L, Clarity C, Newmark L, et al. Usability of commercially available mobile applications for diverse patients. J Gen Intern Med 2016 Dec;31(12):1417-1426. [doi: 10.1007/s11606-016-3771-6] [Medline: 27418347]
109. Rizvi SL, Dimeff LA, Skutch J, Carroll D, Linehan MM. A pilot study of the DBT coach: an interactive mobile phone application for individuals with borderline personality disorder and substance use disorder. Behav Ther 2011 Dec;42(4):589-600. [doi: 10.1016/j.beth.2011.01.003] [Medline: 22035988]
110. Rizvi SL, Hughes CD, Thomas MC. The DBT Coach mobile application as an adjunct to treatment for suicidal and self-injuring individuals with borderline personality disorder: A preliminary evaluation and challenges to client utilization. Psychol Serv 2016 Nov;13(4):380-388. [doi: 10.1037/ser0000100] [Medline: 27797571]
111. Bush NE, Ouellette G, Kinn J. Utility of the T2 mood tracker mobile application among army warrior transition unit service members. Military Medicine 2014 Dec;179(12):1453-1457. [doi: 10.7205/milmed-d-14-00271]
112. Dewar AR, Bull TP, Sproat JM, Reyes NP, Malvey DM, Szalma JL. Testing the reliability of a measure of motivation to engage with telehealth technology. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2016 Sep 15;60(1):1114-1118. [doi: 10.1177/1541931213601261]
113. McCreight SJ, Brinton C, Kinn J, Bush N, Hoyt T. Integration of mCare and T2 Mood Tracker: Illustrating mHealth usability testing. J Technol Behav Sci 2019 Mar 26;4(2):130-138. [doi: 10.1007/s41347-019-00087-w]
114. Dryman MT, McTeague LM, Olino TM, Heimberg RG. Evaluation of an open-access CBT-based Internet program for social anxiety: Patterns of use, retention, and outcomes. J Consult Clin Psychol 2017 Oct;85(10):988-999. [doi: 10.1037/ccp0000232] [Medline: 28650193]
115. Bakker D, Kazantzis N, Rickwood D, Rickard N. A randomized controlled trial of three smartphone apps for enhancing public mental health. Behav Res Ther 2018 Oct;109:75-83. [doi: 10.1016/j.brat.2018.08.003] [Medline: 30125790]
116. Dahne J, Lejuez CW, Diaz VA, Player MS, Kustanowitz J, Felton JW, et al. Pilot randomized trial of a self-help behavioral activation mobile app for utilization in primary care. Behav Ther 2019 Jul;50(4):817-827. [doi: 10.1016/j.beth.2018.12.003] [Medline: 31208690]
117. Kinderman P, Hagan P, King S, Bowman J, Chahal J, Gan L, et al. The feasibility and effectiveness of Catch It, an innovative CBT smartphone app. BJPsych Open 2016 May;2(3):204-209 [FREE Full text] [doi: 10.1192/bjpo.bp.115.002436] [Medline: 27703777]

118. Jayachandran C, Shyamala K. Kohonen SOM deployment in android app-based cognitive behavioral therapy for personality disorders. In: Proeedings of the World Congress on Computing and Communication Technologies (WCCCT). 2017 Presented at: World Congress on Computing and Communication Technologies (WCCCT); Feb. 2-4, 2017; Tiruchirappalli, India. [doi: 10.1109/wccct.2016.69]
119. Flett J, Hayne H, Riordan B, Thompson LJ, Conner T. Mobile mindfulness meditation: a randomised controlled trial of the effect of two popular apps on mental health. Mindfulness 2018 Oct 31;10(5):863-876. [doi: 10.1007/s12671-018-1050-9]
120. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010;51(3-4):214-221 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
121. Jacobson NC, Chung YJ. Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones. Sensors (Basel) 2020 Jun 24;20(12):3572 [FREE Full text] [doi: 10.3390/s20123572] [Medline: 32599801]
122. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. Lancet 2009 Aug 22;374(9690):609-619. [doi: 10.1016/S0140-6736(09)60879-5] [Medline: 19640579]
123. Jacobson NC, Newman MG. Anxiety and depression as bidirectional risk factors for one another: A meta-analysis of longitudinal studies. Psychol Bull 2017 Nov;143(11):1155-1200. [doi: 10.1037/bul0000111] [Medline: 28805400]
124. Chancellor S, De Choudhury M. Methods in predictive techniques for mental health status on social media: a critical review. NPJ Digit Med 2020 Mar 24;3(1):43 [FREE Full text] [doi: 10.1038/s41746-020-0233-7] [Medline: 32219184]
125. Dagum P. Digital biomarkers of cognitive function. NPJ Digit Med 2018;1:10 [FREE Full text] [doi: 10.1038/s41746-018-0018-4] [Medline: 31304295]
126. Jacobson NC, Weingarden H, Wilhelm S. Using digital phenotyping to accurately detect depression severity. J Nerv Ment Dis 2019;207(10):893-896. [doi: 10.1097/nmd.0000000000001042]
127. Jacobson NC, Weingarden H, Wilhelm S. Digital biomarkers of mood disorders and symptom change. NPJ Digit Med 2019;2:3 [FREE Full text] [doi: 10.1038/s41746-019-0078-0] [Medline: 31304353]
128. Saeb S, Lattie EG, Kording KP, Mohr DC. Mobile phone detection of semantic location and its relationship to depression and anxiety. JMIR Mhealth Uhealth 2017 Aug 10;5(8):e112 [FREE Full text] [doi: 10.2196/mhealth.7297] [Medline: 28798010]
129. Saeb S, Lattie E, Schueller S, Kording K, Mohr D. The relationship between mobile phone location sensor data and depressive symptom severity. PeerJ 2016;4:e2537 [FREE Full text] [doi: 10.7717/peerj.2537] [Medline: 28344895]
130. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S. Mobile sensing and support for people with depression: a pilot trial in the wild. JMIR Mhealth Uhealth 2016 Sep 21;4(3):e111 [FREE Full text] [doi: 10.2196/mhealth.5960] [Medline: 27655245]
131. Lu J, Shang C, Yue C, Morillo R, Ware S, Kamath J, et al. Joint modeling of heterogeneous sensing data for depression assessment via multi-task learning. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Mar 26;2(1):1-21. [doi: 10.1145/3191753]
132. Osmani V. Smartphones in mental health: detecting depressive and manic episodes. IEEE Pervasive Comput 2015 Jul;14(3):10-13. [doi: 10.1109/mprv.2015.54]
133. Mehrotra A, Musolesi M, Hendley R, Pejovic V. Designing content-driven intelligent notification mechanisms for mobile applications. In: UbiComp '15: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp '15: The 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 7 - 11, 2015; Osaka Japan p. 813-824. [doi: 10.1145/2750858.2807544]
134. Mishra V, Künzler F, Kramer J, Fleisch E, Kowatsch T, Kotz D. Detecting receptivity for mHealth interventions in the natural environment. Proc ACM Interact Mob Wearable Ubiquitous Technol 2021 Jun 23;5(2):1-24. [doi: 10.1145/3463492]
135. Künzler F, Mishra V, Kramer J, Kotz D, Fleisch E, Kowatsch T. Exploring the state-of-receptivity for mHealth interventions. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Dec 11;3(4):1-27. [doi: 10.1145/3369805]
136. Choi W, Park S, Kim D, Lim Y, Lee U. Multi-stage receptivity model for mobile just-in-time health intervention. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Jun 21;3(2):1-26. [doi: 10.1145/3328910]
137. Sarker H, Sharmin M, Ali A, Rahman M, Bari R, Hossain S, et al. Assessing the availability of users to engage in just-in-time intervention in the natural environment. Proc ACM Int Conf Ubiquitous Comput 2014;2014:909-920 [FREE Full text] [doi: 10.1145/2632048.2636082] [Medline: 25798455]
138. Pielot M, Dingler T, Pedro J, Oliver N. When attention is not scarce - detecting boredom from mobile phone usage. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp '15: The 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 7 - 11, 2015; Osaka Japan p. 825-836. [doi: 10.1145/2750858.2804252]
139. Sucala M, Cuijpers P, Muench F, Cardo R, Soflau R, Dobrean A, et al. Anxiety: There is an app for that. A systematic review of anxiety apps. Depress Anxiety 2017 Jun 15;34(6):518-525. [doi: 10.1002/da.22654] [Medline: 28504859]
140. Van Ameringen M, Turna J, Khalesi Z, Pullia K, Patterson B. There is an app for that! The current state of mobile applications (apps) for DSM-5 obsessive-compulsive disorder, posttraumatic stress disorder, anxiety and mood disorders. Depress Anxiety 2017 Jun 01;34(6):526-539. [doi: 10.1002/da.22657] [Medline: 28569409]

141. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
142. Huckvale K, Torous J, Larsen ME. Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation. JAMA Netw Open 2019 Apr 05;2(4):e192542 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.2542] [Medline: 31002321]
143. Huguet A, Rao S, McGrath PJ, Wozney L, Wheaton M, Conrod J, et al. A systematic review of cognitive behavioral therapy and behavioral activation apps for depression. PLoS One 2016;11(5):e0154248 [FREE Full text] [doi: 10.1371/journal.pone.0154248] [Medline: 27135410]
144. Martinengo L, Van Galen L, Lum E, Kowalski M, Subramaniam M, Car J. Suicide prevention and depression apps' suicide risk assessment and management: a systematic assessment of adherence to clinical guidelines. BMC Med 2019 Dec 19;17(1):231 [FREE Full text] [doi: 10.1186/s12916-019-1461-z] [Medline: 31852455]
145. Wasil AR, Venturo-Conerly KE, Shingleton RM, Weisz JR. A review of popular smartphone apps for depression and anxiety: assessing the inclusion of evidence-based content. Behav Res Ther 2019 Dec;123:103498. [doi: 10.1016/j.brat.2019.103498] [Medline: 31707224]

Abbreviations

JITAI: just-in-time adaptive intervention

Edited by G Eysenbach; submitted 06.04.21; peer-reviewed by Y Zhang, A Teles; comments to author 23.04.21; revised version received 06.07.21; accepted 26.07.21; published 28.09.21.

Please cite as:
Teepe GW, Da Fonseca A, Kleim B, Jacobson NC, Salamanca Sanabria A, Tudor Car L, Fleisch E, Kowatsch T
Just-in-Time Adaptive Mechanisms of Popular Mobile Apps for Individuals With Depression: Systematic App Search and Literature Review
J Med Internet Res 2021;23(9):e29412
URL: https://www.jmir.org/2021/9/e29412
doi:10.2196/29412
PMID:34309569

Original Paper

Finding Relevant Psychoeducation Content for Adolescents Experiencing Symptoms of Depression: Content Analysis of User-Generated Online Texts

Kim K Dysthe[1], MD; Ole R Haavet[1], MD, PhD; Jan I Røssberg[2], MD, PhD; Petter B Brandtzaeg[3,4], PhD; Asbjørn Følstad[4], PhD; Atle Klovning[1], MD, PhD

[1]Department of General Practice/Family Medicine, University of Oslo, Oslo, Norway

[2]Division of Psychiatric Treatment Research, Department of Psychiatry, University of Oslo, Oslo, Norway

[3]Department of Media and Communication, University of Oslo, Oslo, Norway

[4]SINTEF Digital, Software and Service Innovation, Oslo, Norway

Corresponding Author:
Kim K Dysthe, MD
Department of General Practice/Family Medicine
University of Oslo
Kirkeveien 166
Oslo, 0450
Norway
Phone: 47 22 85 05 50
Email: k.k.dysthe@medisin.uio.no

Abstract

Background: Symptoms of depression are frequent in youth and may develop into more severe mood disorders, suggesting interventions should take place during adolescence. However, young people tend not to share mental problems with friends, family, caregivers, or professionals. Many receive misleading information when searching the internet. Among several attempts to create mental health services for adolescents, technological information platforms based on psychoeducation show promising results. Such development rests on established theories and therapeutic models. To fulfill the therapeutic potential of psychoeducation in health technologies, we lack data-driven research on young peoples' demand for information about depression.

Objective: Our objective is to gain knowledge about what information is relevant to adolescents with symptoms of depression. From this knowledge, we can develop a population-specific psychoeducation for use in different technology platforms.

Methods: We conducted a qualitative, constructivist-oriented content analysis of questions submitted by adolescents aged 16-20 years to an online public information service. A sample of 100 posts containing questions on depression were randomly selected from a total of 870. For analysis, we developed an a priori codebook from the main information topics of existing psychoeducational programs on youth depression. The distribution of topic prevalence in the total volume of posts containing questions on depression was calculated.

Results: With a 95% confidence level and a ±9.2% margin of error, the distribution analysis revealed the following categories to be the most prevalent among adolescents seeking advice about depression: self-management (33%, 61/180), etiology (20%, 36/180), and therapy (20%, 36/180). Self-management concerned subcategories on coping in general and how to open to friends, family, and caregivers. The therapy topic concerned therapy options, prognosis, where to seek help, and how to open up to a professional. We also found young people dichotomizing therapy and self-management as opposite entities. The etiology topic concerned stressors and risk factors. The diagnosis category was less frequently referred to (9%, 17/180).

Conclusions: Self-management, etiology, and therapy are the most prevalent categories among adolescents seeking advice about depression. Young people also dichotomize therapy and self-management as opposite entities. Future research should focus on measures to promote self-management, measures to stimulate expectations of self-efficacy, information about etiology, and information about diagnosis to improve self-monitoring skills, enhancing relapse prevention.

(J Med Internet Res 2021;23(9):e28765) doi:10.2196/28765

KEYWORDS

adolescent; depression; internet; education; preventive psychiatry; early medical intervention; self-report; psychoeducation; information content; online; digital health; e-health

Introduction

Background

The highest loss of disability-adjusted life-years (DALYs) due to depression is seen in mid-adolescence [1]. Previous studies report a 5.6%-15% lifetime cumulative prevalence of major depressive disorder (MDD) in youth [2-4]. Recent epidemiologic studies measuring cumulative prevalence in shorter timespans reveal rates ranging from 2% to 17.7% [5-9]. Subthreshold depression and symptoms of depression are found in 9.3%-20% of adolescents [5,6,10], with studies revealing higher rates in girls [6]. The subsequent increased risk of early disability, adverse mental disorders, and suicide later in life [11-16] calls for intervention at adolescent age [17-21]. Although depression is frequent among adolescents, they also appear hesitant to talk openly about mental problems to family, friends, caregivers, and health personnel [22-24]. Instead, young people increasingly use social media, online influencers, and internet search engines to gather health information. Such information may frequently be misleading, not supported by clinical research, and believed to suppress help-seeking behavior [25]. When an intervention opportunity eventually arises, adolescents are inclined to drop out from therapy [26-28]. Consequently, we must consider new ways to design relevant and engaging clinical solutions to stimulate help-seeking behavior and promote therapy adherence [29]. Several attempts to develop such strategies have shown promising results [30].

In general, previous studies suggest that accurate initial information about depression through the realm of psychoeducation could promote help-seeking behavior and treatment compliance, self-management, and social functioning and may alter the course of depression [31,32].

We believe the established clinical principles of psychoeducation could be used to develop diverse technological solutions for youth mental health services [33,34]. However, psychoeducation is characterized by an extensive fauna of therapeutic approaches and theories, masking the efficacy of its various informational components [35,36]. To reveal its full potential as a prerequisite for clinical technologies, such as informational websites, online therapy, or chatbots, we need to develop population-specific psychoeducational content targeting adolescents with depression.

The results of this study could also be applied as a basis for a more extensive toolkit for clinical use in both primary and secondary health care. Psychoeducation is regarded as an integral part of different therapeutic strategies, such as cognitive behavioral therapy (CBT). In the future, we need this kind of research as a cornerstone to develop novel clinical programs based on the principals of joint clinical and technological care.

Prior Work

According to previous studies, psychoeducation seems to be effective in treating depression, although as an integral component of more extensive therapy programs or fields [37-39]. In general, internet-based interventions for depression and anxiety disorders prove efficacious in adults [31,40]. Similar solutions also show promising results in adolescents [39,41,42]. Although 2 systematic reviews suggest that psychoeducation may prevent early onset and support management of adolescent depression [43,44], more research in this area is needed.

In the literature, the definition of psychoeducation varies. For instance, defining it as *different types of information about the disease, essential to the patient's biological, psychological, and social function* [43,44] does not cover all its clinical goals. The boundaries between passive health information and generic therapeutic problem solving remain fuzzy. Instead, a more fruitful path may be to pursue the clinical *aims* of psychoeducational intervention: (1) to provide recipients with health-specific insight and (2) to enhance the level of self-efficacy [40].

First, to provide insight, psychoeducation could be provided passively on websites, in written materials, as technological solutions, as health information prior to further therapy, or in initial classroom lectures [34,43]. To meet this aim, professional consensus claims that health-specific insight should contain crucial elements of information topics about etiology, prevalence, symptoms and self-monitoring, treatment, natural course, and therapy prognosis [45,46].

The second overarching aim of stimulating self-efficacy builds on social learning theory [47,48]. According to this, self-efficacy is regarded as a necessity for self-management and rests on the perceived expectation that actions required for coping and self-preservation are obtainable. Following social learning theory, measures stimulating self-efficacy expectations are just as important for therapy as a sturdy working alliance.

To enhance the experience of self-efficacy, the most frequently used psychoeducational programs extend into the habitat of CBT [33,34,49,50] and interpersonal therapy [51,52]. Examples of 2 commonly used programs are the Interpersonal Psychotherapy program [51] and the multicomponent therapy course Adolescent Coping with Depression [53]. The first focuses on social and interpersonal skills as a road map to symptom relief. The latter developed from the original Coping with Depression course by focusing on information about pleasant activities and relaxation exercise, adding an extra section on parental support and family structure. Later, similar programs introduced elements of metacognitive therapy and positive psychology, as well as lessons about cognitive restructuring, identifying and altering negative beliefs and thought patterns [49].

Building on these 2 clinical aims, existing psychoeducation programs are originally developed from established theories and therapeutic models and their efficacy later tested in clinical trials. However, data-driven research aiming to improve our knowledge about what *content* of psychoeducation is relevant

and engaging to adolescents suffering from depression remains severely limited [34,36,45,49,54].

We found only 2 previous studies aiming to fill this knowledge gap: (1) Bevan-Jones et al. [50] derived psychoeducational content from focus groups and semistructured interviews with young persons with or at risk of getting depression. The study also included parents, caregivers, and professional staff, suggesting attention toward mood versus symptoms of depression, possible reasons, self-management, and where to get help [50]. (2) Bru et al. [49] investigated the user experience of different components of a frequently used program derived from the Adolescent Coping with Depression course. In this study, the program participants found the initial general information about depression useful. Even if they found the latter part concerning cognitive restructuring and identification of thoughts difficult to use, it made them more able to regulate emotional reactions.

Goal of This Study

Our study ultimately aims to specify what information topics regarding depression are relevant to young people. Contrary to the previous studies, we use the spontaneous nature of natural text data in 870 posts related to symptoms of depression on Norway's largest public online information site for young people. To the best of our knowledge, our study will be the first to analyze unmoderated data revealing the information needs of adolescents with symptoms of depression, enabling us to answer the following research question: What information is most relevant to adolescents experiencing symptoms of depression?

This study will contribute knowledge about population-specific information needs, helping developers employ the full potential of psychoeducation for technology-based information platforms and clinical, technological solutions.

Methods

Study Design

In response to the research question and the large amount of text data, we conducted a content analysis [55] using Crabtree and Miller's template approach. The method includes developing an a priori codebook from the findings of previous research or analyses. In this case, to answer the research question, the codebook consists of topics from existing psychoeducation programs. The analysis was chiefly deductive in nature. However, to avoid missing important information, we also inductively developed new codes to a refined codebook during the process of analysis. The choice of study design allowed us to gather insight into what adolescents regard as relevant, described in their own words in posts concerning depression submitted by youth to a public information service.

Participants and Data

The service ung.no is the official youth information website provided by the Norwegian Directorate for Children, Youth and Family Affairs. The website is available for everyone, although Norwegian is the only language. The website's target population is young people in the age range of 13-20 years. The editorial staff provides secure, reliable, and relevant information through online articles and multimedia content such as animated videos, alongside a question-and-answer (Q&A) section where young people can choose from a predefined list of topics, such as economy, education, work, family, social relations, and health.

Through a data transfer agreement between the project and the Norwegian Directorate for Children, Youth, and Family Affairs, we gained access to 277,552 anonymized posts written from the years 2005 to 2018 in Microsoft Excel format. The young people using the service tag their post with age, gender, and a main topic from a predefined drop-down list before submitting it to a staff of nurses, psychologists, and doctors. One of the topics is mental health and emotions. In the entire dataset, there were 14,804 posts belonging to this topic. When replying, the staff tags the questions with 1 or more categories within each main topic. Each post can contain more than 1 category, organized in a given order. One of them is the category *depression*.

Even though the data are preorganized and exported from SQL to Excel format, this specific data source is not commonly used in research. The number of depression posts written by people belonging to the age category of late adolescence (16-20 years) was 870. The posts were registered between the years 2007 to 2018, although most of the posts in the sample were written between 2013 and 2018 (see Figure 1).

Figure 1. Data collection path.

To identify false negatives, that is, relevant posts not included in the initial categorization, we applied the Excel randomization function for a random selection of 150 posts from the entire dataset of 275,731 nondepression posts. We found only 3/150 (2%). A proportion (24%, 24/100) of the posts in the 100-post sample did not reveal any symptoms of depression, described other conditions such as bipolar disorders or personality issues, or portrayed someone else than the writer. We did not include these posts in the content analysis.

For further details about the data characteristics, refer to Table 1 and Figure 1. Some posts contained only a single specific question (see examples of posts in Textbox 1). Others conveyed a description of the situation and symptoms without any question at all. A few posts concerned third persons, such as friends or relatives, suffering from depression, seeking advice on how to relate. There was also some variation in the questions posted: Some of them unambiguously pointed to a specific topic. A significant proportion was polysemantic, disabling analysis based on computerized text mining, dependent on an interpretation of the post as a single unit of meaning.

Overall, the dataset provides unique insight into the information needs and beliefs about mental health during late adolescence and young adulthood. Analyzing this type of data has several advantages for developing technological or clinical programs specific for this group: First, we identify what information the target group demands when experiencing symptoms of depression, asking questions on public information websites, thereby demonstrating an information need. In contrast to data obtained through surveys or interviews, user-generated data from online services or social media are considered less influenced by the study context and environment, the relationship with an interviewing clinician, and the state of mood only at the time of assessment [56]. Second, the posts could describe narratives covering a longer time frame, providing a better opportunity for assessing cognitive processes. Moreover, the writers identify and describe their thoughts in their own words, at a time when they feel a need to express them, thus making the data valid for topic relevance valuation. For these reasons, online user-generated data may be recognized as an important tool for detecting and analyzing mechanisms of depression [57,58].

Table 1. Selected data features.

Data characteristics	Percentage and numbers
Posts included for content analysis	76
Posts revealing no symptoms of depression, other conditions, or describing someone else	Not analyzed: 24% (24/100)
Depression interpretation	Unambiguous: 59% (45/76) Interpretable: 41% (31/76)
Secondary health care level, when specified	18% (14/76)
Comorbidity described	34% (26/76)
Gender (entire sample, 100 posts)	Boys: 23% (22/97) Girls: 77% (75/97) Not specified: 3 posts
Age (entire sample, 100 posts)	16 years: 31% (31/100) 17 years: 24% (24/100) 18 years: 13% (13/100) 19 years: 22% (22/100) 20 years: 10% (10/100)
Mean word count per post (entire sample, 100 posts)	180 (17,997/100)

Textbox 1. Examples of posts, paraphrased before being translated to English.

- "I'm so depressed I don't want to live. My father beat my mom and me. I feel that people fail me. I'm sad. I'm sick and tired of not coping. I started high school, but I can't keep up. My favorite school subjects aren't interesting anymore. When I fail, I just give up. I cry a lot. I'm way too depressed to go to school. I'm in conflict with the others in my class, and cannot go to school when I feel bad, so I stay home. My friends are always together. But I feel like they'll shut me out. Now I'm no longer invited to things. I deleted everyone on Facebook, then they wondered why I don't talk to them anymore. I feel even more sad and want to die. Nobody cares about me. I wonder if I can talk to my girlfriend about that, but I feel like she will not understand me. Everything I try is useless. I don't quite know what to do anymore. I've talked to a counselor at school. He asks me to complete high school, but I can't. I just want an honest answer about what to do without involving psychologists."
- "Hi! I do not feel good enough, not pretty enough. I'm losing myself. I'm not who I once was. Before, I smiled and was never tired. Now I just want to die. I have suffered from depression since primary school. Lately it has actually been better. I have friends, family and a girlfriend who cares for me. I still just want to die. Recent weeks I have cried for nothing. I am thinking of taking my life and I'm a bad person to think this way, that I will leave all the good stuff behind. I have visited a psychologist but cannot open up. The psychologist only tells me things I already know. No therapists will take me seriously. I always smile and behave nicely, but I do not think it is wise, in fact, I've been thinking about hospitalization. I'm unstable. I can commit suicide any time. But if I am admitted, they will never listen anyway. I know that. They're just going to send me on to someone else who's talking about how to get better, just that, to get rid of my anxiety and depression, that simple. All this I know. I'm confused. All I know is that I want to die. That's why I write. What can I do with my life?"

Procedures

We developed an a priori codebook (Table 2) from the information topics about the different psychoeducation programs [45,46,51-53,59], predefining the following upon analysis: etiology of depressive disorders, frequency, natural course, therapy prognosis, therapy options, and diagnosis. From the interpersonal and cognitive theories, we added the topics of self-management and social skills [49-51].

From the 870 posts, we randomly extracted 100 for further analysis using the Excel randomization function. Two raters, a general practitioner (author KKD) and a psychiatrist (author JIR), both possessing cognitive therapy education and training, coded the posts using the NVivo 12 Pro application for qualitative analysis.

Table 2. A priori codebook (8 codes).

Codes: questions	Description/beliefs
1. Etiology: Why is it like this?	Beliefs about causality
2. Frequency: Is it common?	Beliefs about the prevalence of depression
3. Course: How will it develop?	Beliefs about the natural course of depression
4. Prognosis: Will it work?	Beliefs about therapy efficacy
5. Therapy: Will therapy make me better, and where can I seek help?	Beliefs about therapy options
6. Diagnosis: What is wrong with me?	Beliefs about diagnosis
7. Self-management: What can I do myself?	Beliefs about coping
8. Social support and participation: How do I open, and where do I seek social support?	Beliefs about the importance of social support

As visualized in Figure 2, we coded both questions and descriptions into the same corresponding NVivo nodes. To avoid missing information, each rater established novel codes containing topics not described in the codebook. We performed 5 iterations. For every iteration, the 2 raters met to discuss the analyzing process. We developed subcodes for every topic and agreed upon a refined codebook after every meeting. After 34 posts, no new codes were added, leaving an interim codebook of 11 codes. We then used the NVivo 12 Pro function "Aggregate Coding from Children" to calculate the reference count for each of the 11 interim codes and subcodes (Table 3). The interim codebook went through a final refinement, merging overlapping codes and subcodes into categories and subcategories due to similarity of content. The references were summarized into a total count for each category. Using Excel, we calculated each category's reference count proportion within the 100 posts and the percentage distribution of categories in the total number of text references.

Figure 2. Analysis procedure from data selection to results.

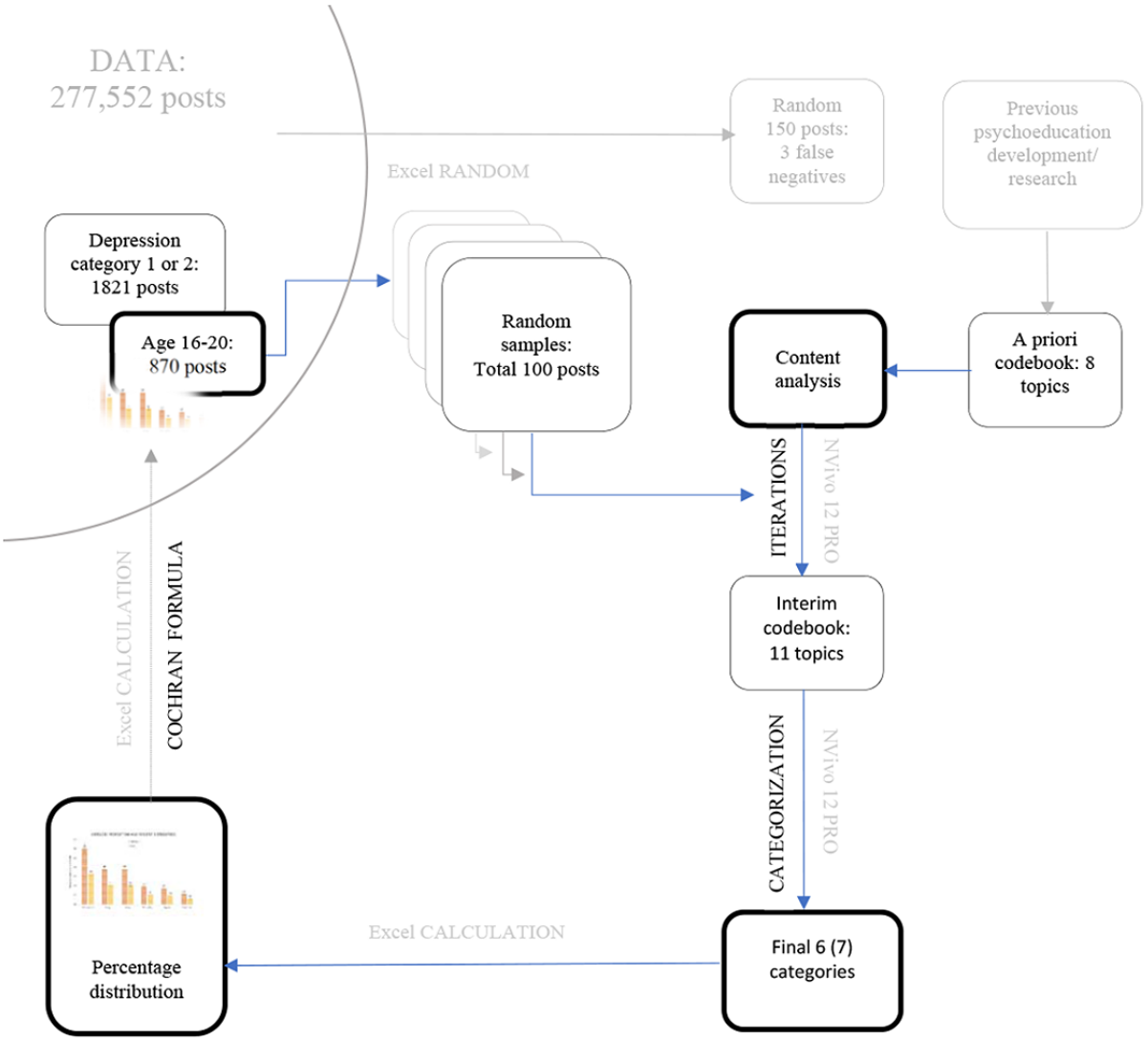


Table 3. Interim codebook and new codes with corresponding numbers of NVivo references (11 codes).

Code	Count	Classification
1. Etiology: Why is it like this? (Beliefs about causality)	35	Initial code
2. Frequency: Is it common? (Beliefs about the prevalence of depression)	1	Initial code
3. Course: How will it develop? (Beliefs about the natural course of depression)	12	Initial code
4. Prognosis: Will it work? (Beliefs about therapy efficacy)	10	Initial code
5. Therapy: What kind of therapy? (Beliefs about therapy options)	45	Initial code
6. Diagnosis: What is wrong with me? (Beliefs about diagnosis)	16	Initial code
7. Self-management: What can I do? (Beliefs about self-management)	33	Initial code
8. Social support and participation (beliefs about the importance of social support)	37	Initial code
9. Beliefs about consequences for their parents	7	New code
10. Beliefs about rights and privileges	7	New code
11. Is it normal? (Beliefs about the normality of symptoms)	9	New code

Statistical Analysis

Applying the Cochran formula [60], we calculated the adequate sample size (n_0) for an infinite population and then used the additional formula to calculate the sample size (n) of the finite population of 870 posts (N). Analyzing 100 posts enabled us to estimate the percentage distribution of each theme in the dataset of 870 posts with a 95% confidence level (Z=1.96) and a 9.2% ± margin of error (e), providing a 0.5 variance value (P=.5). For statistical comparison between the sample and the 870-post dataset, we included all the 100 posts as part of the sample, including those omitted from the content analysis (24%, 24/100).

Reliability

Regarding reliability, the first rater (KKD) coded the posts. The second rater (JIR) then went through all the coded posts. In cases of disagreement, the second rater suggested alternative coding or a new code. We reviewed all alternative or new coding in a code meeting for every iteration and generated new codes in 20% (43/212) of the cases. In all, 3 of the codes remained unresolved. These 3 codes were omitted from the final analysis.

Reflexivity

During the coding process, we discussed all codes that could reveal different interpretations. In numerous cases, both raters could relate connotations to real patients. Such topics were discussed to overcome biased interpretations. In most cases, the interpretation was straightforward, built on descriptions of common symptoms of depression. As far as possible, we strived to anchor the analysis on unambiguous functions of language, alongside a clear reference to previous research and theory.

Results

Preliminary Typology

Adding to the codebook (Table 3), we identified 3 new codes: questions and beliefs about normality, specific rights and privileges given to people with depression, and how depression will affect their parents.

As expected, the data revealed little overlap between the a priori codebook topics, leaving a single category for most of the codes. Based on social learning theory and text meaning, we interpret the actions of social support and participation closely related to the concept of self-efficacy. Hence, we merged the codes and references in the interim categories of self-management and social skills into a new self-management category. We did this for 2 reasons: First, social learning theory does not differentiate between the experience of self-efficacy obtained through actions related to social contact and other types of self-management, such as physical activities, jobs, and school achievements. Second, we also find this relation in the dataset revealing an expectation to manage; however, the adolescents seemed insecure about how. Consequently, ways of initiating social interaction and support are described as something they demand advice about in order to manage on their own, for instance, how to open up about their mental problems to friends and family. For the same reason, we added the new interim code *beliefs about consequences for their parents*, implying advice about involving parents, not only as a protective factor against suicidal and parasuicidal behavior, but also as a measure to support self-management.

We also merged the references to the codes *what kind of therapy* and *therapy prognosis* into the therapy category, due to the cognitive demand for information to alter expectations. In the posts, we found a temporal component in the questions about therapy, asking what to expect from a therapist or hospital admission, effects and side effects of medicines, and an overall demand for information about therapy prognosis in general.

Additionally, the data showed an essential issue in the polysemantic question “What should I do?” Although not frequently asked, the question reflects an important finding discussed later. This category was constructed from references coded to both *self-management* and *therapy* categories in the original a priori codebook and from one single reference from the social support code.

The 2 remaining new codes were much less referred to than the codes in the original codebook. We added 2 of them to the *miscellaneous* category.

Estimated Category Proportion (p^) and Percentage Distribution

In line with the category-constructing process, we applied NVivo 12 Pro to merge the corresponding codes and reference count with the different categories. We then calculated the estimated percentage distribution of the entire dataset of 870 posts, as presented in Figure 3: There was no overlap between the error bars of each column, meaning a ±9.2% margin of error will not affect the interpretation of the overall results. Analyzing a random sample of 100 posts was sufficient to answer the research question. The categories most frequently referred to were *self-management* (33%, 60/180), *therapy* (20%, 36/180), and *etiology* (20%, 36/180) (Table 4).

Figure 3. Proportion (p^) of posts containing references to each category. Percentage distribution of the total count of references, with error bars showing a ±9.2% margin of error.

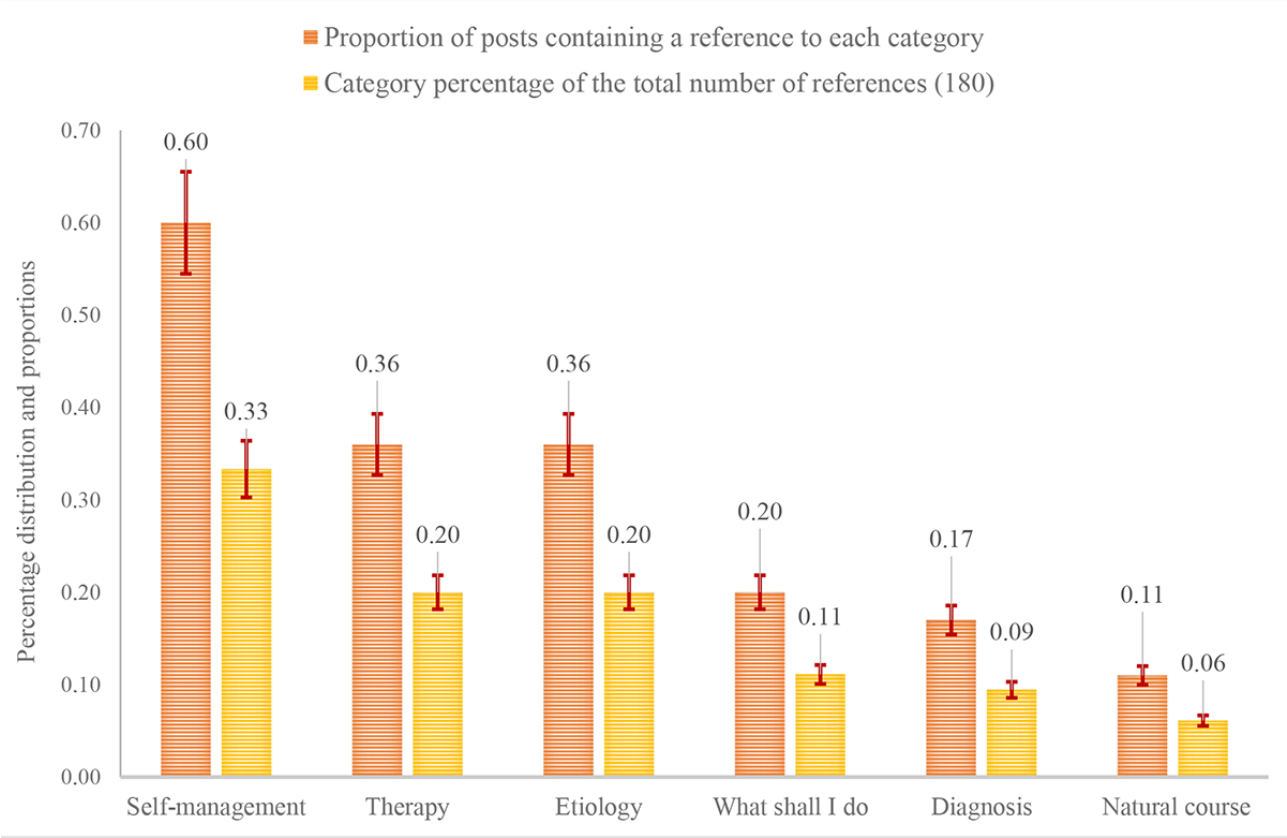


Table 4. Final category content and subcategories after coding and categorization.

Category	Count	Subcategories: beliefs and questions
1. Self-management	60	• Coping in general • How to open to friends and family • Beliefs about consequences for their parents
2. Therapy	36	• Therapy options • Prognosis • Where to seek help • How to open up to a professional
3. Etiology	36	• Risk factors • Stressors • Course of events leading to symptoms and behaviors
4. What should I do?	20	• Manage on my own, as opposed to seeking professional help
5. Diagnosis	17	• Symptoms conflicting with the self-concept • The diagnosis defining the mood
6. Natural course	11	• The course of depression as something out of control
7. Miscellaneous	11	• Rights and entitlements when suffering from depression • Normality regarding common psychological mechanisms

The *self-management* category (category 1) was most frequently referred to, revealing a demand for information about coping in general, how to open to friends and family for support, and how to relate to parents and caregivers. The *etiology* category (category 2) portrayed questions about risk factors and stressors leading to depression, as well as descriptions of events leading to symptoms and behaviors. In contrast, the *therapy* category (category 3) contained questions and beliefs about therapy options and therapy prognosis, where to seek professional aid, and how to open up to a professional regarding mental problems.

The category *What should I do?* (category 4) reflected the equivocalness of therapeutic interventions versus self-management. In a proportion of the posts, the narrative led to the question "What should I do?" This question occurred almost entirely in the context of the codes *what kind of treatment* or *self-management*. In several cases, this distinction was impossible to make, and these ambiguous text references were coded to both categories. One such reference that was coded to *social support and participation* also led to the same type of question. The posts regarding this topic described therapy, self-management, and social support either as different actions the writers believed could lead to improvement or as something they have already tried. Either way, the question did not belong to any of the a priori codes and revealed new beliefs about what actions to take when experiencing symptoms. Hence, we established a separate category for this question, counting each of the references coded to both *what kind of treatment* and *self-management* as one code in the *What should I do?* category (category 4).

The *diagnosis* category (category 5) consisted of questions and beliefs about what the symptoms could signify, describing experiences of being changed and demanding information about symptoms that resemble a disorder as something separate from the internal self-concept. At the same time, beliefs and questions about depression revealed a somewhat fatalistic view about the diagnosis as something defining mood without allowing variation, only susceptible to intervention. We found similar questions related to the *natural course* category. Beliefs about the course and prognosis of depression were regarded as something beyond control, relentlessly worsening the mental state, eventually resulting in suicidal thoughts.

Some of the new codes detected in the data were not frequently referred to; however, they were still worth mentioning due to a possible clinical significance. The code *What are my rights?* consisted of beliefs and questions about public financial support and absence from school and work when experiencing symptoms of depression. The code *Is it common?* consisted of questions and beliefs about disease prevalence, almost nonexistent in the data (count = 1). We generated a new code *Is it normal?* for questions and beliefs about the normality of psychological processes, complaints, and symptoms.

Discussion

Principal Results

We found that text references most frequently referred to the categories *self-management*, *etiology*, and *therapy*. The *therapy* category contains the subcategories of *therapy options* and *prognosis*, including where to seek initial help and how to relate to a health professional. However, measures to stimulate expectations of self-efficacy should also emphasize therapy as an active guiding process. Although the category of *etiology* was 1 of the most referred to, other categories, such as *diagnosis* and *natural course*, could possibly gain higher relative relevance later in the course of depression, when symptoms are less intrusive. This also applies to information about diagnosis, previously thought to improve self-monitoring skills aiming at preventing relapse.

Limitations

Our data were most probably influenced by selection bias, as many of the posts were written by those with self-perceived severe or persistent symptoms of depression. When we investigated the information needs of adolescents experiencing symptoms of depression, we looked for the distribution of information topics among those asking questions on the online service. We had to bear this in mind when generalizing to a population level: These topics were determined relevant to adolescents with symptoms of depression asking for guidance. The information needs may not apply to every adolescent with depression. Rather, when in need of information, this is what they consider to be relevant. Simultaneously, we did not know the cognitive and emotional states of the adolescents submitting the posts. Conditions not apparent in the posts could affect data quality, for instance, drug abuse, personality disorders, social phobia, age-related mood swings, and other conditions commonly associated with unipolar depression, such as anxiety, bipolar disorders, and physical conditions (chronic pain, myalgic encephalomyelitis). Furthermore, categorizing the posts describing depression is only achievable according to descriptions of symptoms, not severity or persistence. The DSM-5 and ICD-10 diagnosis classification systems rate severity according to factors such as duration, loss of function, the presence of core symptoms, and the number of other symptoms. In this dataset, the only reliable hint toward severity assessment is the number of different depression symptoms presented in each post.

Even if the majority of the posts were written in the years 2013 to 2018, some posts were submitted in 2009 (see Figure 1). Young people's language can change, possibly influencing the way they communicate issues about depression as a social construct. However, referring to the research question, we believe the experience of depression symptoms and the conjoining information needs are less influenced by changes in language over time. For example, the way young people regard common variations of mood as symptoms of a disease may be influenced by a proposed psychologization of youth language, nourished by education and social media. Nevertheless, referring to the previous literature, the content of psychoeducation is rooted in established theories of psychology and has not changed drastically, describing depression as a condition with biological presets. We believe the demand for specific information and the cognitive effect of such information, as described in existing psychoeducation development, to be consistent and less dependent on possible changes in language and attitudes.

Comparison With Prior Work

Following the first principal aim of psychoeducation, previous development does not regard any of the information topics to be of superior importance. Nevertheless, we found in our study that psychoeducation should pay extra attention to some of the components, especially during the stages of depression where the level of symptoms is high. Since the topics of self-management and therapy are related to interventions and support, this kind of information is more relevant to adolescents when in despair. Regarding our second aim, that of strengthening self-efficacy, the high prevalence of reference to the *therapy*

and *self-management* categories resonates not only with previous empirical research [45,49-51,53,61] but also with theory-driven development.

Interestingly, the principal results of this and other studies [45,49-51,53,61] could lead us to interpret social learning theory in a rather new perspective: Considering the spontaneous nature of self-reporting text data, our findings may indicate an inherent demand among these adolescents to manage on their own when experiencing symptoms of depression. Social learning theory describes self-efficacy as an expectation acquired through lifelong performance experiences [47]. Self-efficacy, the subsequent belief that different life challenges are manageable, is considered a chief ingredient for help-seeking and therapy motivation. This resonates with a central principle of CBT: the therapist as a guide to self-management.

When interpreting the content of the category *What should I do?*, we suggest that psychoeducation aim at enhancing the close relation between therapy and self-management, confirming the role of the therapist as actively underpinning performance skills [50]. In our study, we interpreted this question as a request for seeking advice about what to do in the current situation, referring to *either* exercises enhancing self-management *or* traditional therapy. This is not found in the 2 preceding studies. The category relates to the research question by conveying demands for guidance about actions to take when there are no apparent other options, revealing beliefs about these 2 types of interactions as opposite entities. Psychoeducation must also aim at overcoming the dichotomy of therapy versus self-management and provide information about the close integration of both.

Our study adds to previous research and psychoeducation development programs [45,50] on the questions and beliefs about *therapy*, as it contains the subcategories of *where to seek professional help* and *how to open up to a professional*. These are topics related to therapy, previously receiving little attention. Bevan-Jones et al. [50] found both young persons and professionals appreciating information about whom to contact for appropriate assistance. The *how to open up to a professional* subcategory reflects previous research about help seeking, suggesting that adolescents do need guidance not only to enhance social and familial openness and support but also to know where to seek initial professional help and what to expect from a therapist. In compliance with the clinical aim of preparing the patient for therapy [62], they require information about the role and communicating skills of a professional helper [22-24].

In our analysis of the questions categorized as concerning *etiology*, we found the adolescents to be particularly interested in risk factors and stressors (Table 3). We also found specific inquiries about etiology, summarized into the question *Why do I suffer from this?* Following previous psychoeducation development, alongside the findings of Bevan-Jones et al. [50], this question meets the aim of providing recipients with health-specific insight. It could also be that the adolescents refer to issues of causality as functions of language constituting an integral component of the narrative sequence of events leading to symptoms of depression. This could possibly explain the relatively high proportion of this topic.

In relation to the *diagnosis* category, CBT emphasizes the importance of self-monitoring skills [63]. Different psychoeducation programs for adolescents stress the ability to differentiate symptoms of depression from normal mood swings, thereby enabling teenagers with depression to seek earlier help and prevent relapse [45,50]. Although previous development regards the topic as highly important, we found a relatively modest attention directed toward symptoms of depression versus mood variations.

There seem to be mainly 2 ways to interpret this. First, as previously discussed, questions and beliefs about symptoms versus mood swings are relatively less relevant when experiencing symptoms. Second, this topic may have received less attention due to a lack of knowledge about the difference between age-related mood swings and symptoms of depression. Considering the self-reporting nature of the data, this could explain the low reference rates [34]. Regarding relapse prevention, we still believe self-monitoring abilities to be an important topic of psychoeducation and should be introduced to gain important insight into depression. It seems, however, more relevant to focus on relapse prevention when symptoms are less severe.

Considering the clinical implications and future research of our study, we believe this study can contribute to bringing more relevant and engaging psychoeducational content to the field of mental health information for adolescents and young adults. Initially, by reducing the risk of clinical information overload, we trust that the empirically derived information content could reduce program and therapy dropout and increase adherence when used in different technological solutions [24]. Then, by paying attention to content relevance, psychoeducation can focus on information about the disease rather than therapeutic techniques and theories, leading to the development of simpler and more efficient technological platforms and applications [64]. Future studies should focus on developing joint clinical and technological platforms targeting this population. Such research involves multiple steps, whereas deriving relevant information content is only the initial stage. Further research on machine learning technology, platform interface, and service design to enhance efficient integrations of data applications and clinical therapy is needed.

Beyond the scope of this study is also the analysis of specific cognitive patterns and relationships. We aim at deriving relevant information. At the same time, we lack empirical knowledge about the cognitive effects of such information. Using the same dataset, future analysis could let us investigate the information topics considering possible cognitive and metacognitive mechanisms and their ability to alter beliefs about depression. We also acknowledge a potential benefit in population- and disease-specific cognitive tools developed from such self-reporting text data. Text mining and machine learning analysis algorithms could be used on the entire dataset of all of the 14,804 posts regarding mental health to analyze combinations of symptoms and risk factors, alongside patterns of associated aspects, such as gender, age, social status, and comorbidity. From this kind of research, we may develop case-finding technological tools to alert clinicians about signs of depression in users of different e-mental health solutions.

This kind of research could possibly enhance our understanding of which patients are best suited for different kinds of therapy.

Finally, additional research on psychoeducation information content using other sources of data is required. Previous studies have used focus groups or individual interviews to ask young people about information preferences undergoing treatment or participating in psychoeducation programs [49,50]. One study also involved the staff, therapists, and parents [50]. We suggest that future research on psychoeducational content be supplemented by using different sources of information. Considering the disadvantages of response bias in interview and focus group data and the shortcomings in data collection control using self-reporting observational data, survey studies gathering information from a broader spectrum of youth could be considered, for instance, from younger adolescents or young people without symptoms of depression. We also suggest supplementary studies using different sources of information from the internet and social media platforms using the advantages of participatory observation, such as active netnographic approaches. Studies such as this could possibly expand our apprehension of young people's information needs and demands. The field requires a broader spectrum of research involving different methodologies, data sources, populations, and people contributing to the complex field of youth mental health.

Conclusions

We analyzed user-generated online text data from adolescents with symptoms of depression to investigate what components of psychoeducation they regard as relevant. For clinical use, we suggest that developers of psychoeducational content and future research focus on the following:

- Measures to promote self-management and information about therapy options and prognosis, including where to seek initial help and how to relate to a health professional.
- Measures to stimulate expectations of self-efficacy and portray therapy as an active guiding process, thereby stimulating help-seeking behavior, motivation, and therapy adherence.
- Information about etiology.
- Information about diagnosis to improve self-monitoring skills and thus to enhance relapse prevention. Emphasis on this topic may be more relevant when symptoms are less severe.

Acknowledgments

This work was supported by the Norwegian Research Council (grant agreement no. 262848).

Authors' Contributions

KKD designed this study within the framework of the multicenter project Social Health Bots, reviewed the literature, drafted the manuscript, and led the content analysis process and statistical calculation. PBB and AF contributed with data collection and file organization and preparation through the Social Health Bots project. AF reviewed and wrote parts of the manuscript. PBB reviewed the interpretation of the results and contributed to the theoretical viewpoints. ORH supervised the project and assisted in the preparation of the manuscript. AK coordinated and reviewed the manuscript and supervised the project. JIR contributed to the content analysis as a rater. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. Lancet 2013 Nov 9;382(9904):1575-1586. [doi: 10.1016/S0140-6736(13)61611-6] [Medline: 23993280]
2. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun 01;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
3. Kramer T, Garralda ME. Psychiatric disorders in adolescents in primary care. Br J Psychiatry 1998 Dec 03;173(6):508-513. [doi: 10.1192/bjp.173.6.508] [Medline: 9926080]
4. Jane Costello E, Erkanli A, Angold A. Is there an epidemic of child or adolescent depression? J Child Psychol Psychiatry 2006 Dec;47(12):1263-1271. [doi: 10.1111/j.1469-7610.2006.01682.x] [Medline: 17176381]
5. Crockett MA, Martínez V, Jiménez-Molina Á. Subthreshold depression in adolescence: gender differences in prevalence, clinical features, and associated factors. J Affect Disord 2020 Jul 01;272:269-276 [FREE Full text] [doi: 10.1016/j.jad.2020.03.111] [Medline: 32553367]
6. Haavet OR, Dalen I, Straand J. Depressive symptoms in adolescent pupils are heavily influenced by the school they go to. A study of 10th grade pupils in Oslo, Norway. Eur J Public Health 2006 Aug;16(4):400-404. [doi: 10.1093/eurpub/cki197] [Medline: 16207727]

7. Merikangas KR, He J, Burstein M, Swanson SA, Avenevoli S, Cui L, et al. Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication--Adolescent Supplement (NCS-A). J Am Acad Child Adolesc Psychiatry 2010 Oct;49(10):980-989 [FREE Full text] [doi: 10.1016/j.jaac.2010.05.017] [Medline: 20855043]
8. Mojtabai R, Olfson M, Han B. National trends in the prevalence and treatment of depression in adolescents and young adults. Pediatrics 2016 Dec 14;138(6):e20161878 [FREE Full text] [doi: 10.1542/peds.2016-1878] [Medline: 27940701]
9. Wiens K, Williams JV, Lavorato DH, Duffy A, Pringsheim TM, Sajobi TT, et al. Is the prevalence of major depression increasing in the Canadian adolescent population? Assessing trends from 2000 to 2014. J Affect Disord 2017 Mar 01;210:22-26. [doi: 10.1016/j.jad.2016.11.018] [Medline: 28012349]
10. Rao W, Xu D, Cao X, Wen S, Che W, Ng CH, et al. Prevalence of depressive symptoms in children and adolescents in China: a meta-analysis of observational studies. Psychiatry Res 2019 Feb;272:790-796. [doi: 10.1016/j.psychres.2018.12.133] [Medline: 30832200]
11. Birmaher B, Ryan ND, Williamson DE, Brent DA, Kaufman J. Childhood and adolescent depression: a review of the past 10 years. Part II. J Am Acad Child Adolesc Psychiatry 1996 Dec;35(12):1575-1583. [doi: 10.1097/00004583-199612000-00008] [Medline: 8973063]
12. Judd LL, Akiskal HS, Paulus MP. The role and clinical significance of subsyndromal depressive symptoms (SSD) in unipolar major depressive disorder. Journal of Affective Disorders 1997 Aug;45(1-2):5-18. [doi: 10.1016/s0165-0327(97)00055-4] [Medline: 9268771]
13. Lewinsohn PM, Rohde P, Seeley JR, Klein DN, Gotlib IH. Natural course of adolescent major depressive disorder in a community sample: predictors of recurrence in young adults. Am J Psychiatry 2000 Oct;157(10):1584-1591. [doi: 10.1176/appi.ajp.157.10.1584] [Medline: 11007711]
14. Yates P, Kramer T, Garralda E. Depressive symptoms amongst adolescent primary care attenders. Levels and associations. Soc Psychiatry Psychiatr Epidemiol 2004 Jul;39(7):588-594. [doi: 10.1007/s00127-004-0792-y] [Medline: 15243698]
15. Fusar-Poli P, Nelson B, Valmaggia L, Yung AR, McGuire PK. Comorbid depressive and anxiety disorders in 509 individuals with an at-risk mental state: impact on psychopathology and transition to psychosis. Schizophr Bull 2014 Jan 22;40(1):120-131 [FREE Full text] [doi: 10.1093/schbul/sbs136] [Medline: 23180756]
16. Häfner H, Maurer K, Trendler G, an der Heiden W, Schmidt M. The early course of schizophrenia and depression*. Eur Arch Psychiatry Clin Neurosci 2005 Jun;255(3):167-173. [doi: 10.1007/s00406-005-0584-8] [Medline: 15995900]
17. Burns JM, Andrews G, Szabo M. Depression in young people: what causes it and can we prevent it? Med J Aust 2002 Oct 07;177(S7):S93-S96. [doi: 10.5694/j.1326-5377.2002.tb04864.x] [Medline: 12358564]
18. Bramesfeld A, Platt L, Schwartz FW. Possibilities for intervention in adolescents' and young adults' depression from a public health perspective. Health Policy 2006 Dec;79(2-3):121-131. [doi: 10.1016/j.healthpol.2005.11.012] [Medline: 16414145]
19. Saxena S, Jané-Llopis E, Hosman C. Prevention of mental and behavioural disorders: implications for policy and practice. World Psychiatry 2006 Feb;5(1):5-14 [FREE Full text] [Medline: 16757984]
20. Birmaher B, Brent D, AACAP Work Group on Quality Issues, Bernet W, Bukstein O, Walter H, et al. Practice parameter for the assessment and treatment of children and adolescents with depressive disorders. J Am Acad Child Adolesc Psychiatry 2007 Nov;46(11):1503-1526. [doi: 10.1097/chi.0b013e318145ae1c] [Medline: 18049300]
21. Carrellas NW, Biederman J, Uchida M. How prevalent and morbid are subthreshold manifestations of major depression in adolescents? A literature review. J Affect Disord 2017 Mar 01;210:166-173. [doi: 10.1016/j.jad.2016.12.037] [Medline: 28049101]
22. Richardson T, Stallard P, Velleman S. Computerised cognitive behavioural therapy for the prevention and treatment of depression and anxiety in children and adolescents: a systematic review. Clin Child Fam Psychol Rev 2010 Sep;13(3):275-290. [doi: 10.1007/s10567-010-0069-9] [Medline: 20532980]
23. Rickwood DJ, Deane FP, Wilson CJ. When and how do young people seek professional help for mental health problems? Med J Aust 2007 Oct 01;187(S7):S35-S39. [doi: 10.5694/j.1326-5377.2007.tb01334.x] [Medline: 17908023]
24. Barney LJ, Griffiths KM, Jorm AF, Christensen H. Stigma about depression and its impact on help-seeking intentions. Aust N Z J Psychiatry 2006 Jan;40(1):51-54. [doi: 10.1080/j.1440-1614.2006.01741.x] [Medline: 16403038]
25. Brandtzaeg PB, Lüders M, Spangenberg J, Rath-Wiggins L, Følstad A. Emerging journalistic verification practices concerning social media. Journalism Practice 2015 Mar 13;10(3):323-342. [doi: 10.1080/17512786.2015.1020331]
26. Lamb C, Murphy M. The divide between child and adult mental health services: points for debate. Br J Psychiatry Suppl 2013 Jan;54:s41-s44. [doi: 10.1192/bjp.bp.112.119206] [Medline: 23288501]
27. Lopez MA, Toprac MG, Crismon ML, Boemer C, Baumgartner J. A psychoeducational program for children with ADHD or depression and their families: results from the CMAP feasibility study. Community Ment Health J 2005 Feb;41(1):51-66. [doi: 10.1007/s10597-005-2599-z] [Medline: 15932052]
28. Singh SP, Paul M, Ford T, Kramer T, Weaver T, McLaren S, et al. Process, outcome and experience of transition from child to adult mental healthcare: multiperspective study. Br J Psychiatry 2010 Oct;197(4):305-312. [doi: 10.1192/bjp.bp.109.075135] [Medline: 20884954]

29. Roche E, O'Sullivan R, Gunawardena S, Cannon M, Lyne JP. Higher rates of disengagement among young adults attending a general adult community mental health team: time to consider a youth-specific service? Early Interv Psychiatry 2020 Jun;14(3):330-335. [doi: 10.1111/eip.12860] [Medline: 31368245]
30. Richardson LP, Katzenellenbogen R. Childhood and adolescent depression: the role of primary care providers in diagnosis and treatment. Curr Probl Pediatr Adolesc Health Care 2005 Jan;35(1):6-24. [doi: 10.1016/j.cppeds.2004.09.001] [Medline: 15611721]
31. Donker T, Griffiths KM, Cuijpers P, Christensen H. Psychoeducation for depression, anxiety and psychological distress: a meta-analysis. BMC Med 2009 Dec 16;7:79 [FREE Full text] [doi: 10.1186/1741-7015-7-79] [Medline: 20015347]
32. Tursi MFDS, Baes CVW, Camacho FRDB, Tofoli SMDC, Juruena MF. Effectiveness of psychoeducation for depression: a systematic review. Aust N Z J Psychiatry 2013 Nov;47(11):1019-1031. [doi: 10.1177/0004867413491154] [Medline: 23739312]
33. Lewinsohn PM, Clarke GN, Hops H, Andrews J. Cognitive-behavioral treatment for depressed adolescents. Behavior Therapy 1990;21(4):385-401. [doi: 10.1016/s0005-7894(05)80353-3]
34. Walsh J. Psychoeducation in Mental Health. Chicago, IL: Lyceum; 2010.
35. Kellett S, Clarke S, Matthews L. Delivering group psychoeducational CBT in primary care: comparing outcomes with individual CBT and individual psychodynamic-interpersonal psychotherapy. Br J Clin Psychol 2007 Jun;46(Pt 2):211-222. [doi: 10.1348/014466506X146188] [Medline: 17524214]
36. Ardito RB, Rabellino D. Therapeutic alliance and outcome of psychotherapy: historical excursus, measurements, and prospects for research. Front Psychol 2011;2:270 [FREE Full text] [doi: 10.3389/fpsyg.2011.00270] [Medline: 22028698]
37. Hollis C, Morriss R, Martin J, Amani S, Cotton R, Denis M, et al. Technological innovations in mental healthcare: harnessing the digital revolution. Br J Psychiatry 2015 Apr;206(4):263-265. [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]
38. Christensen H, Griffiths KM, Jorm AF. Delivering interventions for depression by using the internet: randomised controlled trial. BMJ 2004 Jan 31;328(7434):265 [FREE Full text] [doi: 10.1136/bmj.37945.566632.EE] [Medline: 14742346]
39. Clarke AM, Kuosmanen T, Barry MM. A systematic review of online youth mental health promotion and prevention interventions. J Youth Adolesc 2015 Jan;44(1):90-113. [doi: 10.1007/s10964-014-0165-0] [Medline: 25115460]
40. Lukens EP, McFarlane WR. Psychoeducation as evidence-based practice: considerations for practice, research, and policy. Brief treatment and crisis intervention 2004;4:205-225. [doi: 10.1093/brief-treatment/mhh019]
41. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
42. Cavanagh K, Shapiro DA, Van Den Berg S, Swain S, Barkham M, Proudfoot J. The effectiveness of computerized cognitive behavioural therapy in routine care. Br J Clin Psychol 2006 Nov;45(Pt 4):499-514. [doi: 10.1348/014466505X84782] [Medline: 17076960]
43. Bevan Jones R, Thapar A, Stone Z, Thapar A, Jones I, Smith D, et al. Psychoeducational interventions in adolescent depression: a systematic review. Patient Educ Couns 2018 May;101(5):804-816 [FREE Full text] [doi: 10.1016/j.pec.2017.10.015] [Medline: 29103882]
44. Merry SN, Hetrick SE, Cox GR, Brudevold-Iversen T, Bir JJ, McDowell H. Psychological and educational interventions for preventing depression in children and adolescents. Cochrane Database Syst Rev 2011 Dec 07(12):CD003380. [doi: 10.1002/14651858.CD003380.pub3] [Medline: 22161377]
45. Fristad M, Goldberg AJ, Leffler J. Psychotherapy for Children with Bipolar and Depressive Disorders. New York: The Guilford Press; 2011:6-15.
46. Nangle D, Hansen D, Grover R, Kingery J, Suveg C. Treating Internalizing Disorders in Children and Adolescents: Core Techniques and Strategies. New York: The Guilford Press; Jun 24, 2016:139-161.
47. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [Medline: 847061]
48. Bandura A. On the functional properties of perceived self-efficacy revisited. J Manage 2011 Sep 27;38(1):9-44. [doi: 10.1177/0149206311410606]
49. Bru L, Solholm R, Idsoe T. Participants' experiences of an early cognitive behavioral intervention for adolescents with symptoms of depression. Emot Behav Diffic 2013 Mar;18(1):24-43. [doi: 10.1080/13632752.2012.675138]
50. Bevan Jones R, Thapar A, Rice F, Beeching H, Cichosz R, Mars B, et al. A web-based psychoeducational intervention for adolescent depression: design and development of MoodHwb. JMIR Ment Health 2018 Feb 15;5(1):e13 [FREE Full text] [doi: 10.2196/mental.8894] [Medline: 29449202]
51. Mufson L. Interpersonal Psychotherapy for Depressed Adolescents. 2nd Ed. New York: Guilford Press; 2004.
52. Hall EB, Mufson L. Interpersonal psychotherapy for depressed adolescents (IPT-A): a case illustration. J Clin Child Adolesc Psychol 2009 Jul 14;38(4):582-593. [doi: 10.1080/15374410902976338] [Medline: 20183644]
53. Clarke G, Lewinsohn PM. The Coping With Depression course: a group psychoeducational intervention for unipolar depression. Behav. change 2014 Oct 06;6(2):54-69. [doi: 10.1017/s0813483900007592]
54. Miller WR. Motivational Interviewing: Helping People Change. 3rd ed. Rollnick S., editor. New York: Guilford Press; 2013.

55. Crabtree BL, Miller WF. A template approach to text analysis: developing and using codebooks. In: Crabtree BL, Miller WF, editors. Doing Qualitative Research. Newbury Park, CA: Sage Publications; 1992:93-109.
56. Fatima I, Mukhtar H, Ahmad HF, Rajpoot K. Analysis of user-generated content from online social communities to characterise and predict depression degree. Journal of Information Science 2017 Nov 14;44(5):683-695. [doi: 10.1177/0165551517740835]
57. Seabrook EM, Kern ML, Fulcher BD, Rickard NS. Predicting depression from language-based emotion dynamics: longitudinal analysis of Facebook and Twitter status updates. J Med Internet Res 2018 May 08;20(5):e168 [FREE Full text] [doi: 10.2196/jmir.9267] [Medline: 29739736]
58. Park S, Lee SW, Kwak J, Cha M, Jeong B. Activities on Facebook reveal the depressive state of users. J Med Internet Res 2013 Oct 01;15(10):e217 [FREE Full text] [doi: 10.2196/jmir.2718] [Medline: 24084314]
59. Ellis A, Dryden W. The Practice of Rational Emotive Behavior Therapy. New York: Springer Publishing Company; 1997.
60. Machin D, Machin D. Sample Size Tables for Clinical, Laboratory and Epidemiology Studies. Hoboken, New Jersey: Wiley; 2018.
61. Farrand P, Perry J, Lee C, Parker M. Adolescents' preference towards self-help: implications for service development. Prim Care Community Psychiatr 2006 Oct 01;11(2):73-79 [FREE Full text] [doi: 10.1185/135525706x121129]
62. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. Psychother Theor Res Pract 1979;16(3):252-260. [doi: 10.1037/h0085885]
63. Bandura A. Self-Efficacy: The Exercise of Control. New York: Freeman; 1997.
64. Colom F. Keeping therapies simple: psychoeducation in the prevention of relapse in affective disorders. Br J Psychiatry 2011 May;198(5):338-340. [doi: 10.1192/bjp.bp.110.090209] [Medline: 21525516]

Abbreviations

CBT: cognitive behavioral therapy
DALY: disability-adjusted life-year
MDD: major depressive disorder
Q&A: question and answer

Edited by R Kukafka; submitted 15.03.21; peer-reviewed by R Bevan Jones, X Zhou; comments to author 25.05.21; revised version received 19.07.21; accepted 12.08.21; published 30.09.21.

Please cite as:
Dysthe KK, Haavet OR, Røssberg JI, Brandtzaeg PB, Følstad A, Klovning A
Finding Relevant Psychoeducation Content for Adolescents Experiencing Symptoms of Depression: Content Analysis of User-Generated Online Texts
J Med Internet Res 2021;23(9):e28765
URL: https://www.jmir.org/2021/9/e28765
doi:10.2196/28765
PMID:34591021

Original Paper

Identifying False Human Papillomavirus (HPV) Vaccine Information and Corresponding Risk Perceptions From Twitter: Advanced Predictive Models

Tre Tomaszewski[1], MS; Alex Morales[2], PhD; Ismini Lourentzou[3], PhD; Rachel Caskey[4], MD; Bing Liu[5], PhD; Alan Schwartz[6], PhD; Jessie Chin[1,7], PhD

[1]School of Information Sciences, University of Illinois at Urbana-Champaign, Champaign, IL, United States
[2]Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, United States
[3]Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States
[4]College of Medicine, University of Illinois at Chicago, Chicago, IL, United States
[5]Department of Computer Science, University of Illinois at Chicago, Chicago, IL, United States
[6]Department of Medical Education, University of Illinois at Chicago, Chicago, IL, United States
[7]Cancer Center at Illinois, University of Illinois at Urbana-Champaign, Urbana, IL, United States

Corresponding Author:
Jessie Chin, PhD
School of Information Sciences
University of Illinois at Urbana-Champaign
501 E Daniel St
Champaign, IL, 61820
United States
Phone: 1 217 333 0125
Email: chin5@illinois.edu

Abstract

Background: The vaccination uptake rates of the human papillomavirus (HPV) vaccine remain low despite the fact that the effectiveness of HPV vaccines has been established for more than a decade. Vaccine hesitancy is in part due to false information about HPV vaccines on social media. Combating false HPV vaccine information is a reasonable step to addressing vaccine hesitancy.

Objective: Given the substantial harm of false HPV vaccine information, there is an urgent need to identify false social media messages before it goes viral. The goal of the study is to develop a systematic and generalizable approach to identifying false HPV vaccine information on social media.

Methods: This study used machine learning and natural language processing to develop a series of classification models and causality mining methods to identify and examine true and false HPV vaccine–related information on Twitter.

Results: We found that the convolutional neural network model outperformed all other models in identifying tweets containing false HPV vaccine–related information (*F* score=91.95). We also developed completely unsupervised causality mining models to identify HPV vaccine candidate effects for capturing risk perceptions of HPV vaccines. Furthermore, we found that false information contained mostly loss-framed messages focusing on the potential risk of vaccines covering a variety of topics using more diverse vocabulary, while true information contained both gain- and loss-framed messages focusing on the effectiveness of vaccines covering fewer topics using relatively limited vocabulary.

Conclusions: Our research demonstrated the feasibility and effectiveness of using predictive models to identify false HPV vaccine information and its risk perceptions on social media.

(J Med Internet Res 2021;23(9):e30451) doi:10.2196/30451

KEYWORDS

misinformation; disinformation; social media; HPV; human papillomavirus vaccination; vaccination; causality mining; cause; effect; risk perceptions; vaccine; perception; risk; Twitter; machine learning; natural language processing; cervical cancer

Introduction

About 13,000 women are newly diagnosed with invasive cervical cancer and over 4000 women die from it every year [1]. Cervical cancer is caused by certain types of human papillomavirus (HPV) [2,3]. HPV is the most common sexually transmitted infection in the United States with an estimated 6.2 million new infections every year among persons 14 to 44 years of age [4-6]. In addition to cervical cancer, HPV is the causal mediator in multiple head and neck cancers, genital cancers, and anal cancers [7-9]. The overall burden of HPV-associated cancers has been increasing in the United States [9]. Prevention of HPV is more challenging than most sexually transmitted infections as condoms do not provide complete protection against infection [10]. Hence, prevention through vaccination is critical in decreasing the burden of cancer due to this ubiquitous infection.

The HPV vaccine is universally recommended for all adolescents [10]. Despite the exceptional efficacy (up to 90% protection) in preventing precancerous lesions caused by the targeted HPV types [11-13], only 56.8% of 13 to 17-year-old females and 51.8% of 13 to 17-year-old males in the United States have completed the HPV vaccine series [14]. There are many known barriers to HPV vaccination, including misconceptions about the side effects and adverse events from HPV vaccines, misbeliefs around the need for vaccines, inconsistent advice received from health care givers, costs to complete the vaccination, limited access to clinics, and violations to cultural beliefs [15-21]. Among these barriers, the bias in risk perceptions has not only been associated with low intention of vaccination [22-25] but also with the actual vaccination behavior [16,22,26-30]. The National Immunization Survey revealed the top 3 parental concerns of HPV vaccines to be a lack of knowledge, low perceived usefulness of vaccine (low perceived risk of HPV infection), and high perceived risks of side effects and safety concerns [31], underscoring the importance of risk perceptions in HPV vaccination decisions.

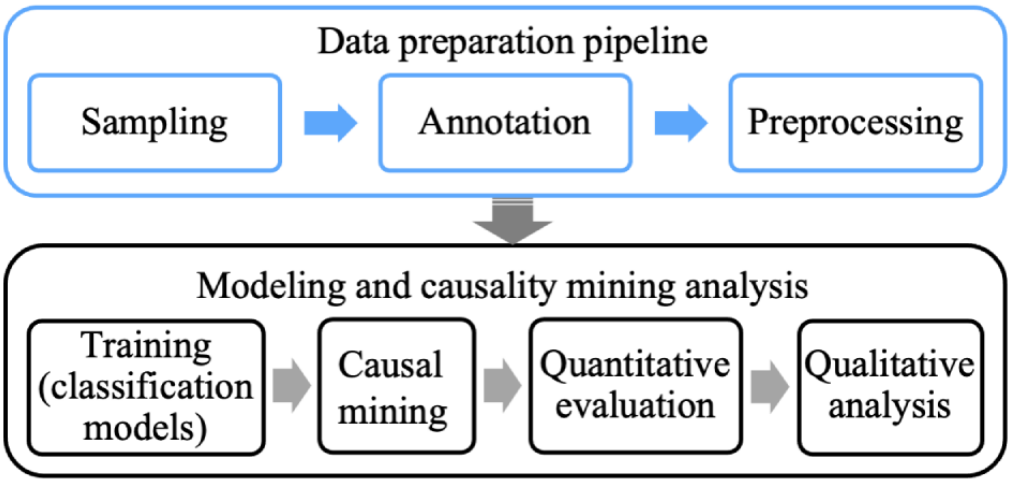
Social media has become an important information source for people to exchange vaccine-related opinions and form their attitudes toward vaccines [32-38]. Its impact is striking, especially for Twitter vaccine information, as HPV vaccine–related opinions on Twitter have been associated with actual vaccine acceptance and coverage [39]. Existing research has investigated the emerging themes and public attitudes toward the pro- and antivaccine online discussions about HPV vaccines [19,24,40-44]. Although multiple false conspiracies and myths around HPV vaccines have been identified, no research has used an automatic computational approach to extract the causal cues of the main vaccination arguments used in circulation of HPV vaccine misinformation. Research has shown that none of content quality, scientific robustness, or the veracity of the information has been found to be indicative of the spread of information, while false or unverified information sometimes becomes more viral than true information [45,46]. As attention to the propagation of false information on social media has surged [45,47-50], an automatic, systematic, and generalizable approach to detect socially endorsed false health information remains understudied. The threats of false information are critical because the reliance (ie, perceived accuracy) on false information can be amplified with each exposure to it and further magnified through social networks [51-53]. People can be especially victimized by the proliferation of user-generated false health information given their lack of health literacy, incompetence in credibility judgments, and the mixed quality of health news despite the sources cited [54-56]. Hence, detecting false health information before it propagates is an important step toward minimizing the threats of false information [57].

Several works have targeted health misinformation [58,59], with most studies using descriptive approaches to study known health misinformation and performing analysis to uncover the common misbeliefs, demographic and geographic patterns, and social media user behaviors [25,60,61]. A few studies have implemented computational models to identify health misinformation from other social medial platforms (such as YouTube and Instagram); however, none of them have attempted to identify health misinformation from short and sometimes incomplete text information, such as tweets [62,63]. In contrast with other related work, we combined a classification model for identifying false HPV vaccine information with unsupervised causality mining to extract the risk perceptions considered to be the attributable causes of HPV antivaccine health concerns based on the content expressed in Twitter messages. To this end, we conducted an infodemiology study to use natural language processing and machine learning methods, such as classification, clustering, dependency parsing, and phrase mining, to identify those false HPV vaccination arguments that frequently appear in social media. Our methodological analysis can be applied to other domains, such as COVID-19 vaccination, food safety, and politics, to extract insightful information regarding the differences and similarities between truthful and misleading claims shared online.

Methods

We collected a corpus related to HPV vaccines with tweets published from December 2013 until December 2017. We used the formerly known Crimson Hexagon's (now Brandwatch) social media analytics application programming interface and a list of HPV-related search terms, including, but not limited to, “HPV vaccine,” “papillomavirus vaccine,” “cervical cancer vaccine,” “HPV shot,” “cervical cancer shot,” and “Gardasil.” Our modeling pipeline consists of several steps: sampling, annotation and data preprocessing, training, and analysis (see Figure 1). The data preprocessing stage includes rule-based lexical normalization and unsupervised pretraining of word embeddings.

Figure 1. Causality mining data collection and modeling pipeline.



First, we randomly sampled 1000 tweets per year and passed them to 2 annotators in 2 rounds. Both annotators received basic training about HPV vaccines (including extensive reading of the verified HPV vaccine-related materials from the National Cancer Institute, Center for Disease Control and Prevention, and American Cancer Society), and 1 had formal educational training in medical sciences. Similar to related work in misinformation detection [64-66], we framed the task as a binary classification, in which each tweet is categorized as true or false information (in which false information includes partial-false or partial-true information). We thus not only asked the annotators to judge the veracity of the content for each tweet but also allowed them to select an additional option as "not applicable" for tweets that did not fall under any of the 2 categories (eg, opinionated text and other nonfactual or irrelevant posts). Tweets labeled as not applicable were filtered out from the annotation pipeline. Any discrepancies of the ratings from the 2 annotators were reconciled through discussion. For the interrater reliability, a Cohen's kappa coefficient (κ) of 0.75, was considered to indicate good agreement on the task [67]. The resulting data set consisted of 5000 labeled and 702,858 unlabeled tweets. Character lengths of the tweets, including all mentions, retweets, and hashtags, ranged from 21 to 826 characters.

To reduce vocabulary size for the lexical normalization steps, words were formatted in lower case and URLs were removed; numerals, and Twitter-specific items, such as user mentions (usernames prefixed by "@") or retweets, were tagged and mapped to a common special token per category (ie, NUMBER, MENTION, RT, respectively). Selected contractions were then replaced with their canonical forms: for example, "Can't" was replaced with "Cannot," "You'll" was replaced with "You will," "&" was replaced with "and," etc. Additionally, hyphens and forward slashes were replaced with spaces, alphanumeric pairings were processed, instances of 2 or more user mentions were reduced to 2 "MENTION" tokens, hashtag quotes and other types of punctuation were removed, and multiple leading or trailing white spaces were replaced with a single one. This process reduced the length of each tweet, which could range between 18 and 295 characters.

The final vocabulary size based on the training set was 4098 terms (including 1 vocabulary term representing a blank space). Analysis of terms weighted by their frequency odds ratio (ie, the ratio of occurrence in each category) showed certain terms were overrepresented in the true category but appeared infrequently in the false category, for example, words that strongly indicated the effectiveness of HPV vaccines on cancer prevention spread online, such as "prevent," "protect," and "effective." On the other hand, false messages contain terms such as "danger," "adverse," and "deadly," and focus more on the negative causal effects that are used as arguments for vaccination.

Results

Classification Model

Word embeddings map discrete word tokens to real-valued vector representations, where semantically similar words have similar vectors and are therefore closer in the embedding space. In general, pretraining of such word embeddings has been found to be beneficial for several natural language processing tasks, allowing for faster model convergence and task performance improvements. Therefore, we trained an unsupervised embedding model, FastText [68], with our full Twitter collection as training data and with the aforementioned preprocessing. Compared to other word representation models, FastText can produce word vectors for out-of-vocabulary words and has been proven to be a strong baseline for short text similarity, with its open-sourced implementation allowing for faster training [69]. More specifically, FastText produced 300-dimensional vector representations for each term in our vocabulary, which was used as the initialization for our model's embedding layer. We also experimented with Wikipedia-pretrained embeddings and without any pretrained embeddings: our experiments showed that the model performed better in terms of accuracy when initialized with HPV-related pretrained word embeddings.

Finally, we divided the annotated data into 60% training, 20% validation, and 20% testing, keeping the same splits across all models for a fair comparison. Deduplication of tweets with exact matches within each set left 3661 tweets in total (2142 for training, 758 for validation, and 761 for the test set). We

experimented with several model architectures, including convolutional neural network (CNN) [70], bidirectional long short-term memory (BiLSTM), and traditional models, including support vector machine and Naive Bayes. We trained with cross-entropy, Adaptive Moment Estimation with a 10^{-4} learning rate, 0.01 decay, and a 32 batch size for the neural models. Hyperparameter tuning was performed using the Tune library [71]. In Table 1, we report the mean and SD of the top-5 performing model variations. Our experimental evaluation showed that CNNs performed better than did the other models (see Figures 2 and 3 for respective confusion matrices and the area under the receiver operating characteristic curve comparisons between neural networks). Of the top-5 best performing models for either of the neural networks, the CNN required less training time than did the BiLSTM. The mean training time per epoch for the CNN was 11.5 ms (SD 1.09, minimum 16, maximum 16, median 12), whereas the mean training time per epoch for the BiLSTM was 51.3 ms (SD 34.07, minimum 14, maximum 88, median 81). Our best-performing CNN model had 256 convolutional filters, including –3 kernels of width (3,4, and 5) and rectified linear unit nonlinearities; a max pooling layer, a fully connected layer of 128 units with rectified linear unit activations and 0.1 dropout, and a final softmax output layer that produced the classification prediction.

Table 1. Identifying false human papillomavirus vaccine information: classification model comparison.

Model	Accuracy	Precision	Recall	*F* score
SVM[a], mean	57.424	57.806	56.721	55.532
Naive Bayes, mean	51.774	52.485	52.301	51.090
CNN[b], mean (SD)	91.958 (0.269)	91.953 (0.272)	91.946 (0.271)	*91.946 (0.270)* [c]
BiLSTM[d], mean (SD)	91.643 (0.432)	91.710 (0.396)	91.574 (0.453)	91.618 (0.438)

[a]SVM: support vector machine.

[b]CNN: convolutional neural network.

[c]Italics indicate the highest *F* score in the table.

[c]BiLSTM: bidirectional long short-term memory.

Figure 2. Confusion Matrix for best-performing CNN model. BiLSTM: bidirectional long short-term memory; CNN: convolutional neural network.

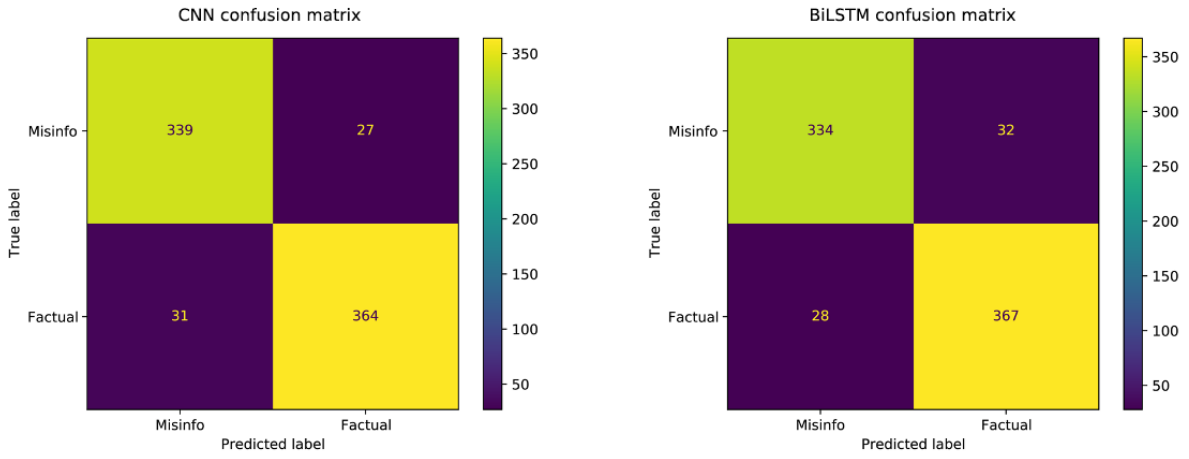
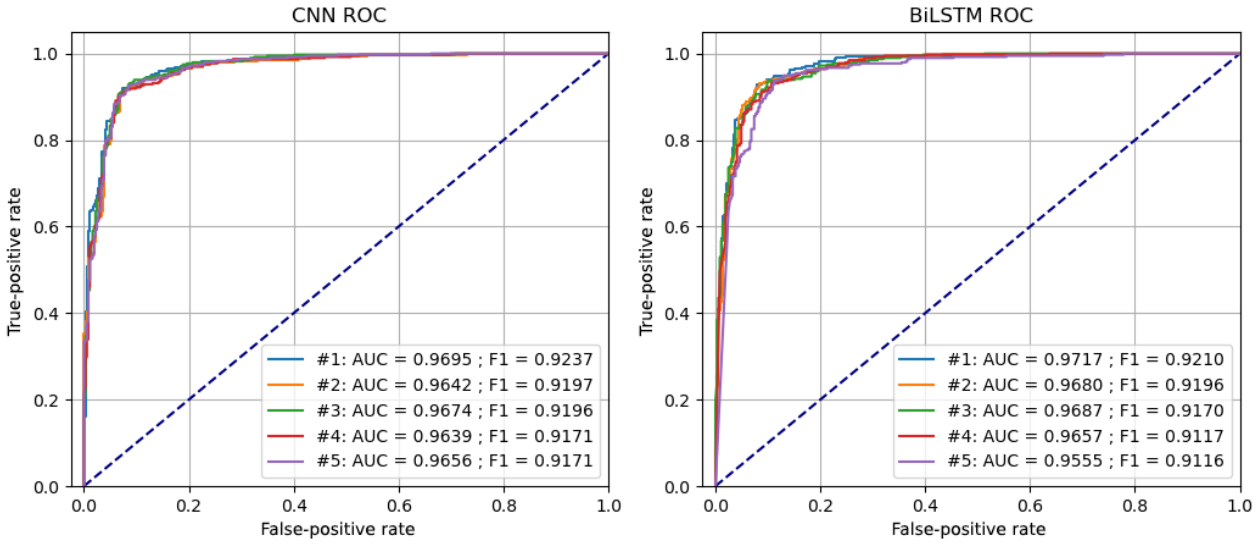


Figure 3. ROC for the best-performing convolutional neural network and bidirectional long short-term memory models. AUC: area under the curve; ROC: receiver operating characteristic.



Causality Mining

To identify the risk perceptions attributed to the HPV vaccines, we first applied our classifier to a set of 291,037 tweets from which we are able to tag 124,031 as false tweets and 167,006 as true tweets. Using a dictionary of causal terms derived by Kayesh et al [72] for Twitter causality detection, we screened for tweets that contained at least 1 of these terms and kept tweets classified as false information if the classification confidence was at least 0.998, as this maintained high fidelity with our classifier. Thus, a total of 9352 tweets were used for the causal relationship mining process (Table 2). We then used a dependency parser for tweets to tag and merge multi-word expressions [73]. As tweets can have multiple utterances (ie, independent sentences or fragments), we kept the noun phrases that appeared with the causal cue regardless of whether they had a dependency related to the causal cue, which is in contrast to the work by Kayesh et al [72]. A candidate causal phrase is a set of terms pertaining to a tweet that contains a causal cue and precipitates the candidate effect phrase.

Table 2. Number of messages after applying several filters.

Model	False	True	Total
No filter, n	124,031	167,006	291,037
+ Confidence threshold, n (%)	72,172 (58.19)	105,166 (62.97)	177,338 (60.93)
+ Contains causal cue, n (%)	3667 (2.96)	5685 (3.40)	9352 (3.21)

We could then compute the pointwise mutual information (PMI) for the causal set $C = \{c_1,...,c_m\}$ and the effect set $E = \{e_1,...,e_m\}$ where the candidate causal phrase, c_i and effect phrase, e_j, are sets that contain terms $w_c \in \boldsymbol{V}_c$ and $w_e \in V_e$, respectively. Here, $\boldsymbol{V}_c$ is the set of terms, noun phrases, and multi-word expressions derived from candidate causal phrases in the tweets (excluding terms with a minimum frequency of 1 and removing stopwords) and V_e is the vocabulary derived from the candidate effect phrases.

To compute the PMI for terms $w_e \in e_j$ and $w_c \in c_i$ we have,

We can apply Laplace smoothing to ensure the probability distributions are nonzero [74] and can compute the normalized pointwise mutual information (NPMI) [75] as follows:

The range of values of NPMI are from –1 to 1, where –1 means the terms never occur together, 0 means they are independent, and 1 is complete co-occurrence.

Collapsing Candidate Effect Phrases and Ranking Effects

As our model was completely unsupervised and included retweets, tweet messages could become very redundant, but our method could detect many near-duplicate candidate effect phrases. To collapse these phrases, we clustered the terms using semantic similarity derived from embedding representations of the candidate effect phrases. In particular, we used the package HuggingFace [76] to acquire the sum of the last 4 layers of the bidirectional encoder representations from transformers model [77]. To compute the word embedding, we then averaged these word embeddings in the candidate effect phrase to produce the embedding vectors.

Density-based spatial clustering of applications with noise (DBSCAN) [78] was then used to cluster the candidate effect phrases. There were 2 parameters of importance: (1)

reachability, which is the max distance between the 2 "points" to be considered in the same cluster; and (2) the minimum number of "points" to be considered clusters. By points here we mean an embedded real-valued vector representation of the effect sets. We set the reachability to 0.1 and the minimum number of points to 1, as we wanted to retrieve only the closest semantically similar words while maintaining meaningful clusters. Note that there are other alternatives to DBSCAN, such as ordering points to identify the clustering structure (OPTICS) [79] and hierarchical density-based spatial clustering of applications with noise (HDBSCAN) [80] for spectral clustering; however, we only needed to reduce the number of effects compared at query time, meaning that DBSCAN was sufficient. We then selected the cluster cores as representatives for each cluster to collapse the effects.

To identify the perceptions associated with different causal words, we formulated this as a retrieval problem. Given a causality-related query q, we ranked the associated effects by using the NPMI. To compute the scoring function, we used the following:

This scoring function computes the average NPMI for all pairs of terms in the query and candidate effect phrase. We can then compute the cumulative NPMI score for a category C_a of effect phrases as follows:

Causality Mining Results

To validate the candidate causal mining approach, we took a lexicon pertaining to risk perceptions (ie, perceived effects) concerning HPV vaccines. The HPV-Vaccine Risk Lexicon (HPVVR) is a consumer-facing lexicon to capture how laymen describe their risk perceptions about HPV vaccines (including their perceived harms and benefits about HPV vaccines) [81]. The HPVVR was developed in 2 stages. The first stage involved adopting the risk expressions and HPV-vaccine–related consumer-facing vocabulary from the Department of Homeland Security Risk Lexicon, MedlinePlus Consumer Health Topic Vocabulary, and Consumer Health Vocabulary (Unified Medical Language System) [82,83]. The second stage was to extract layman language about the descriptions of risk perceptions based on the user-generated content (from randomly sampled user-generated content from 2013 to 2018, including from Twitter and Facebook) by 2 trained annotators (interrater reliability: Cohen's kappa coefficient (κ)=0.80). The HPVVR covers more than 200 terms or phrases across 29 categories of risk perception-related vocabulary.

This gold standard list of effects, G, was then matched with the effect set E. In particular, we defined a partial match to be present if some terms in the ground truth effect phrase were matched with some of the candidate effect phrase (ie, $g \cap e \neq \emptyset$). For example, we mined "prevent throat cancer" from the data, which is a partial match to "prevent cervical cancer" in the HPVVR. There were 2 other kinds of partial match. We defined a match to be proper if the candidate effect phrase was a more specific example of the ground truth effect phrase, $g \in G$. For example, we mined "early onset menopause" from the data, which is a proper match to "menopause" in the HPVVR. We considered a reverse match to be present if the candidate effect phrase was a more general form of the ground truth effect phrase, $e \subseteq g$. For example, we mined "fatigue" from the data, which is a reverse match to "extreme fatigue" in the HPVVR. Out of a total of 136 ground-truth effect phrases, we found 55 (40.4%) matches, 78 (57.4%) partial matches, 48 (35.3%) reverse matches, and 103 (75.7%) either partial or proper matches or both. Meanwhile, there were also some candidate effect phrases which were newly discovered effects.

As the causality mining method is a completely bottom-up, unsupervised method, we could automatically mine candidate effects for any set of tweets. In particular, for the predicted false tweets, one of the largest candidate effect clusters contained terms relating to the reactions of different entities, such as "Japan," "Denmark," and "college," on the potential issues with the HPV vaccine, such as "recall," "lose support," and "banned." Another such cluster contained terms relating to infertility misconceptions of the HPV vaccine, such as "premature ovarian failure" and "early menopause on young girls." Another large candidate effect cluster was about the misconceptions of severe adverse events and complications, such as "sudden death," "paralysis," and "stroke." Note that it is possible that some candidate effect phrases may not be directly related to health effects. Thus, to alleviate this limitation in further analysis, we limited the effects to the terms in the ground truth (ie, $V_e = G$).

Discussion

Principal Results

The performance of the CNN and BiLSTM models used in this study showed the feasibility of discerning misinformation from factual information regarding HPV vaccines using the text of tweets. On average, both models predicted either class with high confidence. Although both models performed almost identically in terms of accuracy (and confidence) during testing, the CNN trained much more expediently than did the BiLSTM model, leading to its choice as the preferred model.

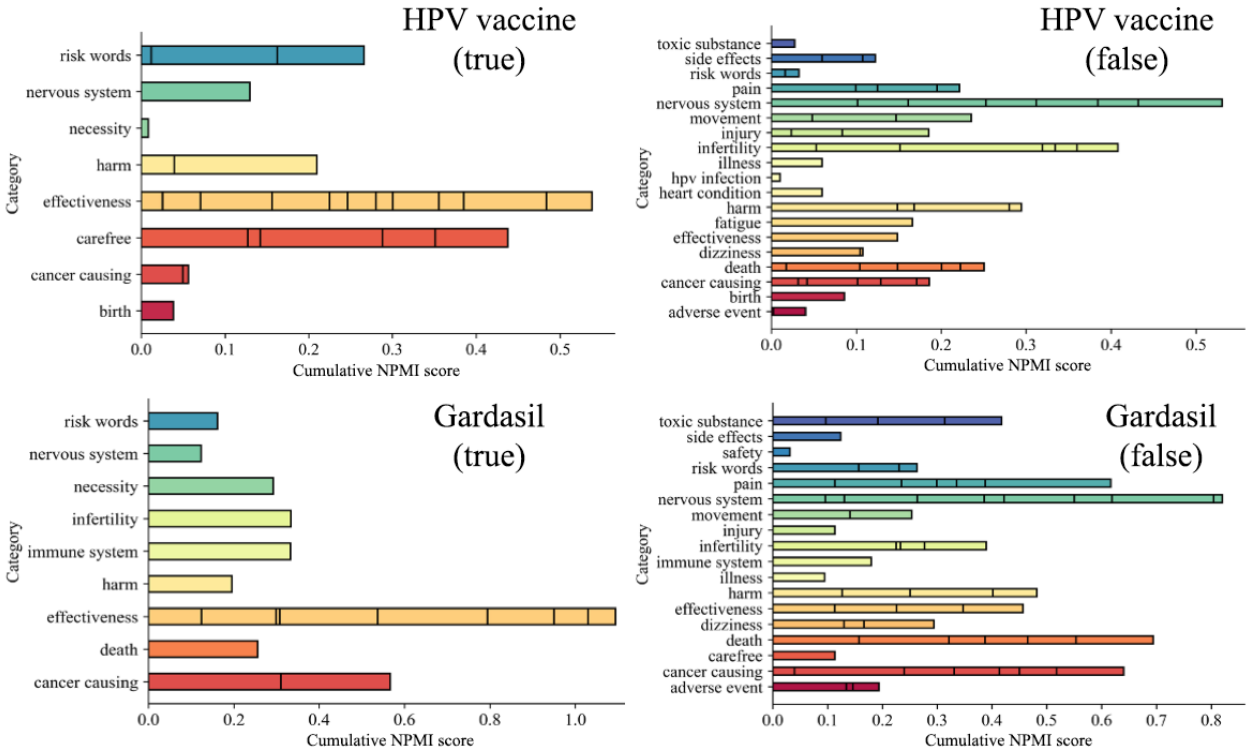
To examine the risk perceptions pertaining to HPV vaccines, we leveraged the false information classifier and the effect ranker. Figure 4 shows the cumulative NPMI scores after our effect ranker querying for both "HPV Vaccine" and "Gardasil" was applied. We could categorize these perceptions around the costs and benefits of HPV vaccines. In general, people discussed the benefits or low risk of harms in the true HPV vaccine tweets and various adverse events in the false HPV vaccine tweets. The main effects associated with the HPV vaccines in the true HPV vaccine tweets were about the prevention of HPV infection–related cancer and the denial of risk of increased unprotected sexual behavior of the vaccinated teens. The main effects associated with the HPV vaccines in the false HPV vaccine tweets were regarding infertility-related conditions (such as ovarian injury), child developmental disorder, death, and toxic ingredients in the HPV vaccines. Following our previous work on the patient-driven HPVVR, the findings from causality mining aided us in identifying the major concerns

related to HPV vaccines, whose solutions could then be prioritized.

The results show that false HPV vaccine messages not only span a wide variety of topics in risk perceptions but also involve a more diverse vocabulary to describe these topics compared to the fewer topics and relatively limited terminology found in true messages. This phenomenon of medical-based term frequency and topic diversity within false or misleading messages has also been noted in similar work regarding anti- and provaccine literature [84]. A possible explanation for this discrepancy is that true information requires an evidentiary consensus, thereby restricting terminology and outcomes to a specific selection of topics or phrases used to describe these topics. Misinformation lacks such restrictions to terms or outcomes and tends to use narrative language or mention novel topics to gauge attention [46,85].

We also observed differences in message framing in true and false HPV vaccine messages through causality mining (see Figure 4). True information contained both gain-framed and loss-framed messages, especially those highlighting the effectiveness of the vaccine at preventing HPV-related cancers, the link between HPV infection and cancer, and negating the potential harms of vaccines such as carefree or unprotected sexual behavior (Figure 4). Conversely, false messages were largely loss-framed, focusing on negative outcomes purportedly caused by the vaccine, such as those causing HPV-related cancer or other serious adverse events (infertility, neurological disorder, or death; Figure 4). The use of the risk-indicating causative verb (eg, vaccines "prevent" versus vaccines "harm" or "cause", etc) might be diagnostic for differentiating the true and false information. Future studies should leverage previous findings on the effectiveness of message framing to examine the impact of misinformation with different framing [86,87].

Figure 4. The cumulative NMPI scores when querying for "HPV Vaccine" and "Gardasil". The sections in the bar width correspond to the NPMI contribution of effect terms for each category. HPV: human papillomavirus; NPMI: normalized pointwise mutual information.



Comparison With Prior Work

Health-related misinformation research spans a broad range of disciplines [58,59], with several studies focusing on different medical domains, such as cancer, sexually transmitted disease and infections, influenza, and more recently, COVID-19 [25,60,61]. In vaccine-related domains, several papers have examined vaccine behavior as well as geographic and demographic patterns on the dissemination of antivaccine and misinformation tweets in social media with respect to autism spectrum disorder [60], influenza (flu) vaccines [88], and cancer treatments [89]. Several research endeavors tackle key issues, such as mitigating label scarcity with additional weak social supervision signals, improving intractability with attention mechanisms, and leveraging network and group or user information [65,90-92]. In general, the distinction between vaccine hesitancy identification and vaccination behavior detection is that the former involves an attitude or stance, while the latter is concerned with detecting the action of getting vaccinated [93]. Our study is more similar to the research in vaccine hesitancy but differs in that we focused on extracting causality from tweets through examining risk perceptions; attributable causes of HPV vaccine–related health concerns or expected gain; and using natural language processing, machine learning, and unsupervised causal mining techniques.

We observed that the convolutional models with multiple filter sizes [70,94] worked better than did BiLSTM models for domains with short text, such as tweets. Intuitively, the CNN

architecture captures the most common n-grams (of lengths 3, 4, and 5) and therefore is better at discovering discriminative text patterns in short text. Although we tested more sophisticated BiLSTM architectures, overall, the CNN model performed better than did the other model variations and was faster to train. These findings can be useful for social media health–related analysis, in particular with regards to the set of models that practitioners in this domain should explore for social media text classification.

With respect to causality mining, early works use hand-coded, domain-specific knowledge bases [95,96]. A challenge in identifying causal relations is the variety of ways in which it can be observed via various linguistic constructions. A previous study [97] showed that a classifier can determine whether a causal-verb expression, automatically extracted from predefined linguistic patterns of the form <noun phrase–verb–noun phrase> is a causal relationship or not. However, supervised methods require large amounts of manually annotated causes and effects and are thus resource demanding. Recent work has compared unsupervised methods for causal relationship mining including co-occurrence methods, such as point-wise mutual information, and discourse cue-based methods, which are based on information retrieval techniques, to count the number of matches in a cause-effect query [98]. Such comparisons were performed on large-scale document collections, and thus their insights are not applicable to our tasks that, in contrast, have limited amounts of data. Finally, event causality detection in tweets restricts the causal relationship mining to certain events of interest. In "Event causality detection in tweets by context word extension and neural net" [72], the authors propose an approach to encode both the candidate causal phrase and the candidate effect phrase for developing a feed-forward network classifier. Our method is not restricted to certain events. Most importantly, we focus on health-related messages pertaining to HPV vaccines, an approach which can be generalized to other health topics.

Limitations

One common bottleneck when applying supervised learning methods is the requirement of large amounts of high-quality annotated data for training. Due to the complex nature of the task-at-hand, and the need for extensive manual effort, our data set size might be restrictive in providing insights that can generalize across other domains and data sources. Additionally, due to frequent linguistic variations found in informal user-generated language, closely worded instances might have evaded deduplication. In the future, we hope to address the shortage of available labeled data by incorporating weak supervision methods and denoising mechanisms. Nevertheless, we chose to continue with supervised learning for higher precision, as weak supervision may result in label noise being injected into the false information detection models and thus affecting the subsequent causality mining steps.

Another limitation stems from the misalignment of model confidence and accuracy. In other words, model confidence might not be indicative of model correctness, a problem that is well-known in the machine learning research community [99]. In our experiments, we observed that the BiLSTM model produced high confidence estimates for most false negatives (ie, it misplaced more confidence when predicting factual text), while the CNN model had an equal number of false positives and false negatives for high-confidence examples. Approximately 20% of CNN's incorrect predictions had low confidence. Overall, the BiLSTM model seems to be overconfident in one direction and could be potentially calibrated better. Further analyses on these high-confidence inaccurate predictions are required to discover interpretable patterns that can identify misinformation subtopics and statements that share strong similarities to factual counterparts.

Finally, we should note that any use of additional metadata requires caution, especially for information that is added by the user, such as user profile characteristics, as well as reported timestamps and social network links, as recent studies show that misinformation spreaders tend to manipulate not only social network structure by forming groups to increase influence [100] but also several types of metadata [101]. In this study, we did not use these types of additional data sources, and thus we can only interpret content-based results and not along any other dimension other than the relationships found in the text.

Conclusions

The study has demonstrated a systematic, automatic approach to developing computational models for identifying false HPV vaccine–related information and its associated effects on social media. This approach could be generalized to other social media health information and provide insights into estimating the potential effects of a given health topic.

Acknowledgments

We thank Chieh-Li (Julian) Chin and Wen-Yuh Su for their assistance in the data collection and preparation for this study. We also thank the Technology Services and the Cancer Center at Illinois at the University of Illinois at Urbana-Champaign for their support of our study. This study was supported by a grant from the National Institutes of Health (#1R21CA237483). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Siegel R, Miller K, Jemal A. Cancer statistics, 2019. CA A Cancer J Clin 2019 Jan 08;69(1):7-34 [FREE Full text] [doi: 10.3322/caac.21551]
2. Schiffman MH, Bauer HM, Hoover RN, Glass AG, Cadell DM, Rush BB, et al. Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. J Natl Cancer Inst 1993 Jun 16;85(12):958-964. [doi: 10.1093/jnci/85.12.958] [Medline: 8388478]
3. Bosch FX, Manos MM, Muñoz N, Sherman M, Jansen AM, Peto J, et al. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International biological study on cervical cancer (IBSCC) Study Group. J Natl Cancer Inst 1995 Jun 07;87(11):796-802. [Medline: 7791229]
4. Dunne EF, Unger ER, Sternberg M, McQuillan G, Swan DC, Patel SS, et al. Prevalence of HPV infection among females in the United States. JAMA 2007 Feb 28;297(8):813-819. [doi: 10.1001/jama.297.8.813] [Medline: 17327523]
5. Myers ER, McCrory DC, Nanda K, Bastian L, Matchar DB. Mathematical model for the natural history of human papillomavirus infection and cervical carcinogenesis. Am J Epidemiol 2000 Jun 15;151(12):1158-1171. [doi: 10.1093/oxfordjournals.aje.a010166] [Medline: 10905528]
6. Weinstock H, Berman S, Cates W. Sexually transmitted diseases among American youth: incidence and prevalence estimates, 2000. Perspect Sex Reprod Health 2004;36(1):6-10. [doi: 10.1363/psrh.36.6.04] [Medline: 14982671]
7. Mork J, Lie AK, Glattre E, Hallmans G, Jellum E, Koskela P, et al. Human papillomavirus infection as a risk factor for squamous-cell carcinoma of the head and neck. N Engl J Med 2001 Apr 12;344(15):1125-1131. [doi: 10.1056/NEJM200104123441503] [Medline: 11297703]
8. Watson M, Saraiya M, Ahmed F, Cardinez CJ, Reichman ME, Weir HK, et al. Using population-based cancer registry data to assess the burden of human papillomavirus-associated cancers in the United States: overview of methods. Cancer 2008 Nov 15;113(10 Suppl):2841-2854 [FREE Full text] [doi: 10.1002/cncr.23758] [Medline: 18980203]
9. Viens LJ, Henley SJ, Watson M, Markowitz LE, Thomas CC, Thompson TD, et al. Human Papillomavirus-Associated Cancers - United States, 2008-2012. MMWR Morb Mortal Wkly Rep 2016 Jul 08;65(26):661-666 [FREE Full text] [doi: 10.15585/mmwr.mm6526a1] [Medline: 27387669]
10. HPV vaccine schedule and dosing. Center for Disease Control and Prevention. URL: https://www.cdc.gov/hpv/hcp/schedules-recommendations.html [accessed 2021-06-17]
11. Garland SM, Kjaer SK, Muñoz N, Block SL, Brown DR, DiNubile MJ, et al. Impact and effectiveness of the quadrivalent human papillomavirus vaccine: a systematic review of 10 years of real-world experience. Clin Infect Dis 2016 Aug 15;63(4):519-527 [FREE Full text] [doi: 10.1093/cid/ciw354] [Medline: 27230391]
12. Villa LL, Costa RLR, Petta CA, Andrade RP, Ault KA, Giuliano AR, et al. Prophylactic quadrivalent human papillomavirus (types 6, 11, 16, and 18) L1 virus-like particle vaccine in young women: a randomised double-blind placebo-controlled multicentre phase II efficacy trial. Lancet Oncol 2005 May;6(5):271-278. [doi: 10.1016/S1470-2045(05)70101-7] [Medline: 15863374]
13. FUTURE II Study Group. Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions. N Engl J Med 2007 May 10;356(19):1915-1927. [doi: 10.1056/NEJMoa061741] [Medline: 17494925]
14. Elam-Evans LD, Yankey D, Singleton JA, Sterrett N, Markowitz LE, Williams CL, et al. National, regional, state, and selected local area vaccination coverage among adolescents aged 13-17 years - United States, 2019. MMWR Morb Mortal Wkly Rep 2020 Aug 21;69(33):1109-1116 [FREE Full text] [doi: 10.15585/mmwr.mm6933a1] [Medline: 32817598]
15. CDC recommends only two HPV shots for younger adolescents. Center for Disease Control and Prevention. 2016. URL: http://www.cdc.gov/media/releases/2016/p1020-hpv-shots.html [accessed 2021-06-17]
16. Muhwezi WW, Banura C, Turiho AK, Mirembe F. Parents' knowledge, risk perception and willingness to allow young males to receive human papillomavirus (HPV) vaccines in Uganda. PLoS One 2014;9(9):e106686 [FREE Full text] [doi: 10.1371/journal.pone.0106686] [Medline: 25203053]
17. Brewer NT, Fazekas KI. Predictors of HPV vaccine acceptability: a theory-informed, systematic review. Prev Med 2007;45(2-3):107-114. [doi: 10.1016/j.ypmed.2007.05.013] [Medline: 17628649]
18. Newman PA, Logie CH, Doukas N, Asakura K. HPV vaccine acceptability among men: a systematic review and meta-analysis. Sex Transm Infect 2013 Nov;89(7):568-574 [FREE Full text] [doi: 10.1136/sextrans-2012-050980] [Medline: 23828943]
19. Larson HJ, Wilson R, Hanley S, Parys A, Paterson P. Tracking the global spread of vaccine sentiments: the global response to Japan's suspension of its HPV vaccine recommendation. Hum Vaccin Immunother 2014;10(9):2543-2550 [FREE Full text] [doi: 10.4161/21645515.2014.969618] [Medline: 25483472]
20. Dempsey AF, Zimet GD, Davis RL, Koutsky L. Factors that are associated with parental acceptance of human papillomavirus vaccines: a randomized intervention study of written information about HPV. Pediatrics 2006 May;117(5):1486-1493. [doi: 10.1542/peds.2005-1381] [Medline: 16651301]
21. Jain N, Euler GL, Shefer A, Lu P, Yankey D, Markowitz L. Human papillomavirus (HPV) awareness and vaccination initiation among women in the United States, National Immunization Survey-Adult 2007. Prev Med 2009 May;48(5):426-431. [doi: 10.1016/j.ypmed.2008.11.010] [Medline: 19100762]

22. Pask EB, Rawlins ST. men's intentions to engage in behaviors to protect against human papillomavirus (HPV): testing the risk perception attitude framework. Health Commun 2016;31(2):139-149. [doi: 10.1080/10410236.2014.940670] [Medline: 26098812]
23. Jolley D, Douglas KM. The effects of anti-vaccine conspiracy theories on vaccination intentions. PLoS One 2014;9(2):e89177 [FREE Full text] [doi: 10.1371/journal.pone.0089177] [Medline: 24586574]
24. Nan X, Madden K. HPV vaccine information in the blogosphere: how positive and negative blogs influence vaccine-related risk perceptions, attitudes, and behavioral intentions. Health Commun 2012 Nov;27(8):829-836. [doi: 10.1080/10410236.2012.661348] [Medline: 22452582]
25. Zimet GD, Rosberger Z, Fisher WA, Perez S, Stupiansky NW. Beliefs, behaviors and HPV vaccine: correcting the myths and the misinformation. Prev Med 2013 Nov;57(5):414-418 [FREE Full text] [doi: 10.1016/j.ypmed.2013.05.013] [Medline: 23732252]
26. Brewer NT, Chapman GB, Gibbons FX, Gerrard M, McCaul KD, Weinstein ND. Meta-analysis of the relationship between risk perception and health behavior: the example of vaccination. Health Psychol 2007 Mar;26(2):136-145. [doi: 10.1037/0278-6133.26.2.136] [Medline: 17385964]
27. Mayhew A, Mullins TLK, Ding L, Rosenthal SL, Zimet GD, Morrow C, et al. Risk perceptions and subsequent sexual behaviors after HPV vaccination in adolescents. Pediatrics 2014 Mar;133(3):404-411 [FREE Full text] [doi: 10.1542/peds.2013-2822] [Medline: 24488747]
28. van der Pligt J. Risk perception and self-protective behavior. European Psychologist 1996 Jan;1(1):34-43. [doi: 10.1027/1016-9040.1.1.34]
29. Weinstein ND, Kwitel A, McCaul KD, Magnan RE, Gerrard M, Gibbons FX. Risk perceptions: Assessment and relationship to influenza vaccination. Health Psychology 2007;26(2):146-151. [doi: 10.1037/0278-6133.26.2.146]
30. Betsch C, Renkewitz F, Betsch T, Ulshöfer C. The influence of vaccine-critical websites on perceiving vaccination risks. J Health Psychol 2010 Apr;15(3):446-455. [doi: 10.1177/1359105309353647] [Medline: 20348365]
31. Reagan-Steiner S, Yankey D, Jeyarajah J, Elam-Evans LD, Singleton JA, Curtis CR, et al. National, regional, state, and selected local area vaccination coverage among adolescents aged 13-17 years--United States, 2014. MMWR Morb Mortal Wkly Rep 2015 Jul 31;64(29):784-792 [FREE Full text] [doi: 10.15585/mmwr.mm6429a3] [Medline: 26225476]
32. Stein RA. The golden age of anti-vaccine conspiracies. Germs 2017 Dec;7(4):168-170 [FREE Full text] [doi: 10.18683/germs.2017.1122] [Medline: 29264353]
33. Covolo L, Ceretti E, Passeri C, Boletti M, Gelatti U. What arguments on vaccinations run through YouTube videos in Italy? A content analysis. Hum Vaccin Immunother 2017 Jul 03;13(7):1693-1699 [FREE Full text] [doi: 10.1080/21645515.2017.1306159] [Medline: 28362544]
34. Zhang C, Gotsis M, Jordan-Marsh M. Social media microblogs as an HPV vaccination forum. Hum Vaccin Immunother 2013 Nov;9(11):2483-2489 [FREE Full text] [doi: 10.4161/hv.25599] [Medline: 23842072]
35. Kang GJ, Ewing-Nelson SR, Mackey L, Schlitt JT, Marathe A, Abbas KM, et al. Semantic network analysis of vaccine sentiment in online social media. Vaccine 2017 Jun 22;35(29):3621-3638 [FREE Full text] [doi: 10.1016/j.vaccine.2017.05.052] [Medline: 28554500]
36. Orr D, Baram-Tsabari A, Landsman K. Social media as a platform for health-related public debates and discussions: the Polio vaccine on Facebook. Isr J Health Policy Res 2016;5:34 [FREE Full text] [doi: 10.1186/s13584-016-0093-4] [Medline: 27843544]
37. Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. PLoS Comput Biol 2011 Oct;7(10):e1002199 [FREE Full text] [doi: 10.1371/journal.pcbi.1002199] [Medline: 22022249]
38. Aquino F, Donzelli G, De Franco E, Privitera G, Lopalco PL, Carducci A. The web and public confidence in MMR vaccination in Italy. Vaccine 2017 Aug 16;35(35 Pt B):4494-4498. [doi: 10.1016/j.vaccine.2017.07.029] [Medline: 28736200]
39. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: an observational study. J Med Internet Res 2015 Jun 10;17(6):e144 [FREE Full text] [doi: 10.2196/jmir.4343] [Medline: 26063290]
40. Keim-Malpass J, Mitchell EM, Sun E, Kennedy C. Using twitter to understand public perceptions regarding the #HPV vaccine: opportunities for public health nurses to engage in social marketing. Public Health Nurs 2017 Mar 06;34(4):316-323. [doi: 10.1111/phn.12318]
41. Surian D, Nguyen DQ, Kennedy G, Johnson M, Coiera E, Dunn AG. Characterizing twitter discussions about HPV vaccines using topic modeling and community detection. J Med Internet Res 2016;18(8):e232 [FREE Full text] [doi: 10.2196/jmir.6045] [Medline: 27573910]
42. Massey PM, Leader A, Yom-Tov E, Budenz A, Fisher K, Klassen AC. Applying multiple data collection tools to quantify human papillomavirus vaccine communication on twitter. J Med Internet Res 2016 Dec 05;18(12):e318 [FREE Full text] [doi: 10.2196/jmir.6670] [Medline: 27919863]
43. Bahk CY, Cumming M, Paushter L, Madoff LC, Thomson A, Brownstein JS. Publicly available online tool facilitates real-time monitoring of vaccine conversations and sentiments. Health Aff (Millwood) 2016 Feb;35(2):341-347. [doi: 10.1377/hlthaff.2015.1092] [Medline: 26858390]

44. Du J, Xu J, Song H, Tao C. Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with Twitter data. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):69 [FREE Full text] [doi: 10.1186/s12911-017-0469-6] [Medline: 28699569]
45. Bessi A, Coletto M, Davidescu GA, Scala A, Caldarelli G, Quattrociocchi W. Science vs conspiracy: collective narratives in the age of misinformation. PLoS One 2015 Feb 23;10(2):e0118093 [FREE Full text] [doi: 10.1371/journal.pone.0118093] [Medline: 25706981]
46. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
47. Acemoglu D, Ozdaglar A, ParandehGheibi A. Spread of misinformation in social networks. arXiv. 2009. URL: http://arxiv.org/abs/0906.5007 [accessed 2021-06-17]
48. Lewandowsky S, Ecker UK, Cook J. Beyond misinformation: understanding and coping with the "post-truth" era. Journal of Applied Research in Memory and Cognition 2017 Dec;6(4):353-369. [doi: 10.1016/j.jarmac.2017.07.008]
49. Bessi A, Caldarelli G, Del Vicario M, Scala A, Quattrociocchi W. Social determinants of content selection in the age of (mis)information. 2014 Presented at: International Conference on Social Informatics; November, 2014; Barcelona, Spain p. 259-268 URL: http://arxiv.org/abs/1409.2651 [doi: 10.1007/978-3-319-13734-6_18]
50. Allcott H, Gentzkow M. Social media and fake news in the 2016 election. National Bureau of Economic Research. 2017. URL: http://www.nber.org/papers/w23089 [accessed 2021-06-17]
51. Dechêne A, Stahl C, Hansen J, Wänke M. The truth about the truth: a meta-analytic review of the truth effect. Pers Soc Psychol Rev 2010 May;14(2):238-257. [doi: 10.1177/1088868309352251] [Medline: 20023210]
52. Unkelbach C, Rom SC. A referential theory of the repetition-induced truth effect. Cognition 2017 Mar;160:110-126. [doi: 10.1016/j.cognition.2016.12.016]
53. Schwartz M. Repetition and rated truth value of statements. The American Journal of Psychology 1982;95(3):393-407. [doi: 10.2307/1422132]
54. Habel MA, Liddon N, Stryker JE. The HPV vaccine: a content analysis of online news stories. J Womens Health (Larchmt) 2009 Mar;18(3):401-407. [doi: 10.1089/jwh.2008.0920] [Medline: 19281323]
55. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002;287(20):2691-2700. [Medline: 12020305]
56. Cline RJ, Haynes KM. Consumer health information seeking on the Internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692 [FREE Full text] [Medline: 11780707]
57. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
58. Dhoju S, Main URM, Ashad KM, Hassan N. Differences in health news from reliable and unreliable media. 2019 Presented at: Companion Proceedings of The 2019 World Wide Web Conference; May 13-17 2019; San Francisco p. 981-987. [doi: 10.1145/3308560.3316741]
59. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. J Med Internet Res 2021 Jan 20;23(1):e17187 [FREE Full text] [doi: 10.2196/17187] [Medline: 33470931]
60. Tomeny TS, Vargo CJ, El-Toukhy S. Geographic and demographic correlates of autism-related anti-vaccine beliefs on Twitter, 2009-15. Soc Sci Med 2017 Oct;191:168-175 [FREE Full text] [doi: 10.1016/j.socscimed.2017.08.041] [Medline: 28926775]
61. Huang X, Smith M, Paul M, Ryzhkov D, Quinn S, Broniatowski D, et al. Examining patterns of influenza vaccination in social media. 2017 Presented at: AAAI Workshops; Feb 4-5, 2017; San Francisco.
62. Hou R, Pérez-Rosas V, Loeb S, Mihalcea R. Towards automatic detection of misinformation in online medical videos. 2019 Presented at: International Conference on Multimodal Interaction; October 2019; Suzhou, China p. 235-243 URL: http://arxiv.org/abs/1909.01543 [doi: 10.1145/3340555.3353763]
63. Wang Z, Yin Z, Argyris Y. Detecting medical misinformation on social media using multimodal deep learning. IEEE J. Biomed. Health Inform 2021 Jun;25(6):2193-2203 [FREE Full text] [doi: 10.1109/jbhi.2020.3037027]
64. Cheng L, Guo R, Shu K, Liu H. Towards causal understanding of fake news dissemination. arXiv. 2020. URL: https://arxiv.org/abs/2010.10580 [accessed 2021-06-17]
65. Ruchansky N, Seo S, Liu Y. CSI: A hybrid deep model for fake news detection. 2017 Presented at: ACM Conference on Information and Knowledge Management; Nov 6-10, 2017; New York p. 797-806. [doi: 10.1145/3132847.3132877]
66. Shu K, Cui L, Wang S, Lee D, Liu H. dEFEND: Explainable Fake News Detection. 2019 Presented at: 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; Aug 4-8, 2019; New York p. 395-405. [doi: 10.1145/3292500.3330935]
67. Hunt R. Percent agreement, Pearson's correlation, and kappa as measures of inter-examiner reliability. J Dent Res 1986 Feb;65(2):128-130. [doi: 10.1177/00220345860650020701] [Medline: 3455967]
68. Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching word vectors with subword information. TACL 2017 Dec;5:135-146. [doi: 10.1162/tacl_a_00051]

69. Joulin A, Grave E, Bojanowski P, Mikolov T. Bag of tricks for efficient text classification. 2017 Presented at: 15th Conference of the European Chapter of the Association for Computational Linguistics; Apr 3-7, 2017; Valencia, Spain p. 427-431. [doi: 10.18653/v1/e17-2068]
70. Kim Y. Convolutional neural networks for sentence classification. : Association for Computational Linguistics; 2014 Presented at: The 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); Oct 24-29, 2014; Doha, Qatar p. 1746-1751. [doi: 10.3115/v1/d14-1181]
71. Liaw R, Liang E, Nishihara R, Moritz P, Gonzalez J, Stoica I. Tune: a research platform for distributed model selection and training. arXiv. 2018. URL: https://arxiv.org/abs/1807.05118 [accessed 2021-06-17]
72. Kayesh H, Islam M, Wang J. Event causality detection in tweets by context word extension and neural networks. 2019 Presented at: 20th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT); Dec 5-7, 2019; Gold Coast, Australia p. 352-357. [doi: 10.1109/pdcat46702.2019.00070]
73. Kong L, Schneider N, Swayamdipta S, Bhatia A, Dyer C, Smith N. A dependency parser for tweets. 2014 Presented at: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP); Oct 25-29, 2014; Doha, Qatar p. 1001-1012. [doi: 10.3115/v1/d14-1108]
74. Zhai C, Lafferty J. A study of smoothing methods for language models applied to ad hoc information retrieval. 2017 Presented at: ACM SIGIR Forum; July 2017; New York p. 268-276. [doi: 10.1145/3130348.3130377]
75. Bouma G. Normalized (pointwise) mutual information in collocation extraction. 2009 Presented at: Proceedings of GSCL; Sep 30, 2009; Potsdam, Germany p. 31-40.
76. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al. HuggingFace's transformers: state-of-the-art natural language processing. arXiv. 2019. URL: https://arxiv.org/abs/1910.03771 [accessed 2021-06-17]
77. Devlin J, Chang M, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv. 2018. URL: https://arxiv.org/abs/1810.04805 [accessed 2021-06-17]
78. Schubert E, Sander J, Ester M, Kriegel H, Xu X. DBSCAN revisited, revisited: why and how you should (Still) use DBSCAN. ACM Trans. Database Syst 2017 Aug 24;42(3):1-21. [doi: 10.1145/3068335]
79. Ankerst M, Breunig M, Kriegel H, Sander J. OPTICS: ordering points to identify the clustering structure. SIGMOD Rec 1999 Jun;28(2):49-60. [doi: 10.1145/304182.304187]
80. McInnes L, Healy J, Astels S. hdbscan: Hierarchical density based clustering. JOSS 2017 Mar;2(11):205. [doi: 10.21105/joss.00205]
81. Chin J, Chin C, Caskey R, Liu B, Schwartz A. Building a patient-driven human papillomavirus (HPV) vaccine risk lexicon. 2018 Presented at: 40th Annual Meeting of the Society for Medical Decision Making; Oct 13-17, 2018; Montreal, Canada. [doi: 10.5005/jp/books/10027_10]
82. Smith C, Stavri P. Consumer health vocabulary. In: Lewis D, Eysenbach G, Kukafka R, Stavri P, Jimison H, editors. Consumer Health Informatics: Informing Consumers and Improving Health Care. New York, NY: Springer; 2005:122-128.
83. Zeng QT, Tse T. Exploring and developing consumer health vocabularies. J Am Med Inform Assoc 2006;13(1):24-29 [FREE Full text] [doi: 10.1197/jamia.M1761] [Medline: 16221948]
84. Xu Z. I don't understand you but I trust you: using computer-aided text analysis to examine medical terminology use and engagement of vaccine online articles. Journal of Communication in Healthcare 2020 May 21;14(1):61-67. [doi: 10.1080/17538068.2020.1755137]
85. Meng J, Peng W, Tan P, Liu W, Cheng Y, Bae A. Diffusion size and structural virality: The effects of message and network features on spreading health information on twitter. Comput Human Behav 2018 Dec;89:111-120 [FREE Full text] [doi: 10.1016/j.chb.2018.07.039] [Medline: 32288177]
86. Gerend MA, Shepherd JE. Using message framing to promote acceptance of the human papillomavirus vaccine. Health Psychol 2007 Nov;26(6):745-752. [doi: 10.1037/0278-6133.26.6.745] [Medline: 18020847]
87. O'Keefe DJ, Nan X. The relative persuasiveness of gain- and loss-framed messages for promoting vaccination: a meta-analytic review. Health Commun 2012;27(8):776-783. [doi: 10.1080/10410236.2011.640974] [Medline: 22292904]
88. Weissenbacher D, Sarker A, Paul M, Gonzalez-Hernandez G. Overview of the third social media mining for health (SMM4H) shared tasks. In: Association for Computational Linguistics. 2018 Presented at: The 2018 EMNLP workshop SMM4H: the 3rd social media mining for health applications workshop & shared task; Oct 2018; Brussels, Belgium. [doi: 10.18653/v1/w18-5904]
89. Ghenai A, Mejova Y. Fake Cures. In: Proc. ACM Hum.-Comput. Interact. 2018 Nov Presented at: Proceedings of the ACM on Human-Computer Interaction; 2018; New York p. 1-20. [doi: 10.1145/3274327]
90. Jin Z, Cao J, Zhang Y, Luo J. News verification by exploiting conflicting social viewpoints in microblogs. 2016 Presented at: Proceedings of the AAAI Conference on Artificial Intelligence; Feb 12-17, 2016; Pheonix, Arizona.
91. Wang Y, Yang W, Ma F, Xu J, Zhong B, Deng Q, et al. Weak supervision for fake news detection via reinforcement learning. In: AAAI. 2020 Apr 03 Presented at: Proceedings of the AAAI Conference on Artificial Intelligence 2020; Feb 7-12; New York p. 516-523. [doi: 10.1609/aaai.v34i01.5389]
92. Lu Y, Li C. GCAN: Graph-aware Co-Attention Networks for Explainable Fake News Detection on Social Media. 2020 Presented at: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics; Jul 6-8, 2020; Online p. 505-514. [doi: 10.18653/v1/2020.acl-main.48]

93. Joshi A, Dai X, Karimi S, Sparks R, Paris C, MacIntyre C. Shot or not: Comparison of NLP approaches for vaccination behaviour detection. 2018 Presented at: Proceedings of the 2018 EMNLP Workshop SMM4H: The 3rd Social Media Mining for Health Applications Workshop & Shared Task; Oct 31, 2018; Brussels, Belgium. [doi: 10.18653/v1/w18-5911]
94. Yin W, Kann K, Yu M, Schütze H. Comparative study of CNN and RNN for natural language processing. arXiv. 2017. URL: https://arxiv.org/abs/1702.01923 [accessed 2021-06-17]
95. Joskowicz L, Ksiezyck T, Grishman R. Deep domain models for discourse analysis. 1989 Presented at: The Annual AI Systems in Government Conference; Mar 27-31, 1989; Washington, DC p. 195. [doi: 10.1109/aisig.1989.47325]
96. Kaplan RM, Berry-Rogghe G. Knowledge-based acquisition of causal relationships in text. Knowledge Acquisition 1991 Sep;3(3):317-337. [doi: 10.1016/1042-8143(91)90009-c]
97. Girju R. Automatic detection of causal relations for question answering. 2003 Presented at: The ACL Workshop on Multilingual Summarization and Question Answering; July 2003; Sapporo, Japan p. 76-83. [doi: 10.3115/1119312.1119322]
98. Hassanzadeh O, Bhattacharjya D, Feblowitz M, Srinivas K, Perrone M, Sohrabi S, et al. Answering binary causal questions through large-scale text mining: an evaluation using cause-effect pairs from human experts. Proceedings of the 28 International Joint Conference on Artificial Intelligence (IJCAI-19) 2019:5003-5009. [doi: 10.24963/ijcai.2019/695]
99. Guo C, Pleiss G, Sun Y, Weinberger K. On calibration of modern neural networks. 2017 Presented at: 34th International Conference on Machine Learning; Aug 6-11, 2017; Sydney, Australia p. 1321-1330.
100. Wu L, Morstatter F, Carley KM, Liu H. Misinformation in social media: Definition, manipulation, and detection. SIGKDD Explor. Newsl 2019 Nov 26;21(2):80-90. [doi: 10.1145/3373464.3373475]
101. Acker A. Data craft: the manipulation of social media metadata. Data & Society Research Institute. 2018. URL: https://datasociety.net/wp-content/uploads/2018/11/DS_Data_Craft_Manipulation_of_Social_Media_Metadata.pdf [accessed 2021-06-16]

Abbreviations

BiLSTM: bidirectional long short-term memory
CNN: convolutional neural network
DBSCAN: density-based spatial clustering of applications with noise
HDBSCAN: hierarchical density-based spatial clustering of applications with noise
HPV: human papillomavirus
HPVVR: The HPV-Vaccine Risk Lexicon
NPMI: normalized pointwise mutual information
OPTICS: ordering points to identify the clustering structure
PMI: pointwise mutual information

Edited by G Eysenbach; submitted 14.05.21; peer-reviewed by R R, D Huang; comments to author 08.06.21; accepted 04.08.21; published 09.09.21.

Please cite as:
Tomaszewski T, Morales A, Lourentzou I, Caskey R, Liu B, Schwartz A, Chin J
Identifying False Human Papillomavirus (HPV) Vaccine Information and Corresponding Risk Perceptions From Twitter: Advanced Predictive Models
J Med Internet Res 2021;23(9):e30451
URL: https://www.jmir.org/2021/9/e30451
doi:10.2196/30451
PMID:34499043

Original Paper

Deep Learning for Identification of Alcohol-Related Content on Social Media (Reddit and Twitter): Exploratory Analysis of Alcohol-Related Outcomes

Benjamin Joseph Ricard[1], MSc; Saeed Hassanpour[1,2,3], PhD

[1]Department of Biomedical Data Science, Dartmouth College, Lebanon, NH, United States

[2]Department of Epidemiology, Dartmouth College, Hanover, NH, United States

[3]Department of Computer Science, Dartmouth College, Hanover, NH, United States

Corresponding Author:
Saeed Hassanpour, PhD
Department of Biomedical Data Science
Dartmouth College
Williamson Translational Research Building
One Medical Center Drive HB 7261
Lebanon, NH, 03756
United States
Phone: 1 603 650 1983
Email: saeed.hassanpour@dartmouth.edu

Abstract

Background: Many social media studies have explored the ability of thematic structures, such as hashtags and subreddits, to identify information related to a wide variety of mental health disorders. However, studies and models trained on specific themed communities are often difficult to apply to different social media platforms and related outcomes. A deep learning framework using thematic structures from Reddit and Twitter can have distinct advantages for studying alcohol abuse, particularly among the youth in the United States.

Objective: This study proposes a new deep learning pipeline that uses thematic structures to identify alcohol-related content across different platforms. We apply our method on Twitter to determine the association of the prevalence of alcohol-related tweets with alcohol-related outcomes reported from the National Institute of Alcoholism and Alcohol Abuse, Centers for Disease Control Behavioral Risk Factor Surveillance System, county health rankings, and the National Industry Classification System.

Methods: The Bidirectional Encoder Representations From Transformers neural network learned to classify 1,302,524 Reddit posts as either alcohol-related or control subreddits. The trained model identified 24 alcohol-related hashtags from an unlabeled data set of 843,769 random tweets. Querying alcohol-related hashtags identified 25,558,846 alcohol-related tweets, including 790,544 location-specific (geotagged) tweets. We calculated the correlation between the prevalence of alcohol-related tweets and alcohol-related outcomes, controlling for confounding effects of age, sex, income, education, and self-reported race, as recorded by the 2013-2018 American Community Survey.

Results: Significant associations were observed: between alcohol-hashtagged tweets and alcohol consumption (P=.01) and heavy drinking (P=.005) but not binge drinking (P=.37), self-reported at the metropolitan-micropolitan statistical area level; between alcohol-hashtagged tweets and self-reported excessive drinking behavior (P=.03) but not motor vehicle fatalities involving alcohol (P=.21); between alcohol-hashtagged tweets and the number of breweries (P<.001), wineries (P<.001), and beer, wine, and liquor stores (P<.001) but not drinking places (P=.23), per capita at the US county and county-equivalent level; and between alcohol-hashtagged tweets and all gallons of ethanol consumed (P<.001), as well as ethanol consumed from wine (P<.001) and liquor (P=.01) sources but not beer (P=.63), at the US state level.

Conclusions: Here, we present a novel natural language processing pipeline developed using Reddit's alcohol-related subreddits that identify highly specific alcohol-related Twitter hashtags. The prevalence of identified hashtags contains interpretable information about alcohol consumption at both coarse (eg, US state) and fine-grained (eg, metropolitan-micropolitan statistical area level and county) geographical designations. This approach can expand research and deep learning interventions on alcohol abuse and other behavioral health outcomes.

(J Med Internet Res 2021;23(9):e27314) doi:10.2196/27314

KEYWORDS

social media; natural language processing; alcohol abuse; machine learning

Introduction

Background

Alcohol-related causes are the third leading preventable cause of death in the United States, and alcohol abuse contributes to many adverse health outcomes, particularly on the developing brain [1-4]. The rise of alcohol-related content on Twitter is alarming, with over half of young adults participating in a study [5] posting alcohol-related content. Social media use and alcohol consumption are common behaviors; the prevalence rates of Twitter, Reddit, and annual alcohol use for US adults are 22%, 11%, and 70%, respectively [6,7]. Internet- and social media–based interventions are scalable and efficient approaches for developing practical tools for treating and monitoring alcohol abuse, especially for at-risk adolescents and young adults [8-14]. However, identifying high-risk areas for efficient and helpful monitoring along with population-level interventions remains a difficult task, in part because of survey bias [15-17].

Text-based *hashtags* are common among many popular social media platforms such as Twitter, Instagram, and TikTok. Individuals use hashtags to categorize, label, organize, and discover posts and content [18]. Previous studies have indicated that study-specific hashtags are useful for mental health research [19]. For example, sexual abuse and harassment (#MeToo), breast cancer (#breastcancer), HIV (#HIV), miscarriages (#ihadamiscarriage), tobacco use (#Vapelife), and viral pandemics (#COVID-19) are some of the many important health outcomes that have been previously studied using hashtags on Twitter [20-28]. Other social media platforms such as Reddit contain specific *themed* communities where interested users discuss a particular topic. In contrast to hashtags, themed communities on websites such as Reddit represent posts related to exactly 1 topic of interest. Like hashtags, these communities, such as *r/cripplingalcoholism*, *r/depression*, or *r/opiates* Reddit subreddits and *HIV* Baidu Tieba bar, contain information that can target and understand behavioral health and disease [29-33]. In addition to hashtags and subreddits, some social media platforms allow for *geotagging* or sharing a user's geographical latitude and longitude coordinates in a post. Geotags have been used in social media research to identify geographically relevant information from social media data [34-36].

Previous Work

Although prior studies have identified specific hashtags or themed communities for studying behavioral health outcomes, many insights are platform-specific. Although helpful information regarding a behavior of interest or themed community may be available on one platform, there may not be such knowledge available on a different platform. Many previous methods examining alcohol content on social media use data from a single platform [5,37-42]. Single-platform analyses may limit discoveries and interventions to only a fraction of the population at risk. There is a growing need for behavioral health researchers working with social media data to incorporate analyses from many sources [43,44]. Although some studies have examined alcohol content on multiple platforms, many methods need survey data from known active users from each source or additional manual annotation [45-47]. The ability and insights gained from using deep learning methods to learn from a large number of posts from specific communities (ie, Reddit subreddits) to predict alcohol-related content on a different platform (ie, Twitter) remain unclear.

Many previous studies that identified alcohol-related language on social media platforms relied on training on extrinsic labels, such as survey responses. Reliance on self-report data is problematic as alcohol consumption is subject to bias, particularly among the youth [15,16,48]. In addition, approaches that use an outcome of interest to both train and evaluate a model (eg, identifying and evaluating alcohol-related hashtags or keywords based on enrichment in regions with higher self-reported alcohol content) may not be generalizable to other related outcomes [49].

Other approaches for studying alcohol content involve identifying a sample as being alcohol-related based on the identification of keywords. Keyword approaches have distinct benefits, such as interpretability. However, identifying text from keywords may rely on standard and predefined terms (eg, searching *drunk*), training on self-report data, or manual review [37,42,49-51]. Classification of social media posts based on previously defined keywords or vector representations (eg, Word2vec) is not as useful when the average length of sequences is small and has out-of-training vocabulary [52-54]. Training on nonspecific platform information alone may fail to capture relevant keywords, especially for rarer outcomes not prominent in the heterogeneity of random and unlabeled social media chatter [55]. In addition, predefined keywords or word vectors may fail to capture slang or the different language structures between Reddit and Twitter [56].

One recent contribution in natural language processing (NLP) is the Bidirectional Encoder Representations From Transformers (BERT) neural network, which has demonstrated superior performance on a wide variety of social media NLP tasks [57-61]. BERT focuses on learning by analyzing sentences with randomly masked words. This masked language model deconstructs larger strings into smaller tokens and is ideal for dealing with hashtags and other platform-unique token structures [57]. Before developing BERT, previous models, such as long-short–term memory networks, logistic regression, Word2vec similarity, and latent Dirichlet allocation, were not well suited to process unknown words and hashtag structures. For example, some previous NLP studies on social media either removed hashtags, represented them as universal tokens, or removed # from strings, with no importance given to hashtags (eg, *#ilovebeer* represented as " " (space), *HASHTAG*, or *ilovebeer*, respectively) [62-64]. In contrast, using hashtags and themed communities as explicit labels in a deep learning

architecture allows for identifying relevant, platform-specific hashtags that can identify posts that indicate the behavior of interest. In addition, the use of these structures adds a layer of interpretability to our trained neural networks, which are commonly criticized as noninterpretable black boxes [65].

Other previous social media text mining methods implementing deep learning often involve training platform-specific models. One issue with this approach is that each platform's training models require an extensive amount of usually labeled data from that platform [66-72]. In addition, although deep learning models have been successful at many tasks, training platform-specific deep networks such as BERT (containing >100 million parameters) is extremely energy- and cost-intensive, and CO_2 emissions from training BERT models have raised concerns about their environmental impact [73]. Optimal methods for translating information from previously trained social media deep learning models to discern insights from separate social media platforms remain a relatively unexplored research area.

The Goal of This Study

We aim to examine the effectiveness of using thematic structures in a deep learning framework to identify alcohol-related behaviors across different social media platforms. First, we trained on Reddit subreddits to identify alcohol-related targets on another social media platform (Twitter) with a different thematic structure (hashtags). Next, we determined whether the hashtags predicted by the model correlate to known alcohol-related outcomes, including self-reported drinking status, alcohol outlet density, and estimated gallons of ethanol consumed, after controlling for confounding effects of age, sex, income, education, and self-reported race. We show that these data-driven hashtags contain interpretable information about alcohol consumption in the United States. Finally, we present validated and queryable hashtags from our model that behavioral health researchers can use as a starting point for the identification of alcohol-related content on Twitter, Reddit, and other social media platforms.

Methods

Overview of the NLP Pipeline

This study fine-tuned a BERT neural network as a binary classifier to predict Reddit post titles as belonging to either alcohol-related communities or a random subreddit. Next, we applied the Reddit-trained network to a smaller set of random, unlabeled Twitter posts to identify 24 hashtags that were significantly associated with alcohol content. We identified 25,558,846 tweets that contained at least one alcohol-related hashtag for the period between 2010 and 2019. A total of 1,412,041 alcohol-related tweets included latitude and longitude data from *geotagging*. The locations of 790,544 geotagged tweets from 2929 US counties and county equivalents were identified using data from the 2017 US Census Shapefiles database [74,75]. Finally, we examined the relationship between the prevalence of alcohol-related tweets per population and various outcome measures related to alcohol consumption, including self-reported alcohol consumption and alcohol outlet density. Figure 1 demonstrates an overview of our NLP pipeline. Figure 2 illustrates the choropleth of population-normalized alcohol-hashtagged tweets for US states and Washington, DC.

Figure 1. Overview of the methodological pipeline. A bidirectional encoder representation from transformers model trained to classify posts as either 18 alcohol-related or control subreddits. The bidirectional encoder representations from transformers model was applied to a set of tweets containing at least one hashtag. The prediction results were analyzed to find 24 significantly enriched hashtags as positive predictions (ie, prediction probability ≥0.5). Tweets posted between 2010 and 2020 with an alcohol-related hashtag were collected and filtered on geotagged location. BERT: Bidirectional Encoder Representations From Transformers.

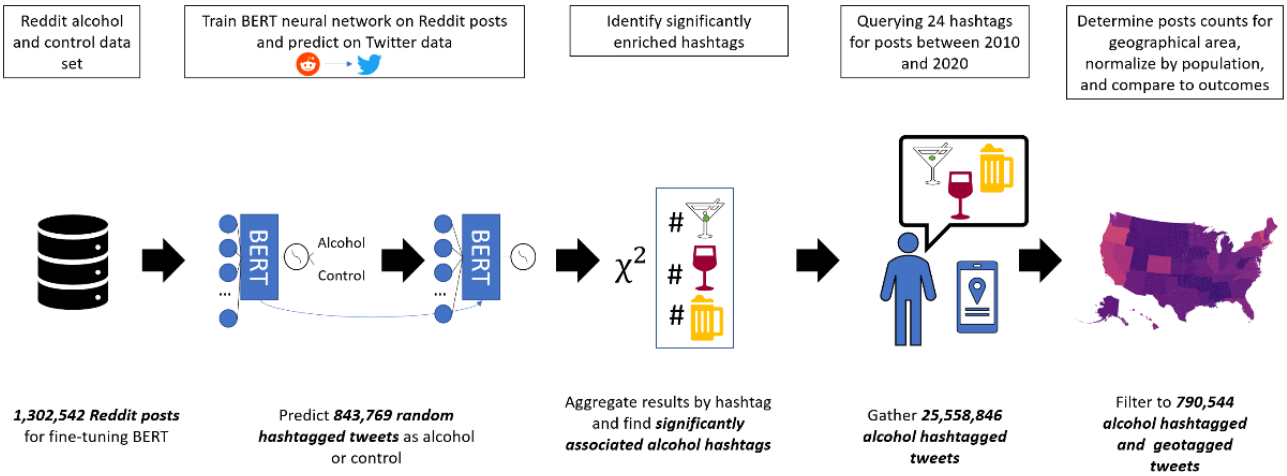
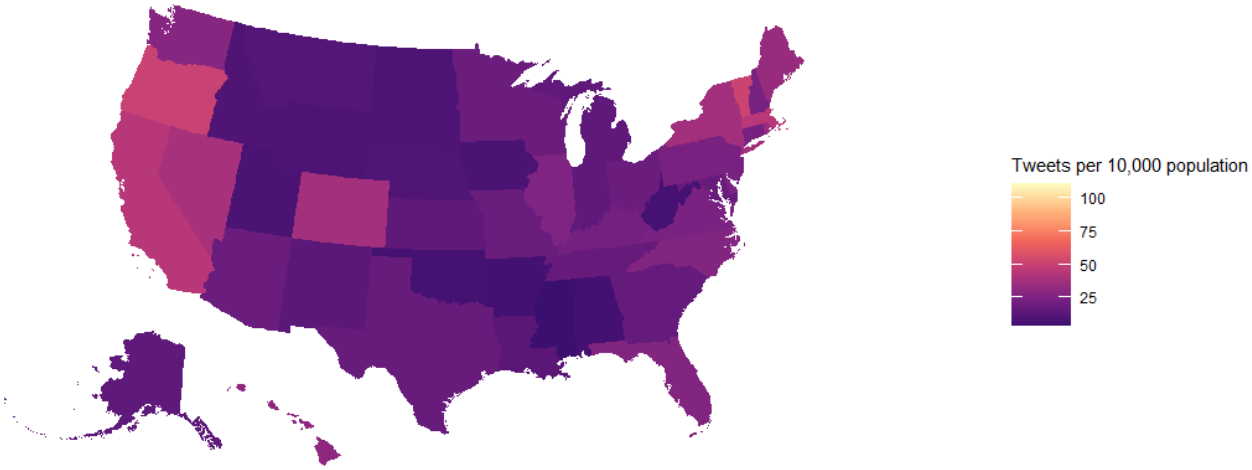


Figure 2. Choropleth of the US state and Washington, DC tweets with alcohol-related hashtags per 10,000 persons.



Reddit Data Set and BERT Training

A large amount of alcohol-related training data were extracted from Reddit subreddits with the pushshift application programming interface (API) previously used in social media research [63,76]. On a subreddit, community moderators create description texts that contain links to other, usually related, subreddits. Scraping the description pages for all subreddits containing at least 1000 posts for any links to *r/drunk*, one of the most popular alcohol-related subreddits, and all links from *r/drunk* to other subreddits yielded 17 alcohol-related subreddits. A total of 651,271 post titles from the following 18 subreddits were used as positive alcohol labels for model training: *r/cripplingalcoholism*, *r/vodka*, *r/oldtimehockey*, *r/alcohol*, *r/beer*, *r/bourbon*, *r/homebrewing*, *r/drinkinggames*, *r/wine*, *r/beercirclejerk*, *r/gin*, *r/scotch*, *r/liquor*, *r/showerbeer*, *r/absinthe*, *r/firewater*, *r/beercanada*, and *r/drunk*.

Negative alcohol (control) posts were gathered by querying 651,271 random posts posted in all other subreddits, excluding the 18 alcohol-related subreddits. Training 79.99% (521,016/651,271), validation 9.99% (65,127/651,271), and testing 9.99% (65,127/651,271), data sets were generated for developing and evaluating the model—a binary classifier trained for posts belonging to either alcohol-related subreddits or other random subreddits. The model fine-tuned a pretrained BERT model with 12 layers and 768 hidden units in PyTorch on an NVIDIA TITAN Xp graphics processing unit using a batch size of 64 for approximately 5 weeks [77].

Twitter Data Set and Identification of Hashtags

The Twitter API provides tweet information from 7 days before a query. Randomly selected tokens in the Twitter GLoVE word embedding dictionary and their respective hashtags (ie, a string that starts with #) were queried using the Twitter API to identify recently posted tweets containing that word or hashtag [78]. Each identified hashtag in the data set was requeried to ensure that it was monitored for at least 2 weeks. The initial *random* Twitter data set comprised 843,769 random hashtag-containing tweets posted between January 2019 and October 2019. The Reddit-trained BERT model was applied to this data set to obtain binary predictions for each tweet. A chi-square test identified 24 significant alcohol-related hashtags from posts predicted to be alcohol-positive (ie, final softmax layer prediction *P* value of ≥0.5) relative to posts predicted as negative (ie, final softmax prediction *P* value of <0.5) using a one-tailed *greater* test. We included only hashtags with 5 or more occurrences and applied the Benjamini-Hochberg algorithm for multiple hypothesis correction using a 0.05 false discovery rate, a common approach for multiple hypothesis corrections in social media data analyses [37,42,79-85]. The analysis resulted in 24 hashtags, as indicated in Textbox 1. GetOldTweets, a Python package widely used in social media research, was used to identify 25,558,846 alcohol-hashtagged tweets posted throughout 10 years (between 2010 and 2020) containing at least one significant alcohol-related hashtag [55,86,87].

Textbox 1. Alcohol-related hashtags extracted by our Reddit-trained classifier according to alcohol category.

Beer hashtags

- *craftbeer*, *beer*, *ncbeer*, *brewery*, *stout*, *beeroclock*, *beergeek*, *beerporn*, *beers*, *instabeer*, *beertime*, *beerstagram*, *beerlover*, *beersnob*

Wine hashtags

- *winetasting*, *wine*, *winelover*, *wines*, *redwine*

Liquor hashtags

- *bourbon*, *whiskey*, *whisky*

Multiple or ambiguous hashtags

- *drinklocal*, *drunktwitter*

Geographical Identification of the Prevalence of Alcohol-Related Hashtags

Next, we tested whether the knowledge of 24 significant alcohol hashtags could uncover information on alcohol-related outcomes in the United States. Alcohol-hashtagged tweets were filtered to 790,544 *geotagged* tweets containing longitude and latitude coordinate locations and mapped to metropolitan-micropolitan statistical areas (MMSAs), US county and county equivalents, and US states and Washington, DC. The total number of alcohol-hashtagged tweets in an area divided by the mean of the population estimates from the 2013-2018 American Community Survey yielded population-normalized alcohol-related hashtag prevalence.

We then tested the association between geographical prevalence of alcohol-related hashtags and alcohol outcomes. Spearman rho, a ranked nonparametric measure that is more robust to outliers than Pearson correlation, is used to report crude (nonadjusted) correlations [88]. Potential confounding variables previously studied in alcohol and social media use include race and sex distribution, median age, education, and income [37,89]. A linear regression analysis evaluated the relationship between the number of tweets per population and alcohol-related outcomes after including terms to control for confounding effects. Specific confounding variables from the 2013-2018 5-year American Community Survey report included *Percent Reporting White*, *Percent Reporting Black*, *Percent Reporting Hispanic*, *Median Income*, *Percent High School Education*, *Percent Bachelor's Degree Education*, and *Males/100 Females* [90]. All alcohol outcomes and confounding variables represented the most recent estimation of alcohol consumption and related behavior at the time of this study.

Metropolitan-Micropolitan Statistical Areas

MMSAs are US Census Bureau designations of concentrated urban centers that may be the integrated areas of multiple cities and states (eg, the single *Washington-Arlington-Alexandria, DC-VA-MD-WV MSA* contains 3 US states and Washington, DC) [74]. The Behavioral Risk Factor Surveillance System publishes reports of survey responses at selected MMSAs for the following categories [91]:

- any alcohol consumption, defined as at least one alcoholic drink in the last 30 days;
- binge drinking behavior, defined as drinking >5 drinks in 1 event for men or >4 drinks in 1 event for women;
- heavy drinking, defined as drinking >1 drink per day for women or >2 drinks per day per man;

All yearly records from 2010-2019 for each MMSA were averaged to obtain a single number for outcome measurements.

US County and County Equivalent

Primary US county outcomes were gathered from the University of Wisconsin Population Health Institute County Health Rankings and Roadmaps 2020 data, which included the estimates of excessive drinking, defined as the percentage reporting either binge or heavy drinking behavior as well as measurements of the percentage of motor vehicle fatalities that involved alcohol for the period between 2013 and 2018 [92]. In addition, data from the North American Industry Classification System provided by 2017 County Business Patterns (US Census) was used for the number of *Drinking Places (Alcohol Beverages)*; *Wineries*; *Breweries*; and *Beer, Wine, and Liquor stores* (North American Industry Classification System codes 722410, 312130, 312120, 445310, respectively) present in each county [93]. Counties were included if they contained at least one tweet and an average reported population >1000 between 2013 and 2018.

US States and Washington DC

Twitter posts containing alcohol-related hashtags were aggregated by state and compared with the National Institute on Alcohol Abuse and Alcoholism's 2018 report, *Apparent Per Capita Alcohol Consumption: National, State, and Regional Trends*. This report predicts gallons of ethanol consumption based on alcohol sales and taxation data, separated for the consumption of wine, beer, or liquor products [94]. To determine which hashtags may be useful for detecting individual preferences of alcohol consumption, we calculated the correlation between the consumption of alcohol from different sources of alcoholic drinks (beer, wine, and liquor) and the prevalence of 19 beer, 5 wine, and 3 liquor-specific hashtags, as indicated in Textbox 1.

Results

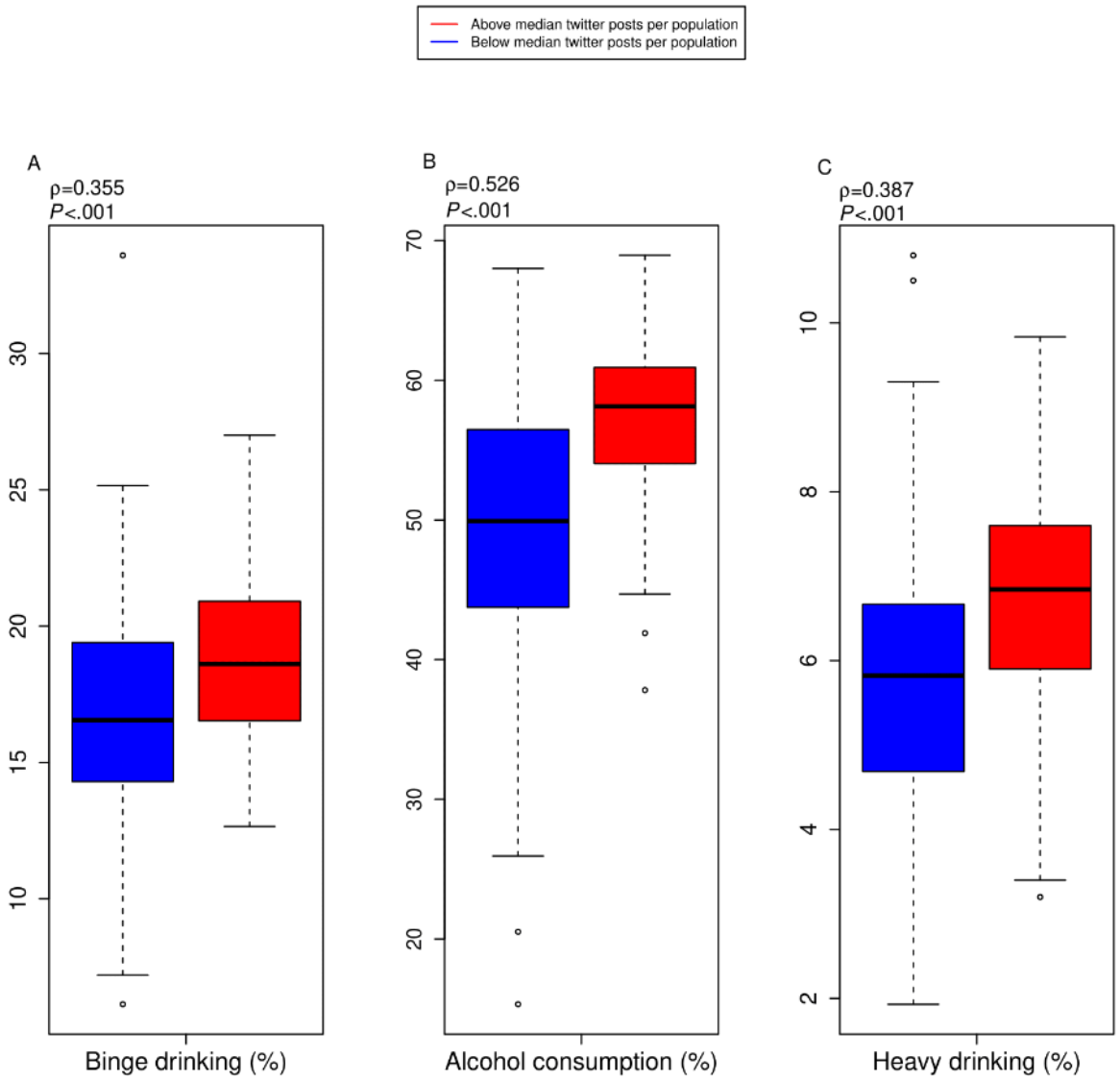
Table 1 demonstrates the results from the analysis of alcohol-related hashtags and alcohol-related outcomes. The

number of geotagged and alcohol-hashtagged tweets per population significantly correlated with many alcohol-related outcomes, including self-reported measures of individuals (1) reporting any alcohol consumption within 30 days (P<.001), (2) meeting the criteria for heavy drinking (P<.001), and (3) meeting the criteria for binge drinking (P<.001) at the MMSA level (Figure 3). However, the relationship between MMSA tweets and binge drinking level was not significant after adjusting for confounding effects (P=.37).

Table 1. Spearman correlation and linear regression results between the number of tweets per population and alcohol-related behavior and health indicators.

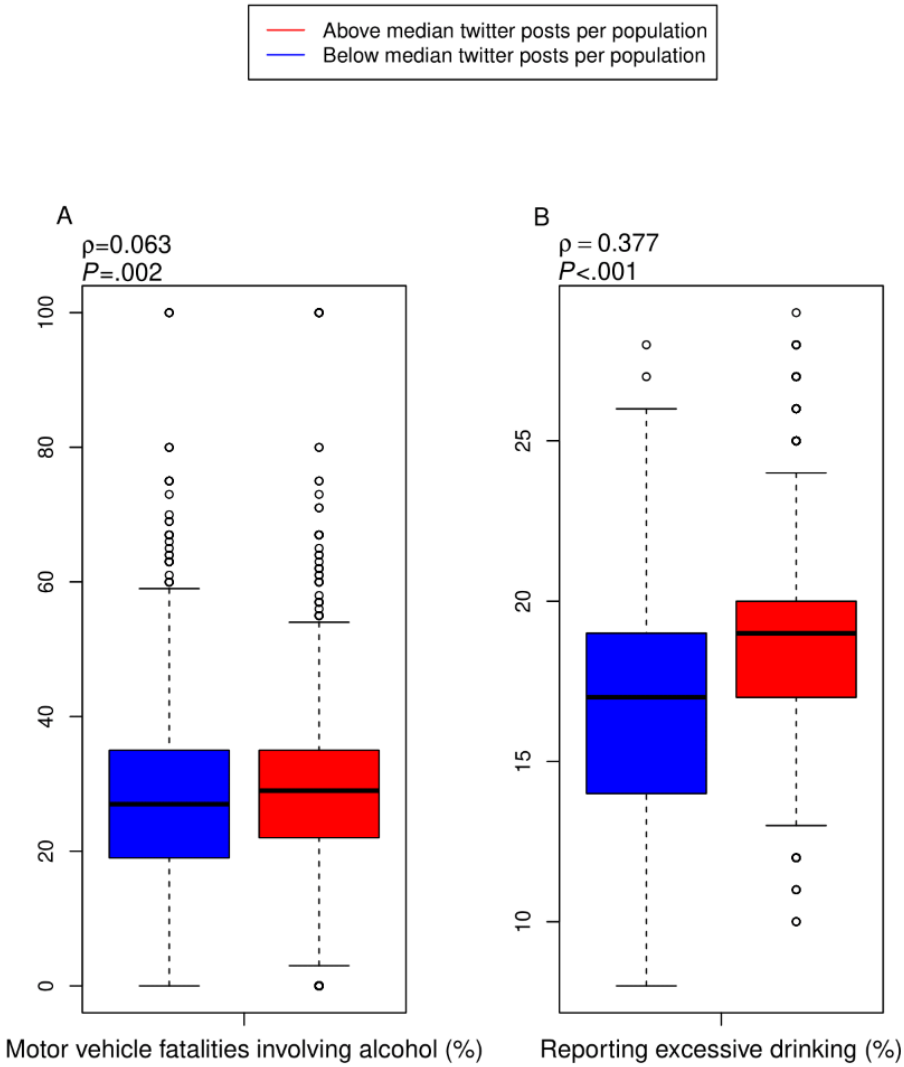
Outcome	Spearman correlation		Adjusted regression		Sample size, n
	ρ	*P* value	Coefficient β	*P* value	
Metropolitan-micropolitan statistical area					
Alcohol consumption	0.526	<.001	1038	.01	179
Binge drinking	0.355	<.001	184.0	.37	179
Heavy drinking	0.387	<.001	244.8	.005	179
County and equivalent					
Excessive drinking	0.377	<.001	32.8	.03	2641
Percentage of alcohol motor vehicle fatality	0.063	.002	110.0	.21	2641
Drinking places (alcoholic beverages) per capita	−0.177	<.001	−2.18e–03	.23	1479
Breweries per capita	0.263	<.001	1.86e–03	<.001	334
Wineries per capita	0.130	.05	2.73e–02	<.001	228
Beer, wine, and liquor stores per capita	−0.043	.11	0.0039	<.001	1444
US states and Washington, DC, all hashtags					
Wine, gallons of ethanol per capita	0.756	<.001	74.11	<.001	51
Beer, gallons of ethanol per capita	−0.050	.73	9.911	.63	51
Liquor, gallons of ethanol per capita	0.320	.01	62.54	.03	51
All sources, gallons of ethanol per capita	0.437	<.001	146.6	<.001	51
US states and Washington, DC, hashtags stratified by alcohol category					
Wine, gallons of ethanol per capita (5 hashtags)	0.754	<.001	214.6	<.001	51
Beer, gallons of ethanol per capita (19 hashtags)	−0.001	.99	16.05	.63	51
Liquor, gallons of ethanol per capita (3 hashtags)	0.140	.33	338.0	.01	51

Figure 3. Metropolitan-micropolitan statistical area correlations for alcohol-hashtagged tweets and percent self-reported alcohol consumption. (A) Number of alcohol-hashtagged tweets and self-reported alcohol consumption within 30 days (N=179); (B) number of alcohol-hashtagged tweets and self-reported binge drinking within 30 days (N=179); (C) number of alcohol-hashtagged tweets and self-reported heavy drinking within 30 days (N=179).



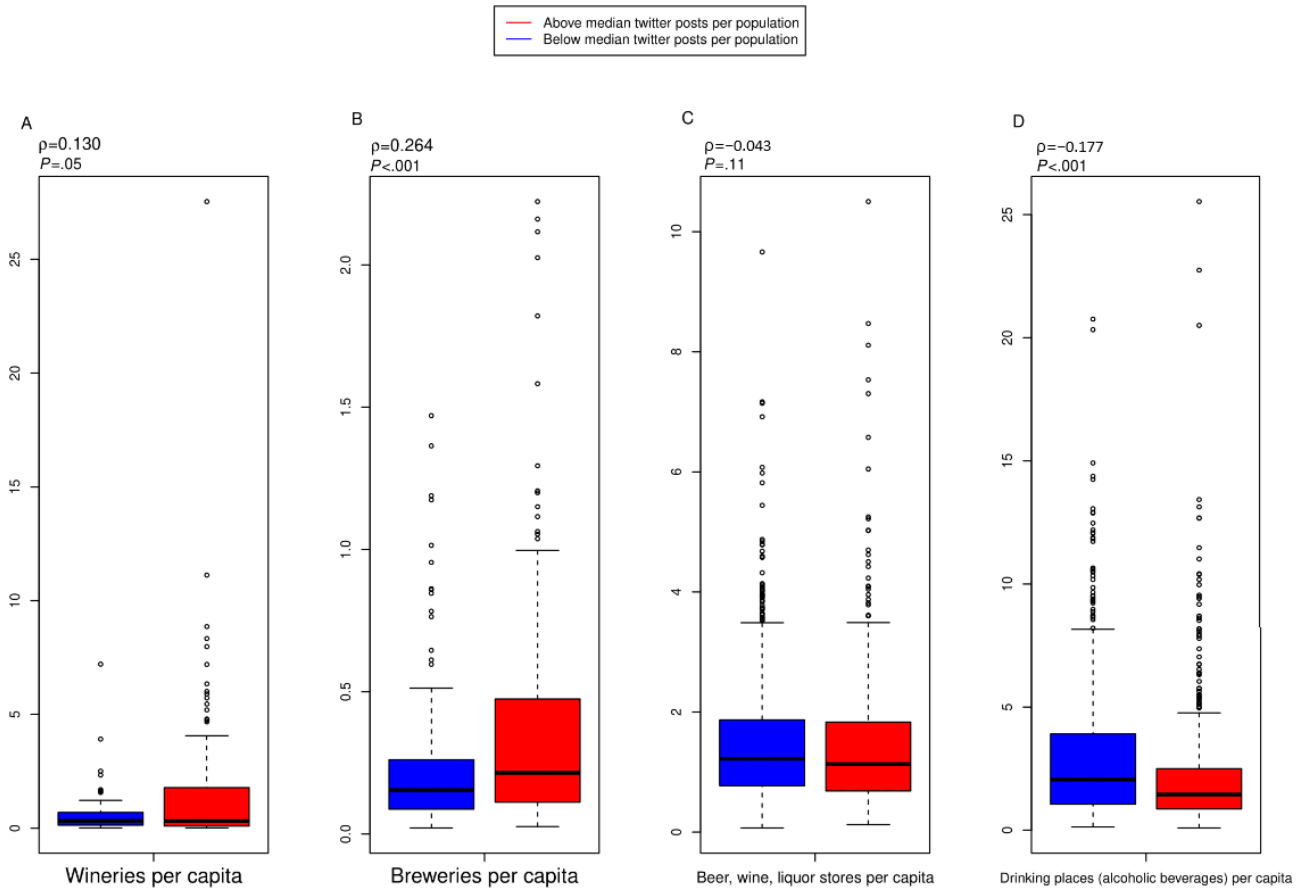
There was a significant correlation between the percentage of motor vehicle deaths reported as involving alcohol (P<.001) and aggregated measures of excessive drinking behavior (P<.001) at the county level (Figure 4). However, the relationship between alcohol-related motor vehicle fatalities and county tweets per population was not significant after adjusting for confounding effects (P=.21).

Figure 4. County- and county-equivalent–level correlations for alcohol-hashtagged tweets, alcohol motor vehicle fatalities, and self-reported excessive alcohol consumption. Counties are included if they contain at least one tweet and 1000 people (N=2641). (A) County-level correlations for alcohol-hashtagged tweets per population and percentage of motor vehicle fatalities involving alcohol. (B) County-level correlations for alcohol-hashtagged tweets per population and percent self-reporting excessive drinking (reporting either more than five alcoholic drinks on a single occasion for men or more than four alcoholic drinks on a single occasion for women, or more than two drinks per day for men or more than one drink per day for women).



There was a significant correlation between the number of alcohol-hashtagged tweets and wineries (P=.05), breweries (P<.001), and drinking places (alcoholic beverages; P<.001) but not beer, wine, and liquor stores (P=.11) per capita at the county level (Figure 5). However, after adjusting for confounding effects, there was a significant association between alcohol-related tweets per population and beer, wine, and liquor stores (P<.001) but not drinking places (P=.23) per capita.

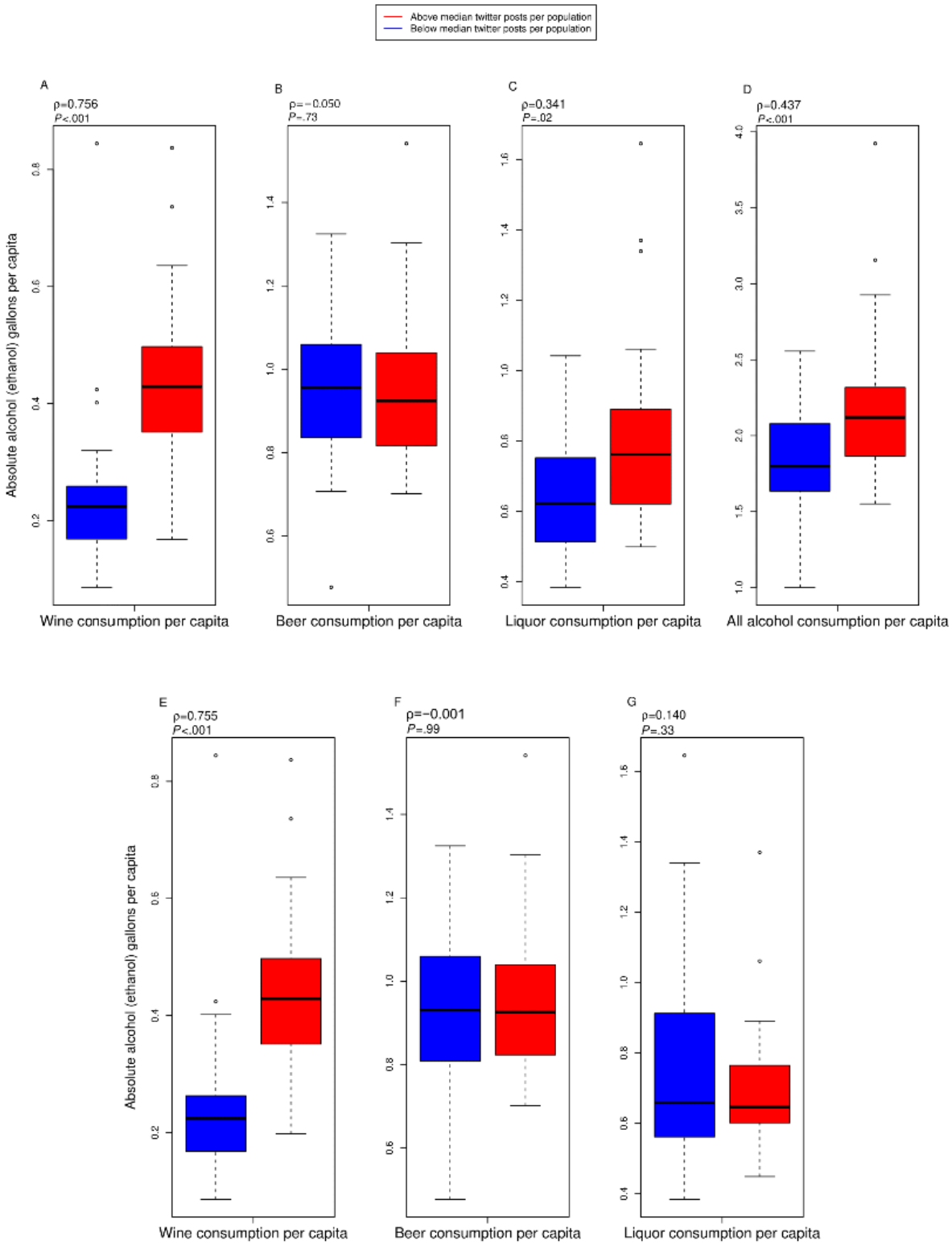
Figure 5. County- and county-equivalent–level correlations for alcohol-hashtagged tweets per person and alcohol-serving outlets, as reported by the North American Industry Classification System. Counties are included if they have at least one (1) tweet in our data set, one (1) alcohol outlet, and contain a population of 1000. (A) Wineries per 10,000 people (n=228 counties); (B) breweries per 10,000 people (n=334); (C) liquor stores per 10,000 people (n=1444); (D) drinking places (alcoholic beverages) per 10,000 people (n=1479).



There was a significant correlation between the prevalence of alcohol-hashtagged tweets and gallons of wine (P<.001), liquor (P=.01), and overall gallons of consumption (P<.001) at the state level (Figure 6). However, the association with all alcohol-hashtagged tweets and gallons of beer consumed was not significant (P=.63).

The prevalence of five wine hashtags had a significant association with wine consumption at the state level (P<.001), but 3 liquor hashtags and 19 beer hashtags did not have a significant relationship with liquor (P=.33) and beer (P=.99) consumption at the state level. However, there was a significant relationship between gallons of liquor consumed and the prevalence of 3 liquor hashtags (P=.01) after controlling for confounding effects (Figure 6).

Figure 6. Gallons of ethanol consumed for US states and alcohol-hashtagged tweets. Population normalization was performed using the average reported population between 2010 and 2018, as reported by the National Institute of Alcohol Abuse and Alcoholism. Top, all 24 hashtags used; gallons of ethanol consumed from (A) wine; (B) beer; (C) liquor; and (D) all alcohol sources and Twitter posts per population. Bottom, only specific hashtags, gallons of ethanol from wine, beer, and liquor, and Twitter posts per population from specific hashtags: (E) wine (5 wine hashtags); (F) beer (14 beer hashtags); (G) liquor (3 liquor hashtags).



Discussion

Principal Findings

We demonstrated that information from alcohol-related subreddits could identify alcohol-related hashtags that correlated with multiple alcohol-related outcomes. The prevalence of geotagged tweets containing these significantly alcohol-related hashtags correlates with alcohol-related behaviors and alcohol outlet density, which are associated with adverse outcomes, including deaths from motor vehicle crashes and excessive drinking [95-97]. This approach has distinct benefits for studying

alcohol-related outcomes. Compared with other approaches to detect alcohol consumption on Twitter, the pipeline presented here is trained using relatively few interpretable input and output parameters, specifically, 18 subreddits and 24 hashtags. The NLP pipeline identified language associated with alcohol abuse on Twitter without manual annotation or predefined keywords and with only the knowledge of relevant subreddits. These results indicate that alcohol-related language, when defined by inclusion into the themed communities of subreddits or containing alcohol-related hashtags, can be used to understand population-level behavior in multiple geographic areas with different population granularity.

Qualitatively, the model presented here detects a wide variety of different alcohol-related hashtags, including slang. Tweets containing these hashtags capture a large amount of information regarding alcohol consumption behavior at a broad population level. The set of hashtags resulting from this study is useful for future alcohol-related research on Twitter and identifying relevant hashtags or community forum labels on other platforms.

A notable benefit of using hashtags and subreddits as platform-specific labels for studying alcohol-related outcomes is interpretability. Although deep learning models excel at understanding massive quantities of data, many NLP and deep learning models rely on complex feature representations to classify or characterize text. However, this approach is not easily interpretable [65,98]. We have demonstrated that by treating subreddits and hashtags as learnable labels, it is possible to directly use hashtags as interpretable features in our NLP pipeline for social media data to understand alcoholic beverage preferences while still learning from a large amount of data.

Consumption of different alcohol types is associated with a variety of both beneficial and detrimental outcomes [99-101]. For example, wine consumption is associated with protection from cardiovascular diseases; however, the cause of this effect may be dietary factors and other lifestyle choices [99,102]. Other studies have associated preferential beer and liquor consumption with adverse outcomes, such as dangerous drinking and other risky behaviors [101,103,104]. Notably, although many studies have examined overall alcohol mentions on Twitter, few models created have been explicitly examined in terms of differences in the types of alcoholic beverages mentioned.

Our study indicates that information capturing consumption of wine and liquor is directly observable using social media data, as shown by the significant associations between the prevalence of 5 wine-related hashtags and the amount of wine consumed, as well as the number of posts containing at least 1 of the 3 liquor-related hashtags and liquor consumed. However, there was no significant relationship between beer consumption per capita and the number of alcohol-hashtagged Twitter posts in an area. The results here indicate that our model can detect certain types of alcohol consumption behavior (wine and liquor consumption) on Twitter using the interpretability of hashtags but not others (beer consumption). It remains unclear whether our results indicate a bias in our model's methodological choice (eg, use of hashtags or training procedure) or a difference in social media populations that prefer different alcohol beverages. The difference in correlations might be because of several variables, including the existence of confounding factors related to the prevalence of social media use in the underlying populations that preferentially consume beer over other alcohol types, differences in perceived acceptance of beer consumption behavior, or because of other factors that may confound alcohol and social media posting [89]. Alcohol preference is an example of having interpretable hashtag representations for a given model that may help identify behavioral differences associated with an outcome of interest. This evidence suggests that similar models trained on Twitter may detect alcohol wine and liquor consumption but not beer.

Comparison With Previous Works

Many previous studies have used hashtags as target labels. However, they mostly rely on a predefined set of hashtags that may not be data-driven or require extensive expert annotation instead of taking advantage of topic-specific sources and social media content on other public social media platforms [40,105-107]. Notably, using a predefined number of hashtags could be biased and too narrow for capturing relevant information, potentially missing informative hashtags for exposure or outcome. Traditional keyword approaches may fail to capture various pieces of information from slang and novel hashtags from platform-specific languages as they are created and popularized. In addition, keyword databases may not exist for all outcomes of interest. Semantic similarity measures, such as Word2vec, may identify hashtags with similar contexts; however, integrating vector representations may lose valuable information relative to training over individual samples, and prediction probabilities or certainty are not readily observable. In contrast, this study indicates that a model trained on a large set of data relevant to the behavior of interest and its application to an unlabeled data set from a different platform can identify data-driven hashtags related to that behavior. This ability to learn hashtags from data is critical, as new hashtags are created every day and may differ substantially between platforms.

Many social media platforms are directly searchable using hashtags, allowing the ability to gather many highly specific posts instead of gathering a large number of nonspecific posts to identify relevant hashtags, keywords, or alcohol content based on available prevalence data. Although the latter approach has shown success in studying alcohol-related behavior previously, methods to extend the analysis to alternative but potentially related outcomes, platforms, or different geographical designations remain unclear [39,40,108]. The generalization of such models necessitates the creation of individual models for each particular outcome and geographic area. Finally, these models may fail to take advantage of extensive research on themed communities to understand alcohol use and other outcomes of interest [30,39,40,76,106,109,110]. In contrast, the method outlined here may help create more efficient public health interventions to analyze the alcohol consumption behavior for a given geographic area of interest, such as a city or hospital catchment area. This approach is particularly useful when the relevant language is dynamic or contains area-specific slang, making previously established dictionary-based methods incomplete or impractical. In particular, our proposed methodology for identifying hashtags using a previously trained

deep learning model can be useful for detecting alcohol consumption behavior on various social media platforms.

Limitations and Future Work

There are some limitations to the model and approach proposed in this study. Other health-related behaviors that are not discussed frequently on social media may be more difficult to ascertain and translate between platforms using hashtags. Any model trained on a limited number of social media platforms may be confounded by differences in user preferences, such as age or socioeconomic status. Furthermore, this method relies on identifying known or previously studied subreddits, which may not be suitable for outcomes without known relevant subreddits. Furthermore, we did not compare our models' performance with the BERT large model or other deep learning alternatives, which can have a different performance for our task. In addition, including additional social media data in model development and using domain knowledge from ontologies, controlled vocabularies, lexicons, and relevant rules and regular expressions can further improve the presented results.

As future work, we plan to extend this study to other mental and behavioral health topics, such as depression and substance use, and other social media platforms that use hashtags, such as Facebook and Instagram. Deep learning has previously been used to combine the analysis of images and texts from social media users [38]. In our future work, we will expand the presented architecture to include other data modalities, such as images and videos, to increase screening capabilities.

Conclusions

These results indicate that using alcohol-related subreddits as learnable labels to train a BERT neural network can capture interpretable, alcohol-related language on Twitter. Our study suggests a significant correlation between the prevalence of alcohol-related geotagged Twitter hashtags and alcohol-related behaviors as measured by self-reported alcohol consumption, alcohol preferences, and alcohol outlet prevalence. This method has the unique advantages of previous methods, including allowing examination at the MMSA, US county, and US state level for different alcohol-related outcomes. These results suggest that using previously studied hashtags and subreddits as learnable targets in a machine learning framework could expand public health outreach efforts and epidemiology research, particularly for monitoring behavior related to alcohol consumption.

Acknowledgments

Financial support for this work was provided by a National Library of Medicine of the National Institutes of Health T32 training grant through the award number T32LM012204. The authors would like to thank Lamar Moss, Natasha de la Rosa-Rivera, and Richard Ricard for their help and suggestions for improving the manuscript.

Authors' Contributions

Study concept and design: BJR and SH; analysis and interpretation of data: BJR and SH; collection or assembly of data: BJR; drafting of the manuscript: BJR and SH; critical revision of the manuscript for important intellectual content: BJR and SH; funding: BJR and SH; administrative, technical, and logistic support: SH; and study supervision: SH.

Conflicts of Interest

None declared.

References

1. Osna NA, Donohue TM, Kharbanda KK. Alcoholic liver disease: pathogenesis and current management. Alcohol Res 2017;38(2):147-161 [FREE Full text] [Medline: 28988570]
2. O'Keefe JH, Bhatti SK, Bajwa A, DiNicolantonio JJ, Lavie CJ. Alcohol and cardiovascular health: the dose makes the poison…or the remedy. Mayo Clin Proc 2014 Mar;89(3):382-393. [doi: 10.1016/j.mayocp.2013.11.005] [Medline: 24582196]
3. Alcohol and public health: Alcohol-Related Disease Impact (ARDI). Centers for Disease Control and Prevention (CDC). URL: https://nccd.cdc.gov/DPH_ARDI/default/default.aspx [accessed 2021-08-24]
4. Zhao Q, Sullivan EV, Honnorat N, Adeli E, Podhajsky S, De Bellis MD, et al. Association of heavy drinking with deviant fiber tract development in frontal brain systems in adolescents. JAMA Psychiatry 2021 Apr 01;78(4):407-415. [doi: 10.1001/jamapsychiatry.2020.4064] [Medline: 33377940]
5. Litt DM, Lewis MA, Spiro ES, Aulck L, Waldron KA, Head-Corliss MK, et al. #drunktwitter: examining the relations between alcohol-related Twitter content and alcohol willingness and use among underage young adults. Drug Alcohol Depend 2018 Dec 01;193:75-82 [FREE Full text] [doi: 10.1016/j.drugalcdep.2018.08.021] [Medline: 30343237]
6. Perrin A, Anderson M. Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/ [accessed 2021-08-24]
7. 2018 National Survey of Drug Use and Health (NSDUH) releases. SAMHSA. URL: https://www.samhsa.gov/data/release/2018-national-survey-drug-use-and-health-nsduh-releases [accessed 2021-08-24]

8. Bonar EE, Schneeberger DM, Bourque C, Bauermeister JA, Young SD, Blow FC, et al. Social media interventions for risky drinking among adolescents and emerging adults: protocol for a randomized controlled trial. JMIR Res Protoc 2020 May 13;9(5):e16688 [FREE Full text] [doi: 10.2196/16688] [Medline: 32401225]
9. Arigo D, Pagoto S, Carter-Harris L, Lillie SE, Nebeker C. Using social media for health research: methodological and ethical considerations for recruitment and intervention delivery. Digit Health 2018 May 7;4:2055207618771757 [FREE Full text] [doi: 10.1177/2055207618771757] [Medline: 29942634]
10. Eysenbach G. Infodemiology: the epidemiology of (mis)information. Am J Med 2002 Dec 15;113(9):763-765. [doi: 10.1016/s0002-9343(02)01473-0] [Medline: 12517369]
11. Mavragani A. Infodemiology and infoveillance: scoping review. J Med Internet Res 2020 Apr 28;22(4):e16206 [FREE Full text] [doi: 10.2196/16206] [Medline: 32310818]
12. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
13. Martino F, Brooks R, Browne J, Carah N, Zorbas C, Corben K, et al. The nature and extent of online marketing by big food and big alcohol during the COVID-19 pandemic in Australia: content analysis study. JMIR Public Health Surveill 2021 Mar 12;7(3):e25202. [doi: 10.2196/25202] [Medline: 33709935]
14. Song T, Qian S, Yu P. Mobile health interventions for self-control of unhealthy alcohol use: systematic review. JMIR Mhealth Uhealth 2019 Jan 29;7(1):e10899 [FREE Full text] [doi: 10.2196/10899] [Medline: 30694200]
15. Dawson DA, Goldstein RB, Pickering RP, Grant BF. Nonresponse bias in survey estimates of alcohol consumption and its association with harm. J Stud Alcohol Drugs 2014 Jul;75(4):695-703 [FREE Full text] [doi: 10.15288/jsad.2014.75.695] [Medline: 24988268]
16. Sartor CE, Bucholz KK, Nelson EC, Madden PA, Lynskey MT, Heath AC. Reporting bias in the association between age at first alcohol use and heavy episodic drinking. Alcohol Clin Exp Res 2011 Aug;35(8):1418-1425 [FREE Full text] [doi: 10.1111/j.1530-0277.2011.01477.x] [Medline: 21438885]
17. Davis CG, Thake J, Vilhena N. Social desirability biases in self-reported alcohol consumption and harms. Addict Behav 2010 Apr;35(4):302-311. [doi: 10.1016/j.addbeh.2009.11.001] [Medline: 19932936]
18. Zappavigna M. Searchable talk: the linguistic functions of hashtags. Soc Semiot 2015 Jan 09;25(3):274-291. [doi: 10.1080/10350330.2014.996948]
19. Berry N, Lobban F, Belousov M, Emsley R, Nenadic G, Bucci S. #WhyWeTweetMH: understanding why people use twitter to discuss mental health problems. J Med Internet Res 2017 Apr 05;19(4):e107 [FREE Full text] [doi: 10.2196/jmir.6173] [Medline: 28381392]
20. Modrek S, Chakalov B. The #MeToo Movement in the United States: text analysis of early Twitter conversations. J Med Internet Res 2019 Sep 03;21(9):e13837 [FREE Full text] [doi: 10.2196/13837] [Medline: 31482849]
21. O'Neil A, Sojo V, Fileborn B, Scovelle AJ, Milner A. The #MeToo movement: an opportunity in public health? Lancet 2018 Jun 30;391(10140):2587-2589. [doi: 10.1016/S0140-6736(18)30991-7] [Medline: 30070210]
22. Kero KM, Puuronen AH, Nyqvist L, Langén VL. Usability of two brief questions as a screening tool for domestic violence and effect of #MeToo on prevalence of self-reported violence. Eur J Obstet Gynecol Reprod Biol 2020 Dec;255:92-97 [FREE Full text] [doi: 10.1016/j.ejogrb.2020.10.024] [Medline: 33113404]
23. Platt JR, Brady RR. #BCSM and #breastcancer: contemporary cancer-specific online social media communities. Breast J 2020 Apr;26(4):729-733. [doi: 10.1111/tbj.13576] [Medline: 31493301]
24. Basch C, MacLean S. Breast cancer on Instagram: a descriptive study. Int J Prev Med 2019 Oct 9;10:166 [FREE Full text] [doi: 10.4103/ijpvm.IJPVM_36_19] [Medline: 32133084]
25. George N, Britto D, Krishnan V, Dass L, Prasant H, Aravindhan V. Assessment of hashtag (#) campaigns aimed at health awareness in social media. J Educ Health Promot 2018 Sep 14;7:114 [FREE Full text] [doi: 10.4103/jehp.jehp_37_18] [Medline: 30271799]
26. Mercier RJ, Senter K, Webster R, Henderson Riley A. Instagram users' experiences of miscarriage. Obstet Gynecol 2020 Jan;135(1):166-173. [doi: 10.1097/AOG.0000000000003621] [Medline: 31809440]
27. Laestadius LI, Wahl MM, Cho YI. #Vapelife: an exploratory study of electronic cigarette use and promotion on instagram. Subst Use Misuse 2016 Oct 14;51(12):1669-1673. [doi: 10.1080/10826084.2016.1188958] [Medline: 27484191]
28. Kudchadkar S, Carroll C. Using social media for rapid information dissemination in a pandemic: #PedsICU and coronavirus disease 2019. Pediatr Crit Care Med 2020 Aug;21(8):538-546 [FREE Full text] [doi: 10.1097/PCC.0000000000002474] [Medline: 32459792]
29. Gkotsis G, Oellrich A, Velupillai S, Liakata M, Hubbard TJ, Dobson RJ, et al. Characterisation of mental health conditions in social media using informed deep learning. Sci Rep 2017 Mar 22;7:45141 [FREE Full text] [doi: 10.1038/srep45141] [Medline: 28327593]
30. Tamersoy A, De Choudhury M, Chau D. Characterizing smoking and drinking abstinence from social media. HT ACM Conf Hypertext Soc Media 2015 Sep;2015:139-148 [FREE Full text] [doi: 10.1145/2700171.2791247] [Medline: 26640831]

31. Liu C, Lu X. Analyzing hidden populations online: topic, emotion, and social network of HIV-related users in the largest Chinese online community. BMC Med Inform Decis Mak 2018 Jan 05;18(1):2 [FREE Full text] [doi: 10.1186/s12911-017-0579-1] [Medline: 29304788]
32. Pandrekar S, Chen X, Gopalkrishna G, Srivastava A, Saltz M, Saltz J, et al. Social media based analysis of opioid epidemic using Reddit. AMIA Annu Symp Proc 2018 Dec 5;2018:867-876 [FREE Full text] [Medline: 30815129]
33. Kochan A, Ong S, Guler S, Johannson KA, Ryerson CJ, Goobie GC. Social media content of idiopathic pulmonary fibrosis groups and pages on Facebook: cross-sectional analysis. JMIR Public Health Surveill 2021 May 31;7(5):e24199 [FREE Full text] [doi: 10.2196/24199] [Medline: 34057425]
34. Gruebner O, Lowe SR, Sykora M, Shankardass K, Subramanian SV, Galea S. A novel surveillance approach for disaster mental health. PLoS One 2017 Jul 19;12(7):e0181233 [FREE Full text] [doi: 10.1371/journal.pone.0181233] [Medline: 28723959]
35. Jurdak R, Zhao K, Liu J, AbouJaoude M, Cameron M, Newth D. Understanding human mobility from Twitter. PLoS One 2015 Jul 8;10(7):e0131469 [FREE Full text] [doi: 10.1371/journal.pone.0131469] [Medline: 26154597]
36. Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, et al. Artificial intelligence-enabled analysis of public attitudes on Facebook and Twitter toward COVID-19 vaccines in the united kingdom and the United States: observational study. J Med Internet Res 2021 Apr 05;23(4):e26627 [FREE Full text] [doi: 10.2196/26627] [Medline: 33724919]
37. Curtis B, Giorgi S, Buffone AE, Ungar LH, Ashford RD, Hemmons J, et al. Can Twitter be used to predict county excessive alcohol consumption rates? PLoS One 2018 Apr 4;13(4):e0194290 [FREE Full text] [doi: 10.1371/journal.pone.0194290] [Medline: 29617408]
38. Hassanpour S, Tomita N, DeLise T, Crosier B, Marsch LA. Identifying substance use risk based on deep neural networks and Instagram social media data. Neuropsychopharmacology 2019 Feb;44(3):487-494 [FREE Full text] [doi: 10.1038/s41386-018-0247-x] [Medline: 30356094]
39. Marengo D, Azucar D, Giannotta F, Basile V, Settanni M. Exploring the association between problem drinking and language use on Facebook in young adults. Heliyon 2019 Oct 9;5(10):e02523 [FREE Full text] [doi: 10.1016/j.heliyon.2019.e02523] [Medline: 31667380]
40. Crocamo C, Viviani M, Bartoli F, Carrà G, Pasi G. Detecting binge drinking and alcohol-related risky behaviours from Twitter's users: an exploratory content- and topology-based analysis. Int J Environ Res Public Health 2020 Feb 26;17(5):1510 [FREE Full text] [doi: 10.3390/ijerph17051510] [Medline: 32111047]
41. Cavazos-Rehg PA, Krauss MJ, Sowles SJ, Bierut LJ. "Hey Everyone, I'm Drunk." An evaluation of drinking-related Twitter chatter. J Stud Alcohol Drugs 2015 Jul;76(4):635-643 [FREE Full text] [doi: 10.15288/jsad.2015.76.635] [Medline: 26098041]
42. Giorgi S, Yaden DB, Eichstaedt JC, Ashford RD, Buffone AE, Schwartz HA, et al. Cultural differences in Tweeting about drinking across the US. Int J Environ Res Public Health 2020 Feb 11;17(4):1125 [FREE Full text] [doi: 10.3390/ijerph17041125] [Medline: 32053866]
43. Stellefson M, Paige SR, Chaney BH, Chaney JD. Social media and health promotion. Int J Environ Res Public Health 2020 May 11;17(9):3323 [FREE Full text] [doi: 10.3390/ijerph17093323] [Medline: 32403215]
44. Sadah SA, Shahbazi M, Wiley MT, Hristidis V. A study of the demographics of web-based health-related social media users. J Med Internet Res 2015 Aug 06;17(8):e194 [FREE Full text] [doi: 10.2196/jmir.4308] [Medline: 26250986]
45. Hendriks H, Van den Putte B, Gebhardt WA, Moreno MA. Social drinking on social media: content analysis of the social aspects of alcohol-related posts on Facebook and Instagram. J Med Internet Res 2018 Jun 22;20(6):e226 [FREE Full text] [doi: 10.2196/jmir.9355] [Medline: 29934290]
46. Laws R, Hunt G, Antin TM. Social media platforms as a photo-elicitation tool in research on alcohol intoxication and gender. Nordisk Alkohol Nark 2018 Aug;35(4):288-303 [FREE Full text] [doi: 10.1177/1455072518781998] [Medline: 30245584]
47. Gupta H, Lam T, Pettigrew S, Tait RJ. The association between exposure to social media alcohol marketing and youth alcohol use behaviors in India and Australia. BMC Public Health 2018 Jun 13;18(1):726 [FREE Full text] [doi: 10.1186/s12889-018-5645-9] [Medline: 29895264]
48. Rehm J, Kilian C, Rovira P, Shield KD, Manthey J. The elusiveness of representativeness in general population surveys for alcohol. Drug Alcohol Rev 2021 Feb;40(2):161-165. [doi: 10.1111/dar.13148] [Medline: 32830351]
49. Nguyen QC, McCullough M, Meng H, Paul D, Li D, Kath S, et al. Geotagged US Tweets as predictors of county-level health outcomes, 2015-2016. Am J Public Health 2017 Nov;107(11):1776-1782. [doi: 10.2105/AJPH.2017.303993] [Medline: 28933925]
50. Krauss MJ, Grucza RA, Bierut LJ, Cavazos-Rehg PA. "Get drunk. Smoke weed. Have fun.": a content analysis of tweets about marijuana and alcohol. Am J Health Promot 2017 May;31(3):200-208 [FREE Full text] [doi: 10.4278/ajhp.150205-QUAL-708] [Medline: 26559715]
51. Allem J, Dormanesh A, Majmundar A, Rivera V, Chu M, Unger JB, et al. Leading topics in Twitter discourse on JUUL and Puff Bar products: content analysis. J Med Internet Res 2021 Jul 19;23(7):e26510 [FREE Full text] [doi: 10.2196/26510] [Medline: 34279236]

52. Lorentzen DG, Nolin J. Approaching completeness: capturing a hashtagged Twitter conversation and its follow-on conversation. Soc Sci Comput Rev 2015 Sep 29;35(2):277-286. [doi: 10.1177/0894439315607018]
53. Le Q, Mikolov T. Distributed representations of sentences and documents. In: Proceedings of the 31st International Conference on Machine Learning. 2014 Presented at: 31st International Conference on Machine Learning; Jun 21-26, 2014; Beijing China p. 1188-1196 URL: http://proceedings.mlr.press/v32/le14.html
54. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv: Computer Science, Computation and Language. 2013 Jan. URL: https://arxiv.org/abs/1301.3781 [accessed 2021-09-02]
55. Saha K, Torous J, Ernala S, Rizuto C, Stafford A, De Choudhury M. A computational study of mental health awareness campaigns on social media. Transl Behav Med 2019 Nov 25;9(6):1197-1207 [FREE Full text] [doi: 10.1093/tbm/ibz028] [Medline: 30834942]
56. Zomick J, Levitan S, Serper M. Linguistic analysis of schizophrenia in Reddit posts. In: Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology. 2019 Presented at: Sixth Workshop on Computational Linguistics and Clinical Psychology; Jun 2019; Minneapolis, Minnesota p. 74-83. [doi: 10.18653/v1/w19-3009]
57. Devlin J, Chang M, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. Assoc Comput Linguist 2019 Jun:4171-4186. [doi: 10.18653/v1/N19-1423]
58. Rogers A, Kovaleva O, Rumshisky A. A primer in BERTology: what we know about how BERT works. Trans Assoc Comput Linguist 2020 Dec;8:842-866. [doi: 10.1162/tacl_a_00349]
59. Mozafari M, Farahbakhsh R, Crespi N. A BERT-based transfer learning approach for hate speech detection in online social media. In: Complex Networks and Their Applications VIII. Basel, Switzerland: Springer International Publishing; Nov 26, 2019.
60. Heidari M, Jones J. Using BERT to extract topic-independent sentiment features for social media bot detection. In: Proceedings of the 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). 2020 Presented at: 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON); Oct 28-31, 2020; New York, NY, USA. [doi: 10.1109/uemcon51285.2020.9298158]
61. Mozafari M, Farahbakhsh R, Crespi N. Hate speech detection and racial bias mitigation in social media based on BERT model. PLoS One 2020 Aug 27;15(8):e0237861 [FREE Full text] [doi: 10.1371/journal.pone.0237861] [Medline: 32853205]
62. Pérez-Pérez M, Pérez-Rodríguez G, Fdez-Riverola F, Lourenço A. Using Twitter to understand the human bowel disease community: exploratory analysis of key topics. J Med Internet Res 2019 Aug 15;21(8):e12610 [FREE Full text] [doi: 10.2196/12610] [Medline: 31411142]
63. Li J, Xu Q, Shah N, Mackey TK. A machine learning approach for the detection and characterization of illicit drug dealers on instagram: model evaluation study. J Med Internet Res 2019 Jun 15;21(6):e13803 [FREE Full text] [doi: 10.2196/13803] [Medline: 31199298]
64. Pereira-Kohatsu JC, Quijano-Sánchez L, Liberatore F, Camacho-Collados M. Detecting and monitoring hate speech in Twitter. Sensors (Basel) 2019 Oct 26;19(21):4654 [FREE Full text] [doi: 10.3390/s19214654] [Medline: 31717760]
65. Chakraborty S, Tomsett R, Raghavendra R, Harborne D, Alzantot M, Cerutti F, et al. Interpretability of deep learning models: a survey of results. In: Proceedings of the IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). 2017 Presented at: IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI); Aug 4-8, 2017; San Francisco, CA, USA. [doi: 10.1109/uic-atc.2017.8397411]
66. Singh T, Roberts K, Cohen T, Cobb N, Wang J, Fujimoto K, et al. Social Media as a Research Tool (SMaaRT) for risky behavior analytics: methodological review. JMIR Public Health Surveill 2020 Nov 30;6(4):e21660 [FREE Full text] [doi: 10.2196/21660] [Medline: 33252345]
67. Wang X, Kou L, Sugumaran V, Luo X, Zhang H. Emotion correlation mining through deep learning models on natural language text. IEEE Trans Cybern 2020 May 12:2987064 (forthcoming). [doi: 10.1109/TCYB.2020.2987064] [Medline: 32413938]
68. Klein AZ, Sarker A, Weissenbacher D, Gonzalez-Hernandez G. Towards scaling Twitter for digital epidemiology of birth defects. NPJ Digit Med 2019 Oct 1;2:96 [FREE Full text] [doi: 10.1038/s41746-019-0170-5] [Medline: 31583284]
69. Klein AZ, Magge A, O'Connor K, Amaro JF, Weissenbacher D, Hernandez GG. Toward using Twitter for tracking COVID-19: a natural language processing pipeline and exploratory data set. J Med Internet Res 2021 Jan 22;23(1):e25314 [FREE Full text] [doi: 10.2196/25314] [Medline: 33449904]
70. Edo-Osagie O, Smith G, Lake I, Edeghere O, De La Iglesia B. Twitter mining using semi-supervised classification for relevance filtering in syndromic surveillance. PLoS One 2019 Jul 18;14(7):e0210689 [FREE Full text] [doi: 10.1371/journal.pone.0210689] [Medline: 31318885]
71. Ru B, Li D, Hu Y, Yao L. Serendipity-a machine-learning application for mining serendipitous drug usage from social media. IEEE Trans Nanobioscience 2019 Jul;18(3):324-334 [FREE Full text] [doi: 10.1109/TNB.2019.2909094] [Medline: 30951476]

72. Nikfarjam A, Sarker A, O'Connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. J Am Med Inform Assoc 2015 May;22(3):671-681 [FREE Full text] [doi: 10.1093/jamia/ocu041] [Medline: 25755127]
73. Strubell E, Ganesh A, McCallum A. Energy and policy considerations for modern deep learning research. AAAI 2020 Apr 03;34(09):13693-13696. [doi: 10.1609/aaai.v34i09.7123]
74. Metropolitan and micropolitan statistical areas population totals and components of change: 2010-2019. United States Census Bureau. URL: https://www.census.gov/data/tables/time-series/demo/popest/2010s-total-metro-and-micro-statistical-areas.html [accessed 2021-08-24]
75. Population, population change, and estimated components of population change: April 1, 2010 to July 1, 2019 (NST-EST2019-alldata). United States Census Bureau. URL: https://www.census.gov/data/tables/time-series/demo/popest/2010s-national-total.html [accessed 2021-08-24]
76. Zhan Y, Zhang Z, Okamoto JM, Zeng DD, Leischow SJ. Underage JUUL use patterns: content analysis of Reddit messages. J Med Internet Res 2019 Sep 09;21(9):e13038 [FREE Full text] [doi: 10.2196/13038] [Medline: 31502542]
77. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: an imperative style, high-performance deep learning library. arXiv: Computer Science, Machine Learning. URL: https://arxiv.org/abs/1912.01703 [accessed 2021-08-24]
78. Pennington J, Socher R, Manning C. GloVe: global vectors for word representation. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014 Presented at: Conference on Empirical Methods in Natural Language Processing (EMNLP); Oct 2014; Doha, Qatar p. 1532-1543 URL: http://www.aclweb.org/anthology/D14-1162 [doi: 10.3115/v1/d14-1162]
79. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 1995;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
80. Hopkins ZH, Moreno C, Secrest AM. Influence of social media on cosmetic procedure interest. J Clin Aesthet Dermatol 2020 Jan;13(1):28-31 [FREE Full text] [Medline: 32082468]
81. Merchant RM, Asch DA, Crutchley P, Ungar LH, Guntuku SC, Eichstaedt JC, et al. Evaluating the predictability of medical conditions from social media posts. PLoS One 2019 Jun 17;14(6):e0215476. [doi: 10.1371/journal.pone.0215476] [Medline: 31206534]
82. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoţiuc-Pietro D, et al. Facebook language predicts depression in medical records. Proc Natl Acad Sci U S A 2018 Oct 30;115(44):11203-11208 [FREE Full text] [doi: 10.1073/pnas.1802331115] [Medline: 30322910]
83. Griffis H, Asch D, Schwartz H, Ungar L, Buttenheim A, Barg F, et al. Using social media to track geographic variability in language about diabetes: analysis of diabetes-related tweets across the United States. JMIR Diabetes 2020 Jan 26;5(1):e14431 [FREE Full text] [doi: 10.2196/14431] [Medline: 32044757]
84. Smith RJ, Crutchley P, Schwartz HA, Ungar L, Shofer F, Padrez KA, et al. Variations in Facebook posting patterns across validated patient health conditions: a prospective cohort study. J Med Internet Res 2017 Jan 06;19(1):e7 [FREE Full text] [doi: 10.2196/jmir.6486] [Medline: 28062392]
85. Guntuku SC, Schneider R, Pelullo A, Young J, Wong V, Ungar L, et al. Studying expressions of loneliness in individuals using twitter: an observational study. BMJ Open 2019 Nov 04;9(11):e030355 [FREE Full text] [doi: 10.1136/bmjopen-2019-030355] [Medline: 31685502]
86. Cecinati F, Matthews T, Natarajan S, McCullen N, Coley D. Mining social media to identify heat waves. Int J Environ Res Public Health 2019 Mar 02;16(5):762 [FREE Full text] [doi: 10.3390/ijerph16050762] [Medline: 30832387]
87. Birnbaum ML, Ernala SK, Rizvi AF, De Choudhury M, Kane JM. A collaborative approach to identifying social media markers of schizophrenia by employing machine learning and clinical appraisals. J Med Internet Res 2017 Aug 14;19(8):e289 [FREE Full text] [doi: 10.2196/jmir.7956] [Medline: 28807891]
88. Spearman C. The proof and measurement of association between two things. Am J Psychol 1904;15(1):72-101. [doi: 10.2307/1412159]
89. Bergman BG, Wu W, Marsch LA, Crosier BS, DeLise TC, Hassanpour S. Associations between substance use and Instagram participation to inform social network-based screening models: multimodal cross-sectional study. J Med Internet Res 2020 Sep 16;22(9):e21916 [FREE Full text] [doi: 10.2196/21916] [Medline: 32936081]
90. United States Census Bureau. Educational Attainment, Income in the Past 12 Months (In 2018 inflation-adjusted dollars), ACS Demographic and Housing Estimates. American Community Survey. 2018. URL: https://www.census.gov/acs/www/data/data-tables-and-tools/subject-tables/
91. Behavioral risk factor surveillance system. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/brfss/index.html [accessed 2021-08-24]
92. 2019 County Health Rankings Key Findings Report. University of Wisconsin Population Health Institute. URL: https://www.countyhealthrankings.org/reports/2019-county-health-rankings-key-findings-report [accessed 2021-08-24]
93. County business patterns: 2017. United States Census Bureau. URL: https://www.census.gov/data/datasets/2017/econ/cbp/2017-cbp.html [accessed 2021-08-24]

94. Apparent per capita alcohol consumption: national, state, and regional trends 1977-2018. Open ICPSR. URL: https://www.openicpsr.org/openicpsr/project/105583/version/V5/view;jsessionid=D71A8F94F8F63A42806C3C45BD042D9D [accessed 2021-08-24]
95. Kavanagh A, Kelly M, Krnjacki L, Thornton L, Jolley D, Subramanian S, et al. Access to alcohol outlets and harmful alcohol consumption: a multi-level study in Melbourne, Australia. Addiction 2011 Oct;106(10):1772-1779. [doi: 10.1111/j.1360-0443.2011.03510.x] [Medline: 21615583]
96. Popova S, Giesbrecht N, Bekmuradov D, Patra J. Hours and days of sale and density of alcohol outlets: impacts on alcohol consumption and damage: a systematic review. Alcohol Alcohol 2009;44(5):500-516. [doi: 10.1093/alcalc/agp054] [Medline: 19734159]
97. Treno AJ, Johnson FW, Remer LG, Gruenewald PJ. The impact of outlet densities on alcohol-related crashes: a spatial panel approach. Accid Anal Prev 2007 Sep;39(5):894-901. [doi: 10.1016/j.aap.2006.12.011] [Medline: 17275773]
98. Belinkov Y, Gehrmann S, Pavlick E. Interpretability and analysis in neural NLP. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts.: Association for Computational Linguistics; 2020 Presented at: 58th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts; Jul, 2020; Online p. 1-5. [doi: 10.18653/v1/2020.acl-tutorials.1]
99. Barefoot JC, Grønbaek M, Feaganes JR, McPherson RS, Williams RB, Siegler IC. Alcoholic beverage preference, diet, and health habits in the UNC Alumni Heart Study. Am J Clin Nutr 2002 Aug;76(2):466-472. [doi: 10.1093/ajcn/76.2.466] [Medline: 12145024]
100. Sluik D, Bezemer R, Sierksma A, Feskens E. Alcoholic beverage preference and dietary habits: a systematic literature review. Crit Rev Food Sci Nutr 2016 Oct 25;56(14):2370-2382. [doi: 10.1080/10408398.2013.841118] [Medline: 25674684]
101. Dey M, Gmel G, Studer J, Dermota P, Mohler-Kuo M. Beverage preferences and associated drinking patterns, consequences and other substance use behaviours. Eur J Public Health 2014 Jun;24(3):496-501. [doi: 10.1093/eurpub/ckt109] [Medline: 23940073]
102. Alcácera MA, Marques-Lopes I, Fajó-Pascual M, Foncillas JP, Carmona-Torre F, Martínez-González MA. Alcoholic beverage preference and dietary pattern in Spanish university graduates: the SUN cohort study. Eur J Clin Nutr 2008 Oct;62(10):1178-1186. [doi: 10.1038/sj.ejcn.1602833] [Medline: 17609695]
103. Siegel MB, Naimi TS, Cremeens JL, Nelson DE. Alcoholic beverage preferences and associated drinking patterns and risk behaviors among high school youth. Am J Prev Med 2011 Apr;40(4):419-426. [doi: 10.1016/j.amepre.2010.12.011] [Medline: 21406275]
104. McCann SE, Sempos C, Freudenheim JL, Muti P, Russell M, Nochajski TH, et al. Alcoholic beverage preference and characteristics of drinkers and nondrinkers in western New York (United States). Nutr Metab Cardiovasc Dis 2003 Feb;13(1):2-11. [doi: 10.1016/s0939-4753(03)80162-x] [Medline: 12772432]
105. Lachmar EM, Wittenborn AK, Bogen KW, McCauley HL. #MyDepressionLooksLike: examining public discourse about depression on Twitter. JMIR Ment Health 2017 Oct 18;4(4):e43 [FREE Full text] [doi: 10.2196/mental.8141] [Medline: 29046270]
106. Alberga A, Withnell S, von Ranson KM. Fitspiration and thinspiration: a comparison across three social networking sites. J Eat Disord 2018 Nov 26;6:39 [FREE Full text] [doi: 10.1186/s40337-018-0227-x] [Medline: 30534376]
107. Shepherd A, Sanders C, Doyle M, Shaw J. Using social media for support and feedback by mental health service users: thematic analysis of a twitter conversation. BMC Psychiatry 2015 Feb 19;15:29 [FREE Full text] [doi: 10.1186/s12888-015-0408-y] [Medline: 25881089]
108. van Swol LM, Chang C, Kerr B, Moreno M. Linguistic predictors of problematic drinking in alcohol-related Facebook posts. J Health Commun 2020 Mar 03;25(3):214-222 [FREE Full text] [doi: 10.1080/10810730.2020.1731632] [Medline: 32096449]
109. Park A, Conway M, Chen AT. Examining thematic similarity, difference, and membership in three online mental health communities from reddit: a text mining and visualization approach. Comput Hum Behav 2018 Jan;78:98-112 [FREE Full text] [doi: 10.1016/j.chb.2017.09.001] [Medline: 29456286]
110. Kavuluru R, Williams AG, Ramos-Morales M, Haye L, Holaday T, Cerel J. Classification of helpful comments on online suicide watch forums. ACM BCB 2016 Oct;2016:32-40 [FREE Full text] [doi: 10.1145/2975167.2975170] [Medline: 28736770]

Abbreviations

API: application programming interface
BERT: Bidirectional Encoder Representations From Transformers
MMSA: metropolitan-micropolitan statistical area
NLP: natural language processing

Edited by G Eysenbach; submitted 21.01.21; peer-reviewed by Y Mao; comments to author 11.02.21; revised version received 30.03.21; accepted 01.08.21; published 15.09.21.

Please cite as:
Ricard BJ, Hassanpour S
Deep Learning for Identification of Alcohol-Related Content on Social Media (Reddit and Twitter): Exploratory Analysis of Alcohol-Related Outcomes
J Med Internet Res 2021;23(9):e27314
URL: https://www.jmir.org/2021/9/e27314
doi:10.2196/27314
PMID:34524095

Original Paper

Trends and Patterns in the Public Awareness of Palliative Care, Euthanasia, and End-of-Life Decisions in 3 Central European Countries Using Big Data Analysis From Google: Retrospective Analysis

Matthias Huemer[1], MD; Daniela Jahn-Kuch[1], MSc, MD; Guenter Hofmann[1], MD; Elisabeth Andritsch[2], MSc; Clemens Farkas[2], MSc; Walter Schaupp[3], PhD, MD; Eva Katharina Masel[4], MSc, PhD, MD; Philipp J Jost[2,5], PhD, MD; Martin Pichler[2], MSc, MD

[1]Division of Oncology with affiliated Unit of Palliative Medicine, Department of Internal Medicine, Medical University of Graz, Graz, Austria

[2]Division of Oncology, Department of Internal Medicine, Medical University of Graz, Graz, Austria

[3]Institute of Moral Theology, University of Graz, Graz, Austria

[4]Clinical Division of Palliative Care, Department of Medicine I and Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria

[5]Department of Internal Medicine III - Hematooncology, Technical University of Munich School of Medicine, Technical University of Munich, Munich, Germany

Corresponding Author:
Martin Pichler, MSc, MD
Division of Oncology
Department of Internal Medicine
Medical University of Graz
Auenbruggerplatz 15
Graz, 8036
Austria
Phone: 43 316385 ext 30196
Fax: 43 31638513355
Email: martin.pichler@medunigraz.at

Abstract

Background: End-of-life decisions, specifically the provision of euthanasia and assisted suicide services, challenge traditional medical and ethical principles. Austria and Germany have decided to liberalize their laws restricting assisted suicide, thus reigniting the debate about a meaningful framework in which the practice should be embedded. Evidence of the relevance of assisted suicide and euthanasia for the general population in Germany and Austria is limited.

Objective: The aim of this study is to examine whether the public awareness documented by search activities in the most frequently used search engine, Google, on the topics of *palliative care*, *euthanasia*, and *advance health care directives* changed with the implementation of palliative care services and new governmental regulations concerning end-of-life decisions.

Methods: We searched for policies, laws, and regulations promulgated or amended in Austria, Germany, and Switzerland between 2004 and 2020 and extracted data on the search volume for each search term topic from Google Trends as a surrogate of public awareness and interest. Annual averages were analyzed using the Joinpoint Regression Program.

Results: Important policy changes yielded significant changes in search trends for the investigated topics. The enactment of laws regulating advance health care directives coincided with a significant drop in the volume of searches for the topic of euthanasia in all 3 countries (Austria: −24.48%, P=.02; Germany: −14.95%, P<.001; Switzerland: −11.75%, P=.049). Interest in palliative care increased with the availability of care services and the implementation of laws and policies to promote palliative care (Austria: 22.69%, P=.01; Germany: 14.39, P<.001; Switzerland: 17.59%, P<.001). The search trends for advance health care directives showed mixed results. While interest remained steady in Austria within the study period, it increased by 3.66% (P<.001) in Switzerland and decreased by 2.85% (P<.001) in Germany.

Conclusions: Our results demonstrate that legal measures securing patients' autonomy at the end of life may lower the search activities for topics related to euthanasia and assisted suicide. Palliative care may be a meaningful way to raise awareness of the different options for end-of-life care and to guide patients in their decision-making process regarding the same.

(J Med Internet Res 2021;23(9):e28635) doi:10.2196/28635

KEYWORDS

Google Trends; end-of-life decisions; assisted suicide; euthanasia; palliative care; health care policy

Introduction

Background

How and to what extent a person may wish to be cared for when terminally ill and shortly before death should depend, without a doubt, solely on the expressed will of the concerned individual. Freedom of choice also includes the personal right to choose death over life [1]. In the case of advance health care directives, the right to a nontreatment decision and the consensual withdrawal of potentially curative treatments are protected by regulating acts in most Western countries [2]. Furthermore, suicide is not considered a crime in such places [2]. However, there is an ongoing debate on the extent to which a second person should assist a person in willingly dying [3]. Legal regulations concerning euthanasia and assisted suicide vary considerably among European countries. Although the Netherlands, Luxembourg, and Belgium allow euthanasia and assisted suicide under specific circumstances, Switzerland has more liberal legislation concerning assisted suicide but not euthanasia. Other countries in Europe strictly prohibit a death that involves a second person [2].

Principally, assisted suicide is not prohibited in Germany, as suicide per se is not considered a crime, but its commercial practice was restricted in 2015 [2]. On the other hand, Austria considers assisted suicide a crime per se and prohibits its practice entirely [2]. However, in 2020, the debate about end-of-life decisions was reignited in Germany and Austria after the constitutional courts in both countries declared the restriction of assisted suicide unconstitutional [4,5]. Both countries are currently in the process of liberalizing their legal regulations concerning this topic. However, voluntary death on demand remains prohibited in both countries [4,5].

Palliative care is often considered an alternative and sometimes even contradictory to the active termination of a patient's life upon voluntary request [6]. The traditional aim of palliative care is to improve the quality of life of a severely ill person and to neither shorten nor unnecessarily prolong that life [7]. Hence, the European Association of Palliative Care (EAPC) dissociated itself from implementing euthanasia and assisted suicide in the context of palliative care in a recently published white paper [6]. Furthermore, the preventive effect of palliative care on the desire to hasten death and the request for euthanasia or assisted suicide has been repeatedly proposed but never confirmed [8,9]. It is more likely that palliative care may act as a filter for patients to differentiate between those who seek euthanasia and assisted suicide because of intrinsic reasons and those who have reactive suicidal thoughts because of severe temporary suffering [8,10]. The desire to die may be either a true conviction of euthanasia being the only acceptable option for the patient or a way to express suffering without the explicit wish to hasten death actively [10]. Suffering could be alleviated through palliative care, potentially leading to a lower desire to die in some patients [11]. However, it seems that palliative care does not prevent the request for euthanasia in patients who are already truly convinced to commit assisted suicide or euthanasia [8]. Hence, patients expressing a desire to die should be referred to palliative care to explore the underlying reasons resulting in a desire to hasten death before granting a request for assisted suicide or euthanasia [8-11].

Surveys among the general public in Germany and Austria show an increasing acceptance of euthanasia and assisted suicide [12-17]. However, these surveys often reflect only a general opinion toward the topic and rarely ask specifically about the demand for such options [12-17]. Usually, hypothetical scenarios of terminally ill persons are presented, and the participants are asked if euthanasia or assisted suicide is an appropriate option in a particular case [12-15]. In addition, some of the studies carry a particular risk of selection bias because of the topic's controversial nature [12,14]. People who are unwilling to participate because of their fundamental beliefs may be underrepresented. Furthermore, proportionally high participation of ideological supporters of euthanasia and assisted suicide may influence the outcome through a self-selection bias in these voluntary surveys [14].

A novel approach for estimating public interests, awareness, and behavior is the analysis of big data obtained from internet search engines such as Google, Yahoo!, and Bing [18]. Recorded Google search volumes are publicly available in anonymized form on the Google Trends platform [19]. The data have already been used to measure public interest in cancer screening [20] and trend changes in the public awareness of palliative care after the implementation of new governmental policies [21] or the death of a celebrity [22]. Google Trends was also used in a variety of other medical and health-related disciplines [23]. It was used to predict and track infectious diseases such as influenza [24] and Lyme disease [25] with comparable accuracy to traditional methods and represents a valuable epidemiological tool.

Compared with traditional surveys that assess opinions, a Google Trends analysis is a useful tool for estimating information needs. Such trends serve as a proxy for the demand for health care services, information needs, and curiosity about health-related topics [23]. A survey across the European Union on internet use found that 58.71% (15,598/26,566) of the respondents used the internet to obtain health-related information [26]. Of those who tried to find specific information about medical treatments and procedures, 76.5% (2724/3561) conducted searches for themselves, and 30.3% (1079/3561) had done so for a family member [26]. A total of 40.51% (1119/2762) of those who rated their health as *bad* used the internet to obtain health-related information [26]. Hence, it is clear that the internet is an important source of information for Europeans, and search volumes represent the subjects of interest of a vast number of European Union citizens.

Objective

In Germany and Austria, no data about the demand for euthanasia or assisted suicide are available to date, as no register or comparable data acquisition is possible because of the restriction and prohibition of the practices, respectively. Therefore, in this study, we hypothesize that internet search behavior might provide a novel approach to gaining essential insights into the general population's needs and interests concerning palliative care, euthanasia, and end-of-life decisions. This study further aims to assess the influence of expanding care services and established governmental health care policies in both countries on the search behavior for each topic.

In addition, we chose Switzerland as a comparator as it is a (partly) German-speaking country and has already established practices for assisted suicide.

Methods

Data Collection

Google Trends data are available from January 2004 onward [19]. Therefore, we searched for relevant governmental policies, enacted or revised laws, and publicized health care strategies that emerged between January 2004 and December 2020 in each country (Germany, Austria, and Switzerland).

Every 6 years (starting in 2007), the EAPC publishes the *EAPC Atlas of Palliative Care in Europe* [27-29]. The report reviews the current status and availability of palliative care services for every European country. The available services per 100,000 citizens were obtained from each report to assess the development of palliative care over time.

Google gathers data on every search query performed on its search engine. The summarized data are publicly available in anonymous form on the Google Trends webpage [19]. Up to 5 search terms or topics can be simultaneously entered on the page. In contrast to search terms, search topics summarize several interrelated search terms and represent interest in a given topic more accurately. In addition, 4 filters can be applied: region, time, category, and search type. The output is the relative volume of queries for the search term or topic of interest to the total volume of queries in the selected region and period. The result is then scaled on a range of 0-100. Repeated search queries of a single user within a short time are not recorded so as to prevent intentional influence and enhance data quality [18].

In our study, we used three search term topics: *palliative care* plus *euthanasia* plus *advance health care directive*. All 3 terms were entered in the German language. The search topic *euthanasia* summarizes queries regarding assisted suicide and voluntary euthanasia. On January 25, 2021, we downloaded the data on the monthly relative search volume from January 2004 to December 2020 for Germany, Austria, and Switzerland. We used *all categories* and *web search* as additional filters. After downloading the data, we calculated the annual mean for each search topic and performed a trend analysis using the Joinpoint Regression Program [30]. Finally, we compared the trends in relative search volumes and the significant turning points (joinpoints) to the implementation or revision of government policy changes.

Data Analysis

The Joinpoint Regression Program allows the analysis of trends over time and identifies apparent changes or turning points in the direction of a trend [30]. We calculated the yearly average from the monthly relative search volumes for each keyword and region and performed a joinpoint analysis using a log-linear Poisson regression model. Model selection was based on the Bayesian information criterion. This approach has already been established in previous studies investigating the impact of governmental policies on public awareness and internet search behavior [31].

Results

Legal Developments After 2004

Table 1 shows a timeline of public health strategies and passed or revised laws for each country. Before 2004, palliative care services were already established in all 3 countries but were still in their infancy [32]. At the turn of the millennium, the first health care policies concerning the care of terminally ill patients were implemented [33-35]. After that, in all 3 countries, continuous advancements were made concerning the financing, education, and implementation strategies for palliative care. Legal regulations about end-of-life decisions and advance health care declarations were first passed in 2006 in Austria [36], followed by Germany in 2009 [37] and Switzerland in 2013 [38]. The provision granted patients a higher degree of autonomy in their health care decisions. The legislature in Austria further discriminates between legally binding and nonbinding advance health care directives [36]. The latter aims to communicate the patient's preferences for end-of-life care and allows the treating physician to reconsider a previously dismissed medical treatment or procedure depending on the situation, whereas the legally binding form provides a detailed and definite description of nontreatment decisions in predefined health conditions [36].

Table 1. Timeline of public health strategies and laws passed or revised after 2004 for Austria, Germany, and Switzerland.

Year	Austria	Germany	Switzerland
2004	Public health care strategy: concept of graded hospice and palliative care in Austria	Register of health care proxies	—[a]
2006	Law: advance health care directive (legally binding and nonbinding)	—	The Swiss Academy of Medical Sciences issues medical ethics guidelines for palliative care
2007	Law: health care proxy	Law: specialized ambulatory palliative care as a mandatory insurance benefit	—
2008	—	—	Novelty: adult guardianship law
2009	—	Law: advance health care directive	National strategy: Palliative Care 2010-2012
2012	—	—	National strategy: Palliative Care 2013-2015
2013	—	—	Law: advance health care directive
2015	A parliamentary commission of inquiry: "Dignity at the end of life" with 51 resolutions for financing, promoting, and establishing palliative care services	Law: hospice and palliative medicine statute; prohibition of services for assisted suicide	Establishment: National Platform for Palliative Care
2018	Novelty: adult guardianship law	—	—
2020	Constitutional court rescinds the prohibition of assisted suicide	Constitutional court rescinds the prohibition of services for assisted suicide	—

[a]No public health strategies and laws were passed or revised during this year.

Assisted suicide has been exempt from punishment since 1918 in Switzerland and 1751 in Germany, whereas it was explicitly prohibited in Austria in 1934 [39]. However, Germany restricted the commercial use of assisted suicide in 2015, meaning that neither a business-like organization nor a nonprofit institution could offer assistance in dying [4]. In 2020, the constitutional courts in Germany and Austria revised their restrictive laws and declared them unjust, paving the way for the legal practice of assisted suicide [4,5]. In both countries, legislators are currently forming meaningful regulations to prevent the duplicitous use of this practice. Voluntary euthanasia is still prohibited in all 3 countries [39].

Development of Palliative Care Services

The results of the regulatory changes and public health strategies are regularly reported in the *EAPC Atlas of Palliative Care in Europe* [27-29]. In summary, the public accessibility of stationary, ambulatory, and voluntary services increased considerably in all 3 countries. Palliative care services per 100,000 citizens doubled from 2007 to 2019 from 0.6 to 2.2 in Austria, 0.4 to 1.1 in Germany, and 0.5 to 1.1 in Switzerland [27-29]. The results are summarized in Table 2.

Table 2. Development of palliative care services between 2007 and 2019 for each country.

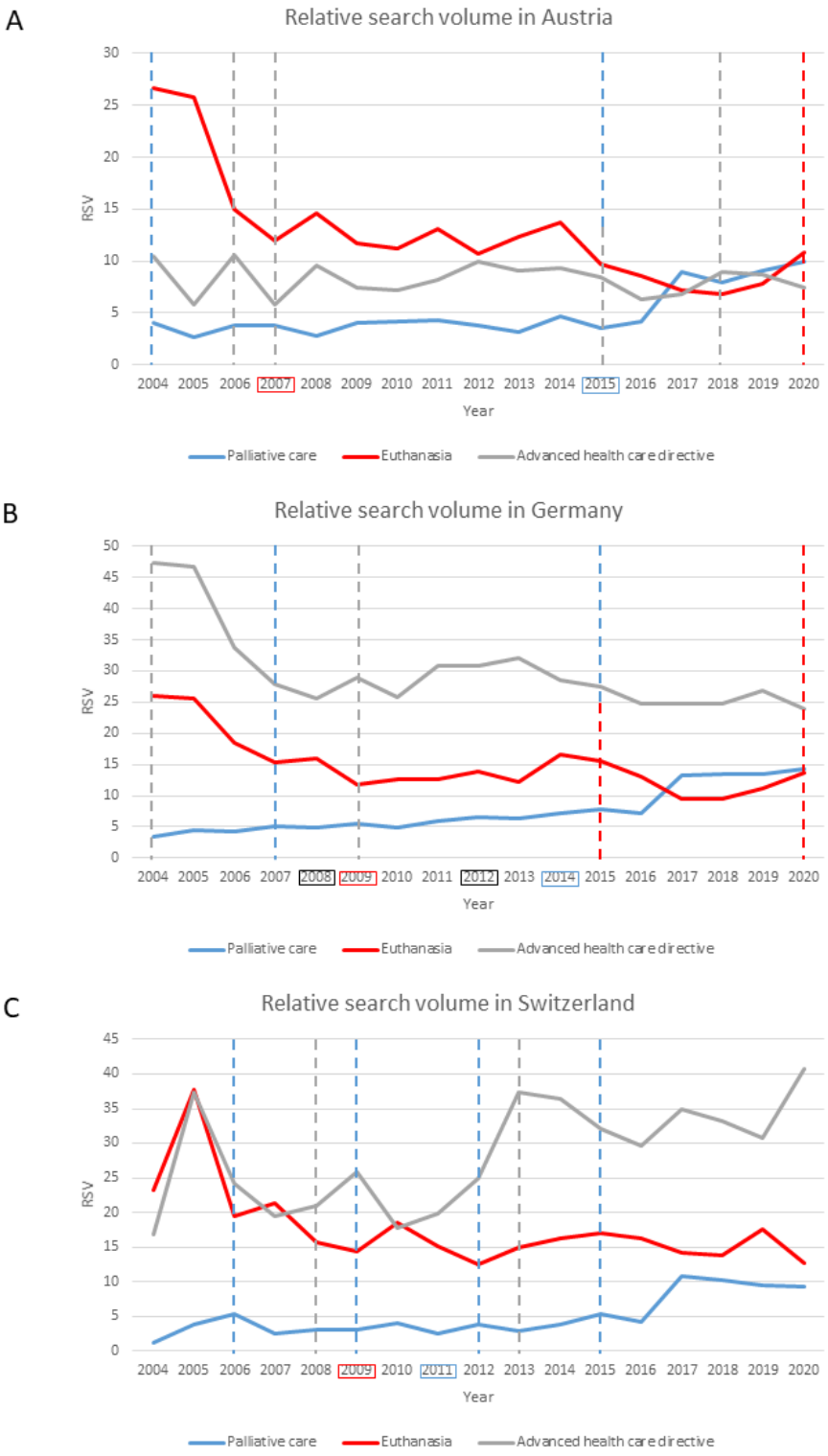
Country	Palliative care services per 100,000 citizens		
	2007	2013	2019
Austria	0.6	1.4	2.2
Germany	0.4	0.7	1.1
Switzerland	0.5	0.8	1.1

Internet Search Behavior

General Findings

The search trends for each keyword and country are shown in Figure 1. In general, the search activity for palliative care increased over the studied period, whereas interest in euthanasia decreased in all 3 countries. The topic of advance health care directives showed mixed results. The greatest change in the average annual relative search volume for palliative care between 2004 and 2020 occurred in Switzerland, where there was an increase of 13.76% (95% CI 1.43-27.60; P=.03), followed by Germany with 9.15% (95% CI 6.12-12.27; P<.001) and then Austria (7.8%, 95% CI 2.96-12.88; P<.001).

Figure 1. Trends in relative search volume for each country. Vertical dashed lines indicate passed or revised public health strategies or laws concerning palliative care, euthanasia, or advance health care directives. The colored frames indicate the years of significant trend changes according to the joinpoint analysis. RSV: relative search volume.



Interest in euthanasia decreased in all 3 countries between 2004 and 2020. However, we found a significant decrease only in Switzerland, with a change of 4.15% (95% CI −8.09 to −0.04; *P*=.047) per year. For Germany and Austria, the relative search volume decreased by 4.13% (95% CI −11.09 to 3.38; *P*=.27) and 6.07% (95% CI −12.78 to 1.15; *P*=.10), respectively.

We found different results for the search behavior of advance health care directives in each country. Although the relative search volume remained at a consistent level in Austria, with an insignificant decrease of 0.16% (95% CI −2.27 to 1.99; *P*=.87) across the study period, the volume significantly decreased by 2.85% (95% CI −4.32 to −1.35; *P*<.001) in

Germany and significantly increased by 3.66% (95% CI 1.18-6.21; *P*<.001) in Switzerland.

Austria

The joinpoint analysis revealed significant trend changes in the search terms' trajectories, which are presented in Table 3. We recorded joinpoints for the interest topics of palliative care and euthanasia; however, none were found for advance health care directives. In 2006 and 2007, two new laws regulating advance health care directives and the appointment of health care proxies were enacted by the Austrian parliament. Following these enactments, the relative search volume for euthanasia dropped by 24.48% (95% CI −39.20 to −6.19). Afterward, 2 periods with an insignificant but continuous reduction occurred, although this trend was more prominent from 2014 to 2018 than from 2007 to 2014 (Table 3). Toward the end of the study period in 2020, we observed a positive trend, although the difference was not significant (Table 3). In 2015, the parliamentary commission of inquiry regarding *dignity at the end of life* adopted 51 resolutions to promote and fund palliative care, followed by a 22.69% (95% CI 7.05-40.61; *P*=.01) increase in the search interest in palliative care. Until then, there had been a positive but small slope (Table 3). Although it fluctuated over the whole study period, the search trend for advance health care directives remained at a steady level without significant trend changes between 2004 and 2020 (Table 3).

Table 3. Trend periods and annual percentage of change in the search volume for each term in Austria.

Trend periods	Annual change (95% CI; %)	*P* value
Palliative care		
2004-2015	1.65 (−2.44 to 5.91)	.40
2015-2020	22.69 (7.05 to 40.61)	.01
Euthanasia		
2004-2007	−24.48 (−39.20 to −6.19)	.02
2007-2014	−0.42 (−7.46 to 7.15)	.89
2014-2018	−15.02 (−31.59 to 5.55)	.12
2018-2020	29.74 (−15.91 to 100.17)	.19
Advance health care directive		
2004-2020	−0.16 (−2.27 to 1.99)	.87

Germany

The search trends in Austria and Germany were similar; however, the latter had joinpoints in all 3 keywords. The results are summarized in Table 3. With the legal regulation of advance health care directives in 2009, the relative search volume for euthanasia dropped by 14.39% (95% CI −21.80 to −7.50; *P*<.001). After a slight positive trend, a negative slope appeared in 2015, followed by a positive trend in 2018. However, none of these changes were significant (Table 4). Within the relative search volume trend for palliative care, one joinpoint was determined in 2014, dividing the overall period into 2 segments of significant positive slopes. Although in the first period the interest grew continuously by 6.12% (95% CI 2.93-9.40; *P*<.001), a sudden rise of 14.39% (95% CI 7.06-22.23; *P*<.001) appeared in 2014 after the enactment of a law regulating hospice and palliative care accessibility. Interestingly, the search trend for advance health care directives followed a pattern comparable with the trend for euthanasia, with a drop of 15.42% (95% CI −22.03 to −8.25; *P*<.001) in 2008. Afterward, the search activity grew between 2008 and 2012 but became negative thereafter (Table 4).

Table 4. Trend periods and annual percentage of change in the search term for each term in Germany.

Trend periods	Annual change (95% CI; %)	*P* value
Palliative care		
2004-2014	6.12 (2.93 to 9.40)	<.001
2014-2020	14.39 (7.06 to 22.23)	<.001
Euthanasia		
2004-2009	−14.95 (−21.80 to −7.50)	<.001
2009-2015	4.13 (−4.26 to 13.26)	.28
2015-2018	−16.83 (−42.88 to 21.08)	.27
2018-2020	24.97 (−14.17 to 81.95)	.20
Advance health care directive		
2004-2008	−15.42 (−22.03 to −8.25)	<.001
2008-2012	5.19 (−7.49 to 19.62)	.39
2012-2020	−3.30 (−5.98 to −0.55)	.01

Switzerland

In Switzerland, similar trends in search activity, presented in Table 5, were found. The adult guardianship law was amended in 2008, and it included the regulation of advance health care directives. This amendment was followed by a decrease of 11.75% (95% CI −22.09 to −0.04; *P*=.049) in the relative volume of searches for euthanasia between 2004 and 2009. In the following years, the trend remained stable and slightly negative (Table 5). However, this result is likely biased by the extraordinary peak in the relative search volume in 2005. At that time, the broad public debate about euthanasia and *suicide tourism* coincided with a higher search activity [40].

Table 5. Trend periods and annual percentage of change in the search volume for each term in Switzerland.

Trend periods	Annual change (95% CI; %)	*P* value
Palliative care		
2004-2006	73.35 (−21.41 to 282.38)	.15
2006-2011	−9.44 (−29.48 to 16.30)	.39
2011-2020	17.59 (9.40 to 26.40)	<.001
Euthanasia		
2004-2009	−11.75 (−22.09 to −0.04)	.049
2009-2020	−0.49 (−4.15 to 3.32)	.78
Advance health care directive		
2004-2020	3.66 (1.18 to 6.21)	<.001

Nevertheless, after 2009 (the calculated joinpoint), no likewise peak was observed, and the overall trend followed a negative slope (Table 5). For palliative care, the search activity increased significantly by 17.59% (95% CI 9.40-26.40; *P*<.001) from 2011 onward after the government proposed its first national strategy for promoting the implementation of palliative care. Interest in advance health care directives increased significantly by 3.66% (95% CI 1.18-6.21; *P*<.001) across the study period.

Discussion

Principal Findings

Our study results show that between 2004 and 2020, substantial legal and public health efforts were made to promote patients' autonomy in regulating end-of-life decisions and develop palliative care services in Austria, Germany, and Switzerland, as summarized in Table 1. According to our analysis, a reduction in search activity for topics related to euthanasia ensued. At the same time, interest in palliative care in all 3 countries increased. Our results support previous findings regarding why individuals consider or eventually choose voluntary death. The fear of losing control over medical decisions and being helpless against a system capable of unnecessarily prolonging a life of suffering puts individuals under pressure to take action themselves [41]. Euthanasia, assisted suicide, and advance health care directives grant individuals a high level of autonomy and self-control, especially at the end of their life. Precise regulations for advance health care directives may have provided a viable alternative option to assisted suicide. For example, the provision of alternatives decreased interest in euthanasia, according to our results. Of note is the significant increase in the search interest in advance health care directives in Switzerland, even though

assisted suicide is available. Specifically, we found significant trend changes that aligned with specific governmental actions.

We found that public awareness of palliative care increased with the realization of public health strategies and the enactment of laws to regulate funding, education, accessibility, and insurance coverage related to palliative care. A previous study using Google Trends data from the United States showed similar results, with increased public awareness for palliative care from January 2005 to December 2015 [21]. The authors also suggest a correlation between this increase in search queries and changes in insurance providers' policies promoting advance care planning and the growth in nationwide palliative care service provision [21]. This result underlines the importance of government efforts to strengthen public awareness and promote access to palliative care services for those who need them. It is important to mention that interest in palliative care also increased in Switzerland, a country with established practices for assisted suicide. However, the government actions concerning palliative care did not align with the trend changes in the search activity for euthanasia, suggesting that there were other factors with a more significant impact on the decreased interest in assisted suicide and death on demand, such as legal regulations concerning the promotion of autonomy at the end of life. The embedment of advance health care directives and the legal appointment of health care proxies into a legal framework was followed by a significant drop in search queries for terms related to euthanasia and assisted suicide.

Before legal regulations were put in place, advance health care directives were a gray area and were theoretically not legally binding, allowing them to be overruled by caregivers and medical staff without facing consequences [42]. The formulation of a sound legal framework eliminated this uncertainty for patients and has potentially led to a higher sense of autonomy, which is in a complex relationship with perceived dignity and personal identity [10]. The desire to maintain a sense of control and autonomy also includes the desire to control different aspects of one's dying process but not necessarily wishing to hasten death actively [43]. However, situations threatening autonomy may be experienced as an undermining of one's dignity, evoking existential fears of patients that may produce a sense of urgency to take self-action [43]. On the other hand, promoting autonomy with the provision of advance care planning leads to higher emotional and mental well-being [44] and quality of life [45] of patients with progressive life-limiting diseases. It appears that the legal regulations of advance health care directives secured the wish to maintain control over the dying process and led to the decreased necessity of seeking information about other opportunities, such as traveling to other countries for assisted suicide.

The significant increase in popularity of search terms related to advance health care directives in Switzerland, the only studied country with established legal practices of assisted suicide, supports the hypothesis that they are a viable alternative option. However, this trend showed different results in the other 2 countries compared with that in Switzerland. At the beginning of the study period, in Germany, the high search volume might have resulted from a high degree of uncertainty about the legal options available, leading to increased search activity to obtain information about the requirements for advance health care directives. Formulating clear regulations could have led to a drop in the queries as it was easier to obtain precise details about such practices. In Austria, the search volume for advance health care directives did not change considerably during the study period. This raises the question of which factors may have contributed to this trend. Although the new legal framework provides clear definitions for legally binding directives and grants a high degree of self-determination through its binding character, Austrians face a cost barrier of up to €500 (US $584.50) for the certifying notary [46]. This is approximately one-third of the average middle-class salary and hinders low-income Austrians from accessing this option [46]. In practice, patients use legally binding directives less frequently than nonlegally binding directives, which require neither notarial certification nor informed consent or discussion with a physician [47]. Therefore, patients may not seek as much information about how to arrange a legally binding directive or may discuss potentially confusing options primarily with their physicians.

Given the current developments in Germany and Austria, which tend to legalize assisted suicide practices, palliative care services could be involved in the decision-making process to find the best solution for each patient. Professional palliative caregivers are experienced in discussing sensitive topics such as end-of-life decisions and can inform patients about the different options in an evidence-based manner. Recent literature reviews have highlighted the benefits of early palliative care implementation in oncology, including an increase in the number of stipulated advance health care directives [48]. Therefore, palliative care may serve as a filter for patients with a genuinely intrinsic wish to hasten death actively and guide them in a meaningful and supportive way [8]. Evaluating the underlying reasoning, fears, and perhaps misunderstandings of individuals about end-of-life care is crucial in preventing the misuse of euthanasia and assisted suicide [49-52]. With the legalization of assisted suicide, it would be reasonable to simultaneously increase the accessibility of palliative care and extend its early integration into the treatment of terminally ill patients.

Limitations

Some limitations of this study need to be discussed. First, this is a correlative retrospective study, and hence, it does not allow for definite conclusions about causal relationships between the enactment of policies and the investigated search trends. However, we tried to minimize the potential bias by comparing the trends in 3 comparable countries and were able to show similar results, even in Switzerland, which has notably liberal regulations concerning the practice of assisted suicide.

Second, Google Trends does not provide absolute numbers and therefore lacks transparency to some extent. Researchers can rely only on the process description that Google publishes on its webpage concerning the handling of its data [19]. Of note are the preventive measures taken against bias by limiting the number of queries per user within a specific time. With this, active manipulation and repeated search queries (eg, during a literature search) did not influence the overall search volume and therefore represented the search activity within a region more accurately.

Third, we used search topics instead of definite search terms to limit the chances of choosing a less commonly used search term. However, this option may have also included nonspecific search terms. The most popular search terms are still included in the output of a trend query, and in our case, these queries predominantly showed specific terms related to the research question in this study. The output (in German) for each country and each search topic are presented in Tables S1-S3 in Multimedia Appendix 1.

Fourth, the trend changes might have been influenced by factors other than legal regulations and health care policies. For example, through the constant improvement of disease-specific and supportive therapies over the past two decades, the disease burden of many life-limiting diseases decreased while simultaneously focusing on enhancing patients' quality of life [53].

Implications

Our study provides novel and essential information about the demands and interests of the residents of Austria, Germany, and Switzerland concerning palliative care, euthanasia, assisted suicide, and end-of-life decisions. In particular, policy makers may use these findings to target potential weaknesses within current regulations (eg, the cost barrier for advance health care directives in Austria). Governmental promotion of the early implementation of palliative care into life-threatening disease trajectories may be another meaningful action. Experts in palliative care can sensitize patients to arranging advance health care directives, inform them about the different options at the end of life, and resolve misunderstandings.

Conclusions

Our study shows, for the first time, that governmental regulations, precise legal definitions, and broader access to advance health care directives might reduce interest in euthanasia, suggesting that loss of control is one of the predominant factors involved in the desire to hasten death. Early palliative care can guide patients through the necessary decision-making processes from the point of the diagnosis of a life-threatening disease to their death and provide advance care planning. After all, it is and has always been a core principle of palliative care to promote and value the autonomy of patients.

Authors' Contributions

MH contributed to the idea formulation, study design, data analysis and interpretation, and writing of the manuscript. DJK and GH contributed to the study design, data interpretation, and writing of the manuscript. EA, CF, WS, EKM, and PJJ contributed to the data interpretation and literature review. MP contributed to the data interpretation, study design, and language editing. All authors reviewed the final manuscript.

Conflicts of Interest

PJJ has had a consulting or advisory role, received honoraria, research funding, travel or accommodation expenses from Abbvie, Bayer, Boehringer, Novartis, Pfizer, Servier, Roche, Bristol Myers Squibb, and Celgene.

Multimedia Appendix 1
Most popular search terms.
[DOC File , 183 KB - jmir_v23i9e28635_app1.doc]

References

1. Horn R. The right to a self-determined death as expression of the right to freedom of personal development: the German Constitutional Court takes a clear stand on assisted suicide. J Med Ethics 2020 Jun;46(6):416-417. [doi: 10.1136/medethics-2020-106197] [Medline: 32220872]
2. Möller HJ. The ongoing discussion on termination of life on request. A review from a German/European perspective. Int J Psychiatry Clin Pract 2021 Mar;25(1):2-18. [doi: 10.1080/13651501.2020.1797097] [Medline: 32729770]
3. Kremeike K, Pralong A, Boström K, Bausewein C, Simon ST, Lindner R, Working Group on Desire to Die of the German Palliative Care Guideline. 'Desire to Die' in palliative care patients-legal framework and recommendations of the national evidence-based guideline on palliative care in Germany. Ann Palliat Med 2021 Mar;10(3):3594-3610 [FREE Full text] [doi: 10.21037/apm-20-381] [Medline: 33440974]
4. Hyde R. Germany overturns ban on assisted suicide. Lancet 2020 Mar 07;395(10226):774. [doi: 10.1016/S0140-6736(20)30533-X] [Medline: 32145781]
5. Es gilt das gesprochene Wort!. Verfassungsgerichtshof Österreich. URL: https://www.vfgh.gv.at/downloads/G_139.2019_Verkuendung.pdf [accessed 2021-08-31]
6. Radbruch L, Leget C, Bahr P, Müller-Busch C, Ellershaw J, de Conno F, Board Members of EAPC. Euthanasia and physician-assisted suicide: a white paper from the European Association for Palliative Care. Palliat Med 2016 Feb;30(2):104-116. [doi: 10.1177/0269216315616524] [Medline: 26586603]
7. Sepúlveda C, Marlin A, Yoshida T, Ullrich A. Palliative care: the World Health Organization's global perspective. J Pain Symptom Manage 2002 Aug;24(2):91-96 [FREE Full text] [doi: 10.1016/s0885-3924(02)00440-2] [Medline: 12231124]

8. Dierickx S, Deliens L, Cohen J, Chambaere K. Involvement of palliative care in euthanasia practice in a context of legalized euthanasia: a population-based mortality follow-back study. Palliat Med 2018 Jan;32(1):114-122 [FREE Full text] [doi: 10.1177/0269216317727158] [Medline: 28849727]
9. Gordijn B, Janssens R. The prevention of euthanasia through palliative care: new developments in The Netherlands. Patient Educ Couns 2000 Aug;41(1):35-46. [doi: 10.1016/s0738-3991(00)00113-0] [Medline: 10900365]
10. Rodríguez-Prat A, Monforte-Royo C, Porta-Sales J, Escribano X, Balaguer A. Patient perspectives of dignity, autonomy and control at the end of life: systematic review and meta-ethnography. PLoS One 2016 Mar 24;11(3):e0151435 [FREE Full text] [doi: 10.1371/journal.pone.0151435] [Medline: 27010323]
11. Lindner R, Foerster R, von Renteln-Kruse W. Physical distress and relationship problems: exploring the psychosocial and intrapsychic world of suicidal geriatric patients. Z Gerontol Geriatr 2014 Aug;47(6):502-507. [doi: 10.1007/s00391-013-0563-z] [Medline: 24202298]
12. Stolz E, Burkert N, Großschädl F, Rásky É, Stronegger WJ, Freidl W. Determinants of public attitudes towards euthanasia in adults and physician-assisted death in neonates in Austria: a national survey. PLoS One 2015 Apr 23;10(4):e0124320 [FREE Full text] [doi: 10.1371/journal.pone.0124320] [Medline: 25906265]
13. Stolz E, Großschädl F, Mayerl H, Rásky É, Freidl W. Determinants of acceptance of end-of-life interventions: a comparison between withdrawing life-prolonging treatment and euthanasia in Austria. BMC Med Ethics 2015 Dec 01;16(1):81 [FREE Full text] [doi: 10.1186/s12910-015-0076-y] [Medline: 26625908]
14. Stolz E, Mayerl H, Gasser-Steiner P, Freidl W. Attitudes towards assisted suicide and euthanasia among care-dependent older adults (50+) in Austria: the role of socio-demographics, religiosity, physical illness, psychological distress, and social isolation. BMC Med Ethics 2017 Dec 07;18(1):71 [FREE Full text] [doi: 10.1186/s12910-017-0233-6] [Medline: 29212490]
15. Stolz E, Mayerl H, Waxenegger A, Rásky É, Freidl W. Attitudes towards end-of-life decisions in case of long-term care dependency: a survey among the older population in Austria. J Med Ethics 2017 Jun;43(6):413-416. [doi: 10.1136/medethics-2016-103731] [Medline: 28235885]
16. Anneser J, Jox RJ, Thurn T, Borasio GD. Physician-assisted suicide, euthanasia and palliative sedation: attitudes and knowledge of medical students. GMS J Med Educ 2016 Feb 15;33(1):Doc11 [FREE Full text] [doi: 10.3205/zma001010] [Medline: 26958648]
17. Jahn-Kuch D, Domke A, Bitsche S, Stöger H, Avian A, Jeitler K, et al. End-of-life decision making by Austrian physicians - a cross-sectional study. BMC Palliat Care 2020 Jan 04;19(1):4 [FREE Full text] [doi: 10.1186/s12904-019-0509-3] [Medline: 31901225]
18. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. PLoS One 2014 Oct 22;9(10):e109583 [FREE Full text] [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]
19. Google Trends. URL: https://trends.google.com/trends/?geo=IN [accessed 2021-08-13]
20. Schootman M, Toor A, Cavazos-Rehg P, Jeffe DB, McQueen A, Eberth J, et al. The utility of Google Trends data to examine interest in cancer screening. BMJ Open 2015 Jun 08;5(6):e006678 [FREE Full text] [doi: 10.1136/bmjopen-2014-006678] [Medline: 26056120]
21. McLean S, Lennon P, Glare P. Internet search query analysis can be used to demonstrate the rapidly increasing public awareness of palliative care in the USA. BMJ Support Palliat Care 2019 Mar;9(1):40-44. [doi: 10.1136/bmjspcare-2016-001171] [Medline: 28130324]
22. Lane T, Ramadurai D, Simonetti J. Public awareness and perceptions of palliative and comfort care. Am J Med 2019 Feb;132(2):129-131. [doi: 10.1016/j.amjmed.2018.07.032] [Medline: 30076826]
23. Mavragani A, Ochoa G, Tsagarakis KP. Assessing the methods, tools, and statistical approaches in Google Trends research: systematic review. J Med Internet Res 2018 Nov 06;20(11):e270 [FREE Full text] [doi: 10.2196/jmir.9366] [Medline: 30401664]
24. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
25. Seifter A, Schwarzwalder A, Geis K, Aucott J. The utility of "Google Trends" for epidemiological research: Lyme disease as an example. Geospat Health 2010 May;4(2):135-137. [doi: 10.4081/gh.2010.195] [Medline: 20503183]
26. Flash Eurobarometer 404 (European citizens´ digital health literacy). GESIS Data Archive. 2015. URL: https://search.gesis.org/research_data/ZA5948 [accessed 2021-08-13]
27. Arias-Casais N, Garralda E, Rhee J, de Lima L, Pons J, Clark D, et al. EAPC Atlas of Palliative Care in Europe 2019. Vilvoorde: EAPC vzw Press; May 2019.
28. Centeno C, Clark D, Lynch T, Rocafort J, Flores L, Greenwood A, et al. EAPC Atlas of Palliative Care in Europe. Houston: IAHPC Press; 2007.
29. Centeno C, Rocafort J, Lynch C, Clark D, Donea O. EAPC Atlas of Palliative Care in Europe 2013. Milano: European Association for Palliative Care; Jun 2013.
30. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. Stat Med 2000 Feb 15;19(3):335-351. [doi: 10.1002/(sici)1097-0258(20000215)19:3<335::aid-sim336>3.0.co;2-z] [Medline: 10649300]

31. Lin R, Cheng Y, Jiang Y. Exploring public awareness of overwork prevention with big data from Google Trends: retrospective analysis. J Med Internet Res 2020 Jun 26;22(6):e18181 [FREE Full text] [doi: 10.2196/18181] [Medline: 32589160]
32. Centeno C, Clark D, Lynch T, Racafort J, Praill D, De Lima L, EAPC Task Force. Facts and indicators on palliative care development in 52 countries of the WHO European region: results of an EAPC Task Force. Palliat Med 2007 Sep;21(6):463-471. [doi: 10.1177/0269216307081942] [Medline: 17846085]
33. Lang A. The good death and the institutionalisation of dying: an interpretive analysis of the Austrian discourse. Soc Sci Med 2020 Jan;245:112671. [doi: 10.1016/j.socscimed.2019.112671] [Medline: 31734481]
34. Melching H. [New legal regulations for palliative care with implications for politics and practice]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2017 Jan;60(1):4-10. [doi: 10.1007/s00103-016-2480-y] [Medline: 27995269]
35. Framework for palliative care in Switzerland. Federal Office of Public Health FOPH, Swiss Conference of the Cantonal Ministers of Public Health. URL: https://www.bag.admin.ch/bag/de/home/strategie-und-politik/nationale-gesundheitsstrategien/strategie-palliative-care.html [accessed 2021-08-13]
36. Körtner HJ. [Patient advance directives in Austria--ethical discussions]. Bull Soc Sci Med Grand Duche Luxemb 2008(3):279-287. [Medline: 19069726]
37. Evans N, Bausewein C, Meñaca A, Andrew EV, Higginson IJ, Harding R, project PRISMA. A critical review of advance directives in Germany: attitudes, use and healthcare professionals' compliance. Patient Educ Couns 2012 Jun;87(3):277-288. [doi: 10.1016/j.pec.2011.10.004] [Medline: 22115975]
38. Haesen S, Shaw D. Directing citizens to create advance directives. Swiss Med Wkly 2018 May 16;148:w14628 [FREE Full text] [doi: 10.4414/smw.2018.14628] [Medline: 29767827]
39. Grosse C, Grosse A. Assisted suicide: models of legal regulation in selected European countries and the case law of the European Court of Human Rights. Med Sci Law 2015 Oct;55(4):246-258. [doi: 10.1177/0025802414540636] [Medline: 25143341]
40. Shondell Miller D, Gonzalez C. When death is the destination: the business of death tourism – despite legal and social implications. Int J Culture Tourism Hosp Res 2013 Aug 02;7(3):293-306. [doi: 10.1108/ijcthr-05-2012-0042]
41. Chapple A, Ziebland S, McPherson A, Herxheimer A. What people close to death say about euthanasia and assisted suicide: a qualitative study. J Med Ethics 2006 Dec;32(12):706-710 [FREE Full text] [doi: 10.1136/jme.2006.015883] [Medline: 17145910]
42. Wiesing U, Jox RJ, Hessler H, Borasio GD. A new law on advance directives in Germany. J Med Ethics 2010 Dec;36(12):779-783. [doi: 10.1136/jme.2010.036376] [Medline: 21112938]
43. Monforte-Royo C, Villavicencio-Chávez C, Tomás-Sábado J, Mahtani-Chugani V, Balaguer A. What lies behind the wish to hasten death? A systematic review and meta-ethnography from the perspective of patients. PLoS One 2012;7(5):e37117 [FREE Full text] [doi: 10.1371/journal.pone.0037117] [Medline: 22606338]
44. Dyar S, Lesperance M, Shannon R, Sloan J, Colon-Otero G. A nurse practitioner directed intervention improves the quality of life of patients with metastatic cancer: results of a randomized pilot study. J Palliat Med 2012 Aug;15(8):890-895 [FREE Full text] [doi: 10.1089/jpm.2012.0014] [Medline: 22559906]
45. Bakitas M, Lyons KD, Hegel MT, Balan S, Brokaw FC, Seville J, et al. Effects of a palliative care intervention on clinical outcomes in patients with advanced cancer: the Project ENABLE II randomized controlled trial. J Am Med Assoc 2009 Aug 19;302(7):741-749 [FREE Full text] [doi: 10.1001/jama.2009.1198] [Medline: 19690306]
46. Schaden E, Herczeg P, Hacker S, Schopper A, Krenn CG. The role of advance directives in end-of-life decisions in Austria: survey of intensive care physicians. BMC Med Ethics 2010 Oct 21;11:19 [FREE Full text] [doi: 10.1186/1472-6939-11-19] [Medline: 20964852]
47. Köstenberger M, Diegelmann S, Terlutter R, Bidmon S, Neuwersch S, Likar R. Advance directives in Austrian intensive care units: an analysis of prevalence and barriers. Resusc Plus 2020 Jun 28;3:100014 [FREE Full text] [doi: 10.1016/j.resplu.2020.100014] [Medline: 34223298]
48. Hui D, Hannon BL, Zimmermann C, Bruera E. Improving patient and caregiver outcomes in oncology: team-based, timely, and targeted palliative care. CA Cancer J Clin 2018 Sep;68(5):356-376 [FREE Full text] [doi: 10.3322/caac.21490] [Medline: 30277572]
49. Breitbart W, Pessin H, Rosenfeld B, Applebaum AJ, Lichtenthal WG, Li Y, et al. Individual meaning-centered psychotherapy for the treatment of psychological and existential distress: a randomized controlled trial in patients with advanced cancer. Cancer 2018 Aug 01;124(15):3231-3239 [FREE Full text] [doi: 10.1002/cncr.31539] [Medline: 29757459]
50. Breitbart W, Rosenfeld B, Gibson C, Pessin H, Poppito S, Nelson C, et al. Meaning-centered group psychotherapy for patients with advanced cancer: a pilot randomized controlled trial. Psychooncology 2010 Jan;19(1):21-28 [FREE Full text] [doi: 10.1002/pon.1556] [Medline: 19274623]
51. Breitbart W, Rosenfeld B, Pessin H, Applebaum A, Kulikowski JS, Lichtenthal WG. Meaning-centered group psychotherapy: an effective intervention for improving psychological well-being in patients with advanced cancer. J Clin Oncol 2015 Mar 01;33(7):749-754 [FREE Full text] [doi: 10.1200/JCO.2014.57.2198] [Medline: 25646186]
52. Julião M, Oliveira F, Nunes B, Carneiro AV, Barbosa A. Effect of dignity therapy on end-of-life psychological distress in terminally ill Portuguese patients: a randomized controlled trial. Palliat Support Care 2017 Dec;15(6):628-637. [doi: 10.1017/S1478951516001140] [Medline: 28166861]

53. Santos JV, Souza J, Valente J, Alonso V, Ramalho A, Viana J, et al. The state of health in the European Union (EU-28) in 2017: an analysis of the burden of diseases and injuries. Eur J Public Health 2020 Jun 01;30(3):573-578. [doi: 10.1093/eurpub/ckz203] [Medline: 31821479]

Abbreviations

EAPC: European Association of Palliative Care

Edited by R Kukafka; submitted 11.03.21; peer-reviewed by K Kremeike, S Kardes; comments to author 27.04.21; revised version received 25.05.21; accepted 05.07.21; published 20.09.21.

Please cite as:
Huemer M, Jahn-Kuch D, Hofmann G, Andritsch E, Farkas C, Schaupp W, Masel EK, Jost PJ, Pichler M
Trends and Patterns in the Public Awareness of Palliative Care, Euthanasia, and End-of-Life Decisions in 3 Central European Countries Using Big Data Analysis From Google: Retrospective Analysis
J Med Internet Res 2021;23(9):e28635
URL: https://www.jmir.org/2021/9/e28635
doi:10.2196/28635
PMID:34542419

Original Paper

Predicting Norovirus in the United States Using Google Trends: Infodemiology Study

Kai Yuan[1*], MD, PhD; Guangrui Huang[1*], PhD; Lepeng Wang[2*], MD; Ting Wang[1], MD; Wenbin Liu[1], MD; Haixu Jiang[1], MD; Albert C Yang[3,4,5], MD, PhD

[1]School of Life Sciences, Beijing University of Chinese Medicine, Beijing, China

[2]School of Humanities, Beijing University of Chinese Medicine, Beijing, China

[3]Digital Medicine Center, National Yang Ming Chiao Tung University, Taiwan, Republic of China

[4]Department of Medical Research, Taipei Veterans General Hospital, Taiwan, Republic of China

[5]Division of Interdisciplinary Medicine and Biotechnology, Beth Israel Deaconess Medical Center/Harvard Medical School, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Albert C Yang, MD, PhD
Digital Medicine Center
National Yang Ming Chiao Tung University
155 Li-Nong Street, Section 2
Peitou District
Taipei
Republic of China
Phone: 886 28267995
Email: accyang@gmail.com

Abstract

Background: Norovirus is a contagious disease. The transmission of norovirus spreads quickly and easily in various ways. Because effective methods to prevent or treat norovirus have not been discovered, it is important to rapidly recognize and report norovirus outbreaks in the early phase. Internet search has been a useful method for people to access information immediately. With the precise record of internet search trends, internet search has been a useful tool to manifest infectious disease outbreaks.

Objective: In this study, we tried to discover the correlation between internet search terms and norovirus infection.

Methods: The internet search trend data of norovirus were obtained from Google Trends. We used cross-correlation analysis to discover the temporal correlation between norovirus and other terms. We also used multiple linear regression with the stepwise method to recognize the most important predictors of internet search trends and norovirus. In addition, we evaluated the temporal correlation between actual norovirus cases and internet search terms in New York, California, and the United States as a whole.

Results: Some Google search terms such as gastroenteritis, watery diarrhea, and stomach bug coincided with norovirus Google Trends. Some Google search terms such as contagious, travel, and party presented earlier than norovirus Google Trends. Some Google search terms such as dehydration, bar, and coronavirus presented several months later than norovirus Google Trends. We found that fever, gastroenteritis, poison, cruise, wedding, and watery diarrhea were important factors correlated with norovirus Google Trends. In actual norovirus cases from New York, California, and the United States as a whole, some Google search terms presented with, earlier, or later than actual norovirus cases.

Conclusions: Our study provides novel strategy-based internet search evidence regarding the epidemiology of norovirus.

(J Med Internet Res 2021;23(9):e24554) doi:10.2196/24554

KEYWORDS

norovirus; Google Trends; correlation; outbreak; predictors

Introduction

Norovirus, also named stomach bug or winter vomiting bug, is a kind of contagious virus leading to gastroenteritis [1]. It is responsible for almost 20% of acute gastroenteritis cases worldwide [2]. Especially in cruise ships, norovirus is the most common cause of gastrointestinal illness [3]. The symptoms of norovirus include diarrhea, nausea, and vomiting, and it can be

transmitted by contaminated water, contaminated food, and person-to-person contact [4,5]. Polymerase chain reaction and enzyme-linked immunosorbent assay are useful methods to diagnose norovirus infection [6,7]. Thus far, there have been no specific drugs to treat norovirus infection. In the prevention of norovirus, hand washing and disinfectants are important and effective methods. In addition, avoiding contaminated water and contaminated food in restaurants are useful methods to prevent norovirus.

It is important to rapidly recognize and report norovirus infection. In the United States, the US Centers for Disease Control and Prevention (CDC) report and monitor the outbreak of gastroenteritis with an internet search system. In addition, other surveillance systems, such as the Foodborne Diseases Active Surveillance Network, can be used to report the trends of foodborne disease outbreaks [8]. However, these surveillance and reporting systems have a lag of 1 to 2 weeks after an outbreak due to the reporting and verification procedure. Internet search has been a useful method for people to access information immediately. Infodemiology is an emerging field based on public attention, knowledge, opinion, etc [9]. Data from infodemiology could be acquired and analyzed to enhance public health [10] and predict outbreaks of disease such as influenza [11]. In addition, infodemiology could also be used to monitor foodborne disease, gastroenteritis, dengue, etc [12]. Apart from Google search, search engines from other platforms including Facebook and Twitter can benefit infoveillance [13]. Yang et al [14-16] revealed the correlation between internet search trends and suicide incidence and depression. In addition, internet searches have been used to monitor infectious diseases, including influenza and Lyme disease [17,18]. Mavragani et al [19] analyzed 104 published papers published from 2006 to 2016 related to Google Trends and found that the monitoring of online queries can provide insight into human behavior. In addition, Nuti et al [20] analyzed 70 papers published from 2009 to 2013 and discovered the topic domains related to Google Trends were general population behavior, infectious disease, noncommunicable diseases, mental health, and substance use.

In this study, we discovered a correlation between internet search terms and norovirus. The search trend of Google Insight was extracted and used in our study. Certain keywords in the Google search trend included “diarrhea,” “vomiting,” “fever,” etc. These keywords could reflect the characteristics of norovirus infection. We also conducted a monthly time series analysis to discover the relationship between internet search terms and norovirus. We hoped to provide new insight to reveal the epidemic and outbreak of norovirus using internet search data.

Methods

Internet Search Approach

The internet search data relating to norovirus were obtained from Google Trends [21]. Google Trends is a useful website to discover search queries from various regions and countries. This tracking system could also provide detailed analysis of Google searches. Google Trends could provide long time series and geographical search locations. Google Trends is in widespread use to investigate medical topics. We downloaded the data of various terms in monthly granularity. The normalization of data indicated that the values vary from 0 to 100. The value 0 indicated very low search volumes that are not included in the results, while the value 100 indicated very high search volumes. The data are retrieved directly from the Google Trends Explore page in .csv format after keywords are entered and the region, period, and category are selected (see Multimedia Appendix 1) [22]. We chose geographic locations for search volumes as the states of New York and California and the United States as a whole. A time range from January 1, 2004, to December 31, 2018, was chosen. The category was limited to Health to avoid unrelated results of Google searches. We used 50 search terms to reflect a broad sense of norovirus: “norovirus,” “gastroenteritis,” “diarrhea,” “vomiting,” “dehydration,” “contaminated,” “norovirus infection,” “watery diarrhea,” “contagious,” “contaminated water,” “Norwalk virus,” “noroviruses,” “acute gastroenteritis,” “stomach flu,” “viral gastroenteritis,” “winter vomiting disease,” “stomach bug,” “travel,” “party,” “barbecue,” “cruise,” “oyster,” “bar,” “restaurant,” “wedding,” “hotel,” “motel,” “virus,” “infectious,” “outbreak,” “rotavirus,” “coronavirus,” “influenza,” “food poisoning,” “incubation period,” “fever,” “poison,” “CDC,” “vaccine,” “Chipotle,” “ship,” “hand sanitizer,” “wash hand,” “flu symptom,” “streptococcus,” “antibiotics,” “candidiasis,” “otitis media,” “skin rash,” and “coxsackie virus.” These terms were chosen to indicate the related symptoms of and risk factors for norovirus. These terms could provide positive and negative information about norovirus. The Google Trends data were exhibited on a monthly basis. No patient records or personal information were included in the search results of Google Trends. Google Trends normalizes search data to make comparisons between terms easier. Search results are normalized to the time and location of a query. Each data point is divided by the total searches of the geography and time range it represents to compare relative popularity. The resulting numbers are then scaled on a range of 0 to 100 based on a topic’s proportion to all searches on all topics [23].

Actual Norovirus Cases Acquisition

Actual norovirus cases from California, New York, and the country as a whole were obtained from the National Outbreak Reporting System (NORS) at the CDC [24,25]. We limited the etiology to “norovirus” ranging from 2004 to 2018. In addition, the illnesses in every month varied from January 2004 to December 2018. Cross-correlation analysis was conducted to reveal the relationship between actual norovirus cases and internet search terms.

Statistical Analysis

SPSS for Windows (version 20.0, IBM Corp) was used for statistical analysis. A 2-stage modeling process was conducted to evaluate the correlation between norovirus and internet search trends. Cross-correlation analysis was conducted to reveal the relationship between 2 parameters. Cross-correlation is a method to discover the correlation between 2 time series parameters contemporaneously and with various lagged values. Cross-correlation could help researchers understand the relationship between 2 parameters. In addition, cross-correlation analysis could also help researchers understand whether one

parameter was preceded by or followed another. Thus, cross-correlation could estimate the time lag of norovirus and other internet search trend terms. The results of cross-correlation were reported in our study.

We used a multiple linear regression method to recognize the most important predictors of internet search trends. Multiple linear regression was used in the second stage to analyze the internet search terms. A stepwise method in multiple linear regression was conducted to discover significant parameters correlated with norovirus. It also could help researchers evaluate the variation in monthly norovirus counts explained with significant parameters. A stepwise method could reduce the collinearity that internet search might be intercorrelated. The variance inflation factor was used to evaluate the collinearity of internet search data. Those terms with variance inflation factor >10 were excluded from the model. The summary of multiple linear regression was reported as results. A 2-tailed $P<.05$ was regarded as significant in all analyses.

Results

Basic Features Description

The results of Google Trends in the state of New York from January 2004 to December 2018 were downloaded in our study. Except for the search term “norovirus,” a total of 50 keywords from Google Trends were downloaded. Visual inspection indicated that the value of norovirus in Google Trends peaked around December to February in most years (Figure 1). The basic features of actual norovirus cases are shown in Figure 2.

Figure 1. The basic features of the norovirus Google search.

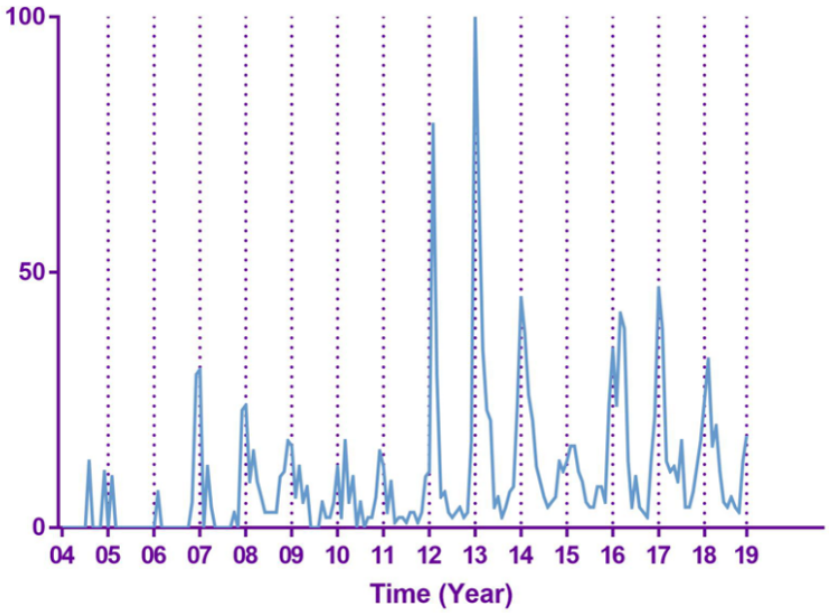
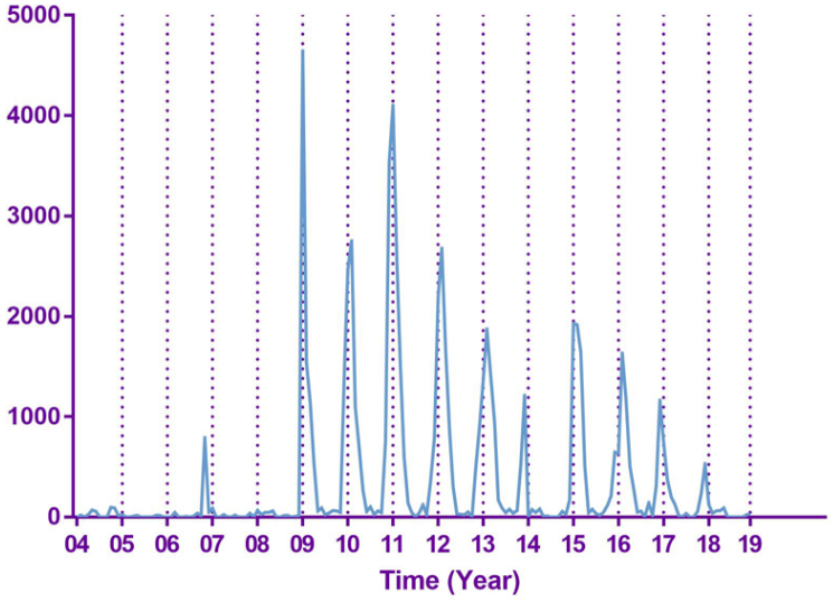


Figure 2. The basic features of actual norovirus cases.



Temporal Correlation in New York

We used cross-correlation analysis to discover the temporal correlation between norovirus and other terms. The internet search terms were included for cross-correlation analysis with norovirus. Multimedia Appendix 2 indicates the results of the cross-correlation of norovirus and other internet search terms. The gray markup shows the lead pattern for norovirus and other internet search terms. The search terms "gastroenteritis," "watery diarrhea," "stomach flu," "winter vomiting disease," "stomach bug," and "food poisoning" were coincided with norovirus in Google Trends. The search terms of "contagious," "travel," "party," "restaurant," "wedding," "hotel," "infectious," and "poison" presented earlier than norovirus Google Trends. The search terms of "dehydration," "bar," "Chipotle," and "coronavirus" presented several months later than norovirus Google Trends.

After discovering the temporal correlation between norovirus and other terms, we used multiple linear regression with the stepwise method to recognize the most important predictors of internet search trends and norovirus. The search terms that did not have a relationship with norovirus were excluded from the cross-correlation analysis. The internet search terms in Multimedia Appendix 2 were included for multiple linear regression analysis with a norovirus search. The results indicated that searches for "stomach flu," "stomach bug," "winter vomiting disease," "gastroenteritis," and "cruise" were significant predictors of norovirus search and accounted for 55% (R^2) of the variation in norovirus search (Table 1).

Table 1. Stepwise regression model of norovirus and correlated internet search trends in New York State.

Search term	Model summary		
	Standardized β	SE	*P* value
Stomach flu	0.24	0.06	<.001
Stomach bug	0.29	0.06	<.001
Winter vomiting disease	0.24	0.05	<.001
Gastroenteritis	0.13	0.05	.01
Cruise	0.16	0.07	.03

Norovirus also refers to stomach flu, stomach bug, or winter vomiting disease. We excluded these terms to identify the possible predictors of norovirus. Therefore, "stomach flu," "stomach bug," "winter vomiting disease," and "norovirus infection" were excluded in this multiple linear regression analysis. The results showed that searches for "fever," "gastroenteritis," "poison," "cruise," "wedding," "watery diarrhea," and "acute gastroenteritis" accounted for 40% (R^2) of the variation in the norovirus search (Table 2). We found that the symptoms of gastroenteritis, including vomiting and watery diarrhea, were important factors that were significantly correlated with norovirus. In addition, activities such as cruise and wedding were important in norovirus outbreak.

Table 2. Stepwise regression model of norovirus and correlated internet search trends excluded alternate names of norovirus in New York State.

Search term	Model summary		
	Standardized β	SE	*P* value
Fever	0.33	0.08	<.001
Gastroenteritis	0.19	0.06	.001
Poison	–0.16	0.05	.002
Cruise	0.32	0.10	.001
Wedding	–0.34	0.11	.002
Watery diarrhea	0.09	0.04	.02
Acute gastroenteritis	0.09	0.04	.04

Identifying Predictors of Actual Norovirus Cases

In this study, we evaluated the temporal correlation between actual norovirus cases in New York and internet search terms in New York. The data from actual norovirus cases were downloaded from the CDC. As shown in Multimedia Appendix 3, the search terms “norovirus,” “gastroenteritis,” “norovirus infection,” “contagious,” and “acute gastroenteritis” were coincided with actual norovirus cases in New York. The search terms “travel,” “party,” “barbecue,” and “cruise” presented earlier than actual norovirus cases in New York. The search terms “dehydration,” “winter vomiting disease,” “outbreak,” and “rotavirus” presented one or several months later than actual norovirus cases in New York.

Furthermore, we evaluated the temporal correlation between actual norovirus cases in California with internet search terms in California. The search terms including “barbecue,” “oyster,” “bar,” and “infectious” were presented one or several months earlier than actual norovirus cases in California. The search terms including “diarrhea,” “vomiting,” “contagious,” “contaminated water,” and “Norwalk virus” were presented one or several months later than actual norovirus cases in California. The detailed information is shown in Multimedia Appendix 4.

In addition to New York and California, we also evaluated the temporal correlation between actual norovirus cases with internet search terms within the United States. The search terms “vomiting,” “contaminated,” “norovirus infection,” and “contagious” coincided with actual norovirus cases in the United States. Search terms including “travel,” “party,” “cruise,” “restaurant,” and “wedding” presented one or several months earlier than actual norovirus cases in the United States. The search terms “gastroenteritis,” “diarrhea,” “dehydration,” and “watery diarrhea” presented one or several months later than actual norovirus cases in the United States. The detailed information is shown in Multimedia Appendix 5.

Discussion

Principal Findings

In our study, we found that norovirus was correlated with the search terms “gastroenteritis,” “diarrhea,” “vomiting,” “dehydration,” “contagious,” “stomach flu,” “stomach bug,” “restaurant,” and “food poisoning.” In the total related search terms, gastroenteritis, diarrhea, watery diarrhea, fever, vomiting, and dehydration were typical symptoms of norovirus. Stomach flu, winter vomiting disease, and stomach bug are other names of norovirus. Restaurant and bar were food service settings that occurred with norovirus outbreaks. Chipotle Mexican Grill is a specific restaurant that has experienced norovirus outbreaks in the United States. Apart from the previous search terms, streptococcus and coxsackievirus were related to norovirus. Infections with streptococcus and coxsackievirus have characteristics similar to those of norovirus, such as fever and contagious features.

Infodemiology could be discovered and analyzed in near real time. Many disease outbreaks could be predicted and monitored by infodemiology. The applications of infodemiology included queries analysis with internet search engines, people’s status updates monitor, publications related to public health monitors, etc. Patthi et al [26] used Google Trends from 2004 to 2016 to investigate the global search trends of oral problems. Their method provided an insight to analyze and compare oral disease over time on a mass population. Shimkhada et al [27] used Twitter Chat to identify barriers and responsive policy of patients with metastatic breast cancer care. Klein et al [28] conducted Twitter data to track the spread of COVID-19. Johnson et al [29] performed Google Trends to monitor sexually transmitted infections Chicago. Thus, infodemiology could provide valuable insights into health-related behavior of populations [9].

In this study, we discovered the correlation between norovirus and other search terms in Google search trends. Norovirus is a kind of infectious disease leading to diarrhea and vomiting [30]. Extremely or chronic diarrhea and vomiting can lead to dehydration in some cases of norovirus illness patients [31,32]. There are various transmission methods for norovirus [33]. First, noroviruses are transmitted to healthy people by contaminated food [34,35]. Foods such as oysters or fruits might be contaminated by water containing norovirus particles [36]. Second, noroviruses can be transmitted to healthy people by contaminated water [37]. Last, the norovirus could spread to healthy people by surfaces contaminated by infected people [38]. Therefore, norovirus particles easily infect healthy people via contaminated food, water, and surfaces. Because an effective vaccine to prevent norovirus has not been discovered, it is important for healthy people to prevent norovirus infection by washing hands thoroughly [39]. The transmission of norovirus could be effectively decreased by practicing proper hand

hygiene, preparing food safely, sterilizing surfaces, and washing laundry thoroughly. In this study, we used reasonable prior knowledge of norovirus-related risks to define our list of search terms. In order to cover a wider range of norovirus-related search, we adopted a loose criteria to choose the terms used in this paper that can be used to predict norovirus outbreak.

This study discovered the temporal correlation between norovirus and other terms. Dehydration would be caused by chronic diarrhea or vomiting in older adults or young children. Thus, the search terms including “dehydration” were presented several months later than norovirus in Google Trends. Interestingly, the search term “contagious” was presented earlier than “norovirus” in Google Trends. This indicated that people might contact some contaminated water or food. In addition, people might contact directly with other people who are infected with norovirus. Therefore, they search the term “contagious” before “norovirus.” Apart from typical symptoms, the stepwise regression model showed that internet searches for “cruise” and “wedding” accounted for predictors in the norovirus search. Thus, the internet search terms of some activities might be useful indicators of norovirus outbreaks. In these cases, internet search might occur several months earlier than actual activity for the purpose of reservation and planning ahead of these activities.

Apart from norovirus Google Trends, we also used the data of actual norovirus cases to validate the conclusion. In New York, the cross-correlation analysis of actual norovirus cases and internet search terms showed that the results were similar to norovirus in Google Trends. The Google Trends results of “travel,” “party,” and “wedding” presented one or several months earlier than actual norovirus cases. This result indicated that the search terms “party” and “travel” might be factors of norovirus outbreaks. “Party” and “travel” would lead to people gathering and cause norovirus transmission. This search strategy could be a useful method to predict and monitor outbreaks of norovirus in New York and the United States. Later, we also evaluated the temporal correlation between actual norovirus cases in California with internet search terms in California. However, the results of California were different from those of New York and the United States. The geographic area of California was much bigger than New York. The California illnesses in every norovirus outbreak might be less than New York. It could explain the reason for the difference between New York and California.

Limitations

There are some limitations in this work. First, we tried to discover the correlation between internet search terms and norovirus infection in New York, California, and the United States. Other countries in the world were not included in this study. The results of the United States with Google Trends might not be applied to other countries worldwide. Second, the internet search engine Google was used in this study, other internet search engines and social networking services such as Bing, Twitter, and Facebook were not included in this study. In the future, the social networking services could be included to evaluate outbreaks of norovirus. Last, the norovirus data and Google Trends data were available monthly. Thus, we have not calculated the correlation between daily data on norovirus infections or hospitalizations with Google Trends data in this study. In the future, the correlation between daily data on norovirus infections or hospitalizations with Google Trends data could be explored if the daily data are available.

Conclusions

In this study, we used Google search trends to investigate the correlation between internet searches and norovirus. The data were downloaded from Google Trends and NORS at the CDC. We used cross-correlation analysis to discover the temporal correlation between norovirus and other terms. We found that search trend data from Google is useful to predict norovirus outbreak. Our study provides a novel strategy based on internet search data to investigate the infodemiology of norovirus and monitor the outbreak of norovirus in the future.

Authors' Contributions

KY and GH performed statistical analysis and drafted the manuscript. LW, TW, WL, and HJ contributed to database building. ACY conceived the design of the study and revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The average CSV about the search terms in Google Trends.
[DOCX File , 18 KB - jmir_v23i9e24554_app1.docx]

Multimedia Appendix 2
Cross-correlation analysis of norovirus and Internet search terms–New York.
[DOCX File , 25 KB - jmir_v23i9e24554_app2.docx]

Multimedia Appendix 3
Cross-correlation analysis of actual norovirus cases and Internet search terms–New York.
[DOCX File , 22 KB - jmir_v23i9e24554_app3.docx]

Multimedia Appendix 4
Cross-correlation analysis of actual norovirus cases and Internet search terms–California.
[DOCX File , 25 KB - jmir_v23i9e24554_app4.docx]

Multimedia Appendix 5
Cross-correlation analysis of actual norovirus cases and Internet search terms–US.
[DOCX File , 28 KB - jmir_v23i9e24554_app5.docx]

References

1. Robilotti E, Deresinski S, Pinsky BA. Norovirus. Clin Microbiol Rev 2015 Jan;28(1):134-164 [FREE Full text] [doi: 10.1128/CMR.00075-14] [Medline: 25567225]
2. Ahmed SM, Hall AJ, Robinson AE, Verhoef L, Premkumar P, Parashar UD, et al. Global prevalence of norovirus in cases of gastroenteritis: a systematic review and meta-analysis. Lancet Infect Dis 2014 Aug;14(8):725-730 [FREE Full text] [doi: 10.1016/S1473-3099(14)70767-4] [Medline: 24981041]
3. Isakbaeva ET, Widdowson M, Beard RS, Bulens SN, Mullins J, Monroe SS, et al. Norovirus transmission on cruise ship. Emerg Infect Dis 2005 Jan;11(1):154-158 [FREE Full text] [doi: 10.3201/eid1101.040434] [Medline: 15705344]
4. Karst SM, Tibbetts SA. Recent advances in understanding norovirus pathogenesis. J Med Virol 2016 Nov;88(11):1837-1843 [FREE Full text] [doi: 10.1002/jmv.24559] [Medline: 27110852]
5. de Graaf M, van Beek J, Koopmans MPG. Human norovirus transmission and evolution in a changing world. Nat Rev Microbiol 2016 Jul;14(7):421-433. [doi: 10.1038/nrmicro.2016.48] [Medline: 27211790]
6. Switaj TL, Winter KJ, Christensen SR. Diagnosis and management of foodborne illness. Am Fam Physician 2015 Sep 01;92(5):358-365 [FREE Full text] [Medline: 26371569]
7. Thorne L, Goodfellow I. Norovirus gene expression and replication. J Gen Virol 2014 Feb;95(Pt 2):278-291. [doi: 10.1099/vir.0.059634-0] [Medline: 24243731]
8. Tack DM, Ray L, Griffin PM, Cieslak PR, Dunn J, Rissman T, et al. Preliminary incidence and trends of infections with pathogens transmitted commonly through food - foodborne diseases active surveillance network, 10 U.S. Sites, 2016-2019. MMWR Morb Mortal Wkly Rep 2020 May 01;69(17):509-514 [FREE Full text] [doi: 10.15585/mmwr.mm6917a1] [Medline: 32352955]
9. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
10. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
11. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006:244-248 [FREE Full text] [Medline: 17238340]
12. Bernardo TM, Rajic A, Young I, Robiadek K, Pham MT, Funk JA. Scoping review on search queries and social media for disease surveillance: a chronology of innovation. J Med Internet Res 2013 Jul 18;15(7):e147 [FREE Full text] [doi: 10.2196/jmir.2740] [Medline: 23896182]
13. Mavragani A. Infodemiology and infoveillance: scoping review. J Med Internet Res 2020 Apr 28;22(4):e16206 [FREE Full text] [doi: 10.2196/16206] [Medline: 32310818]
14. Yang AC, Tsai S, Huang NE, Peng C. Association of Internet search trends with suicide death in Taipei City, Taiwan, 2004-2009. J Affect Disord 2011 Jul;132(1-2):179-184. [doi: 10.1016/j.jad.2011.01.019] [Medline: 21371755]
15. Yang AC, Huang NE, Peng C, Tsai S. Do seasons have an influence on the incidence of depression? The use of an internet search engine query data as a proxy of human affect. PLoS One 2010 Oct 28;5(10):e13728 [FREE Full text] [doi: 10.1371/journal.pone.0013728] [Medline: 21060851]
16. Yang AC, Tsai S, Yang C, Shia B, Fuh J, Wang S, et al. Suicide and media reporting: a longitudinal and spatial analysis. Soc Psychiatry Psychiatr Epidemiol 2013 Mar 10;48(3):427-435. [doi: 10.1007/s00127-012-0562-1] [Medline: 22878833]
17. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
18. Seifter A, Schwarzwalder A, Geis K, Aucott J. The utility of "Google Trends" for epidemiological research: Lyme disease as an example. Geospat Health 2010 May 01;4(2):135-137. [doi: 10.4081/gh.2010.195] [Medline: 20503183]
19. Mavragani A, Ochoa G, Tsagarakis KP. Assessing the methods, tools, and statistical approaches in Google Trends research: systematic review. J Med Internet Res 2018 Nov 06;20(11):e270 [FREE Full text] [doi: 10.2196/jmir.9366] [Medline: 30401664]
20. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. PLoS One 2014 Oct 22;9(10):e109583 [FREE Full text] [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]
21. Google Trends. URL: https://trends.google.com/ [accessed 2021-07-26]

22. Mavragani A, Ochoa G. Google Trends in Infodemiology and infoveillance: methodology framework. JMIR Public Health Surveill 2019 May 29;5(2):e13439 [FREE Full text] [doi: 10.2196/13439] [Medline: 31144671]
23. FAQ about Google Trends data. URL: https://support.google.com/trends/answer/4365533?hl=en [accessed 2021-07-26]
24. Hall AJ, Wikswo ME, Manikonda K, Roberts VA, Yoder JS, Gould LH. Acute gastroenteritis surveillance through the National Outbreak Reporting System, United States. Emerg Infect Dis 2013 Aug;19(8):1305-1309 [FREE Full text] [doi: 10.3201/eid1908.130482] [Medline: 23876187]
25. NORS Dashboard. URL: https://wwwn.cdc.gov/norsdashboard/ [accessed 2021-07-26]
26. Patthi B, Kumar J, Singla A, Gupta R, Prasad M, Ali I, et al. Global search trends of oral problems using Google Trends from 2004 to 2016: an exploratory analysis. J Clin Diagn Res 2017 Sep;11(9):ZC12-ZC16 [FREE Full text] [doi: 10.7860/JCDR/2017/26658.10564] [Medline: 29207825]
27. Shimkhada R, Attai D, Scheitler A, Babey S, Glenn B, Ponce N. Using a Twitter chat to rapidly identify barriers and policy solutions for metastatic breast cancer care: qualitative study. JMIR Public Health Surveill 2021 Jan 15;7(1):e23178 [FREE Full text] [doi: 10.2196/23178] [Medline: 33315017]
28. Klein AZ, Magge A, O'Connor K, Flores Amaro JI, Weissenbacher D, Gonzalez Hernandez G. Toward using Twitter for tracking COVID-19: a natural language processing pipeline and exploratory data set. J Med Internet Res 2021 Jan 22;23(1):e25314 [FREE Full text] [doi: 10.2196/25314] [Medline: 33449904]
29. Johnson AK, Bhaumik R, Tabidze I, Mehta SD. Nowcasting sexually transmitted infections in Chicago: predictive modeling and evaluation study using Google Trends. JMIR Public Health Surveill 2020 Nov 05;6(4):e20588 [FREE Full text] [doi: 10.2196/20588] [Medline: 33151162]
30. Romero C, Tinoco Y, Loli S, Razuri H, Soto G, Silva M, et al. Incidence of norovirus-associated diarrhea and vomiting disease among children and adults in a community cohort in the Peruvian Amazon basin. Clin Infect Dis 2017 Sep 01;65(5):833-839. [doi: 10.1093/cid/cix423] [Medline: 29017284]
31. Godoy P, Alsedà M, Bartolomé R, Clavería D, Módol I, Bach P, et al. Norovirus gastroenteritis outbreak transmitted by food and vomit in a high school. Epidemiol Infect 2016 Jul;144(9):1951-1958. [doi: 10.1017/S0950268815003283] [Medline: 26759924]
32. Shah MP, Hall AJ. Norovirus illnesses in children and adolescents. Infect Dis Clin North Am 2018 Mar;32(1):103-118 [FREE Full text] [doi: 10.1016/j.idc.2017.11.004] [Medline: 29406972]
33. Gaythorpe KAM, Trotter CL, Lopman B, Steele M, Conlan AJK. Norovirus transmission dynamics: a modelling review. Epidemiol Infect 2018 Jan;146(2):147-158 [FREE Full text] [doi: 10.1017/S0950268817002692] [Medline: 29268812]
34. Hardstaff JL, Clough HE, Lutje V, McIntyre KM, Harris JP, Garner P, et al. Foodborne and food-handler norovirus outbreaks: a systematic review. Foodborne Pathog Dis 2018 Oct;15(10):589-597 [FREE Full text] [doi: 10.1089/fpd.2018.2452] [Medline: 30109958]
35. Centers for Disease Control and Prevention (CDC). Norovirus activity—United States, 2002. MMWR Morb Mortal Wkly Rep 2003 Jan 24;52(3):41-45 [FREE Full text] [Medline: 12570319]
36. Wang J, Deng Z. Modeling and prediction of oyster norovirus outbreaks along Gulf of Mexico Coast. Environ Health Perspect 2016 May;124(5):627-633 [FREE Full text] [doi: 10.1289/ehp.1509764] [Medline: 26528621]
37. El-Senousy WM, Costafreda MI, Pintó RM, Bosch A. Method validation for norovirus detection in naturally contaminated irrigation water and fresh produce. Int J Food Microbiol 2013 Oct 01;167(1):74-79. [doi: 10.1016/j.ijfoodmicro.2013.06.023] [Medline: 23866911]
38. Repp K, Hostetler T, Keene W. A norovirus outbreak related to contaminated surfaces. J Infect Dis 2013 Jul 15;208(2):295-298. [doi: 10.1093/infdis/jit148] [Medline: 23559462]
39. Assab R, Temime L. The role of hand hygiene in controlling norovirus spread in nursing homes. BMC Infect Dis 2016 Aug 09;16(1):395 [FREE Full text] [doi: 10.1186/s12879-016-1702-0] [Medline: 27507065]

Abbreviations

CDC: Centers for Disease Control and Prevention
NORS: National Outbreak Reporting System

Edited by R Kukafka; submitted 24.09.20; peer-reviewed by A Mavragani, A Sudaryanto, JK Kumar; comments to author 21.12.20; revised version received 26.02.21; accepted 05.07.21; published 29.09.21.

Please cite as:
Yuan K, Huang G, Wang L, Wang T, Liu W, Jiang H, Yang AC
Predicting Norovirus in the United States Using Google Trends: Infodemiology Study
J Med Internet Res 2021;23(9):e24554
URL: https://www.jmir.org/2021/9/e24554
doi:10.2196/24554
PMID:34586079

Original Paper

Using an Individual-Centered Approach to Gain Insights From Wearable Data in the Quantified Flu Platform: Netnography Study

Bastian Greshake Tzovaras[1,2], PhD; Enric Senabre Hidalgo[1], PhD; Karolina Alexiou, MSc; Lukaz Baldy, BSc; Basile Morane[3], MSc; Ilona Bussod[4], BSc; Melvin Fribourg[4], BSc; Katarzyna Wac[5], PhD; Gary Wolf[6], MA; Mad Ball[2,4], PhD

[1]Center for Research & Interdisciplinarity, INSERM U1284, Université de Paris, Paris, France

[2]Open Humans Foundation, Sanford, NC, United States

[3]École centrale d'électronique, Paris, France

[4]Center for Research & Interdisciplinarity, Paris, France

[5]Quality of Life Technologies, GSEM/CUI, University of Geneva, Geneva, Switzerland

[6]Article 27 Foundation, Berkeley, CA, United States

Corresponding Author:
Bastian Greshake Tzovaras, PhD
Center for Research & Interdisciplinarity
INSERM U1284
Université de Paris
8bis Rue Charles V
Paris, 75004
France
Phone: 33 766752149
Email: bgreshake@googlemail.com

Abstract

Background: Wearables have been used widely for monitoring health in general, and recent research results show that they can be used to predict infections based on physiological symptoms. To date, evidence has been generated in large, population-based settings. In contrast, the Quantified Self and Personal Science communities are composed of people who are interested in learning about themselves individually by using their own data, which are often gathered via wearable devices.

Objective: This study aims to explore how a cocreation process involving a heterogeneous community of personal science practitioners can develop a collective self-tracking system for monitoring symptoms of infection alongside wearable sensor data.

Methods: We engaged in a cocreation and design process with an existing community of personal science practitioners to jointly develop a working prototype of a web-based tool for symptom tracking. In addition to the iterative creation of the prototype (started on March 16, 2020), we performed a netnographic analysis to investigate the process of how this prototype was created in a decentralized and iterative fashion.

Results: The Quantified Flu prototype allowed users to perform daily symptom reporting and was capable of presenting symptom reports on a timeline together with resting heart rates, body temperature data, and respiratory rates measured by wearable devices. We observed a high level of engagement; over half of the users (52/92, 56%) who engaged in symptom tracking became regular users and reported over 3 months of data each. Furthermore, our netnographic analysis highlighted how the current Quantified Flu prototype was a result of an iterative and continuous cocreation process in which new prototype releases sparked further discussions of features and vice versa.

Conclusions: As shown by the high level of user engagement and iterative development process, an open cocreation process can be successfully used to develop a tool that is tailored to individual needs, thereby decreasing dropout rates.

(J Med Internet Res 2021;23(9):e28116) doi:10.2196/28116

KEYWORDS

symptom tracking; COVID-19; wearable devices; self-tracking; citizen science; netnographic analysis; cocreation

Introduction

Background

Patient- or participant-led research has been suggested to improve self-management capabilities [1] and provide ways to generate otherwise undone science [2,3]. A particular subtype of participant-led research is personal science, which involves the use of empirical methods by individuals to pursue personal health questions [4]. Personal science is a distinct category of citizen science that has emerged from the Quantified Self community and its efforts to advance participant-led research [5,6]. In personal science, practitioners almost always take the lead in all stages of the research process by definition [4]. Owing to this high level of individual engagement and tailoring to individuals' interests, personal science has the potential to deliver novel insights relevant to its practitioners [7], which can lead to an improved sense of agency and quality of life [8]. Furthermore, the insights and self-expertise generated by these types of participant-led processes have potential relevance for professional and scientific research, both topically as a source of ideas and methodologically as a source of tools, analytical approaches, and workflows [9].

Wearable devices—from wristbands to smartwatches and other personalized, miniaturized on- and around-body devices—are frequently used by self-trackers. These devices are becoming increasingly common and are used for a wide spectrum of well-being, fitness, and health-related purposes [10]. This is further facilitated by the fact that the number of sensors used in these devices is growing rapidly. In addition to accelerometers and gyroscopes to track physical activity, sensors to measure physiological signals such as heart rate, body temperature, respiratory rate, and blood oxygen saturation, which may correspond to the health or sickness state of the human body [11], are also frequently found in wearables [12,13]. Consequently, even outside the realms of personal science, wearables have long been seen as promising tools for facilitating health-related monitoring and enabling personalized medicine [14,15] and have been proposed or used to monitor conditions as diverse as cardiovascular disease [16,17], Alzheimer [18], and graft-versus-host disease [19].

In response to the COVID-19 pandemic, interest in using wearable technology for infection prediction and surveillance has increased [20-22]. Anecdotal reports from self-trackers suggest that wearables may provide evidence of COVID-19 infection [23]. During the first year of the COVID-19 pandemic, a small number of studies appeared, highlighting that wearable devices, often along with self-reported symptoms, might indeed be used for the early detection of COVID-19 infections and to assess physiological symptoms [24-27]. The majority of these studies take a crowdsourcing-based approach—in which participants are invited to contribute by providing their own wearable data along with regular symptom reports and COVID-19 test results—as the main way of engaging individuals. The goal of the data collection process in these studies is to create big data sets to interrogate.

In contrast, there have been limited efforts to engage personal scientists in cocreating such symptom tracking efforts. Personal science practices are largely done in isolation, and the Quantified Self movement has consequently accumulated limited knowledge so far [8]. To fill this gap, we present a case study of Quantified Flu (QF), a project cocreated by a community of personal science practitioners in response to the COVID-19 pandemic.

Objective

The goals of this work are twofold. First, we documented the contrasting cocreation approach of QF with its focus on personal science rather than large-scale research. To this end, we used netnographic methods to document how the cocreation process developed and generated a citizen science platform prototype over a relatively short period. Second, we explored the consequences of the projects' contrasting cocreation approach and focused on personal science, particularly with respect to the ultimate design of the QF tool and its use.

Methods

Overview

The cocreation process of this study is based on an action research approach [28], simultaneously for developing a useful community resource while also generating shared knowledge about the process. In our case, action research was implemented through practical work to support the participatory design of a digital platform [29,30] under open-source principles [31], followed by netnographic data collection and analysis to understand its development and usefulness as a cocreation process [32]. For this, all authors except ESH were involved as participants in the cocreation process, in collaboration with the rest of the participants, during the iterative prototyping of the QF platform.

Community Cocreation Process

QF began with a discussion on the monthly Open Humans (OH) community call at the beginning of the COVID-19 pandemic on March 10, 2020. OH is a platform for empowering individuals around their personal data, to explore and share research processes for the purposes of education, health, and science in general [33]. The community calls involved 83 individuals so far (until September 3, 2020), and the monthly calls are frequented by a mix of citizen science and personal science practitioners; usually, around 10 individuals take part in each call. Following an initial brainstorming, the discussions and planning stages were continued through the following community calls and a dedicated communication channel of the OH community Slack [34]. Furthermore, over the evolution of the project, other communities such as Quantified Self [35] and OpenCovid19 Initiative [36] were engaged and involved in different aspects of the development of the project.

In parallel to 10 additional community calls between March 10 and September 3, 2020, the main coordination tool for the QF project was a specific Slack channel, with a total of 146 subscribers and 34.2% (50/146) active users over time with different levels of involvement and activity. During this timeframe, this openly accessible channel gathered a total of 844 messages from these users, with a total count of 26,691 words (and 3917 unique words).

Although the planning, coordination, and social aspects of the cocreation process mainly took place on the mentioned project's Slack channel, technical collaboration and software development occurred through GitHub and the git repository of the QF. Due to the iterative nature of open-source collaboration, no up-front requirement analysis was performed. Instead, prototypes were developed over time according to community discussions by iteratively adding and testing implementations. On GitHub, 7 contributors created a total of 316 commits since March 12, 2020, leading to the technical prototype outlined later. The source code for the project is available under an open license on GitHub [37].

Netnographic Content Analysis

To investigate and analyze the cocreation stages that led to the QF prototype, we performed a netnographic analysis of its iterative communication process, similar to previous studies on cocreation in health-related community settings [38]. Netnography is an interpretive research method derived from ethnography, usually applied to social interaction processes in digital channels and platforms, and focused on digital traces of public conversations as analyzable data. As a qualitative technique broadly applied to the study of web-based communities [32], Netnography allows capturing and reflecting interactions as an observational, inductive, and unobtrusive approach while combining it with participatory methods [39]. In particular, we examine how individuals engaged in the QF Slack channel for the collaborative development of the QF platform as a case study setting [40].

For this part of data collection, one of the researchers (ESH) developed a codebook combining key concepts of cocreation and collaboration in communities of practice (Textbox 1). The codebook was cross-checked for validity by 2 other authors (BGT and MB). Following this, it was applied to the QF Slack channel posthoc without this specific researcher (ESH) having participated in the previous community discussions.

Textbox 1. Codebook for Quantified Flu Slack communication message content analysis.

Communities of practice-related messages

- Socialization
  - Support or coordination: parallel messages regarding overall coordination and personal and empathic support interventions
  - Possible collaborations: ideas regarding potential collaborators and connections to other organizations or experts who can support or contribute to the project
  - Outreach: messages related to the visibility of the project, possible dissemination, or alliances for spreading the process
  - Off topic: nonrelated messages to any of the previous (eg, about personal issues or intention-to-buy wearables)

Cocreation-related messages

- Ideas
  - Inspiring or similar initiatives: mentions to other COVID-19-related projects being developed or known externally
  - COVID-19 related: links to news or updates regarding the COVID-19 pandemic and its evolution
  - Mention to tool or wearable: references to a specific wearable for its potential connection to the Quantified Flu project
  - Scientific knowledge or papers: mentions or links to studies or publications and elaborated scientific knowledge
- The Quantified Flu concept
  - Goal setting or discussion: concept-related interventions about the objectives of the project
  - Protocol or tool design: mentions to how the protocol and tool should work or specific aspects of its possible design
  - Feature suggestion: interventions suggesting specific characteristics or new possible features of the tool
  - Pattern or data observation: statements regarding the observation of data in relation to the goals or possible functioning of the project
- The Quantified Flu prototype
  - Incremental development or updates: messages informing about new implementations and code development of features applied to the prototype
  - Technical issues: specific technical issues to solve or observations about needed improvements for correct use
  - Help testing: interventions asking or offering support in testing the tool by community members
  - Help developing: interventions asking or offering technical support for the development of the tool

This part of data analysis was used to determine the typology of messages regarding the cocreation of the QF platform, from idea to concept to prototype [29], and other types of messages relevant from a communicational and empathy-needed dialogic process in communities of practice [41]. Each Slack message was assigned up to three top tags based on the aforementioned

codebook categories, depending on its text density and characteristics. The researcher (ESH) assessments of types and categories of messages were subsequently reviewed and discussed by another coauthor (BGT), who was actively involved in the analyzed cocreation process.

Results

We present the current prototype of the QF platform [42], as a result of the described technical development, before analyzing the cocreation process that led to it.

Community-Based Development

The first overview derived from our netnographic analysis of the four main categories of messages interchanged during the cocreation of the QF prototypes on its dedicated Slack channel (March 10 to September 3, 2020) shows a relative balance in the topics of the web-based messages among the 1171 message fragments that were annotated (Textbox 2).

Overall, during the development of QF, the *Prototyping* and *Socialization* messages were slightly more common than the *Concept* and *Ideas* ones (Figure 1). On the level of the tags or subcategories, the most frequent ones are *Support or coordination* (227), *Protocol or tool design* (109), *Technical issues* (107), and *Help developing* (106).

Focusing on these more specific tags, as defined in the codebook (Textbox 1), within each category over time (Figure 1), we observe that all the four main categories, as well as the individual tags, are present over the whole time frame of cocreation from early April to September 2020. In particular, messages regarding *Support or coordination* are present throughout the entire time range. Other recurrent message types during the analyzed time span fall within the categories *Ideas*, *Concept*, and *Prototyping*, highlighting the iterative design, implementation, and testing participatory processes that took place to develop and improve the QF prototype over time.

Importantly, the *Protocol or tool design*, *Mention to tool or wearable*, and *Feature suggestions* categories, which are indicative of the cocreation process, appear early on but remain active in bursts throughout the full observed time span, often following new releases of the QF prototype. In addition, the *Help developing* and *Help testing* categories remain active over the whole duration of the prototype development, with the former showing a more constant activity (mean 1.1 tags per day, SD 1.9) whereas the latter appears in bursts (mean 0.76 tags per day, SD 2.1) around new feature releases.

Textbox 2. Examples of messages that were tagged according to the codebook used for the netnographic analysis.

Socialization

- Support or coordination
  - "Very good community call focused on quantified flu this morning. Glad I participated"
  - "Great updates, thank you and I am definitely staying tuned...how can I help other than visualisations. Have a good rest of the day!"
- Possible collaborations
  - "[Person outside the community] is usually sitting in the office right next to mine (working from home these days) and we'll do a call tomorrow to chat about synergies!"
  - "It's possible we could address these with support from the company, which in turn depends on our convincing them to prioritize this support. We have some close contacts there that could lead to success"
- Outreach
  - "We're (very briefly) featured in the latest UCSD newsletter"
  - "Oh, and we already got some media coverage in the german 'digital living' magazine t3n"
- Off topic
  - "Not sure if it's appropriate to ask but does anyone have a way to get discount on the Oura ring?"
  - "BTW, semi-related to this project: Just coinciding with the general lockdown in Paris I stopped smoking and could nicely see my resting heart rate drop, my heart rate variability grow, etc. within the first few days"

Ideas

- Inspiring or similar initiatives
  - "Another flu-tracking app, from Duke:...[URL]"
  - "Looks like Michael Snyder's famous self-tracking lab at Stanford is doing something similar"
- COVID-19 related
  - "Placing this link here because it was an interesting symptom diary someone shared on Twitter they made"
  - "Btw. during the community call yesterday, [community member] shared this symptom report of a contributor who thinks he has covid19"
- Mention to tool or wearable
  - "How far is the Fitbit Intraday integration? As of right now the Fitbit Graph seems quite a bit less detailed than the Oura one"
  - "The Garmin devices are a bit tricky, as their API is locked off unless you apply for access with them"
- Scientific knowledge or papers
  - "August 31 Webinar from hlth.com: Wearable Technology's Potential to Help Detect Illness"
  - "Stanford's 2017 paper was all about longitudinal health data and health outcomes"

The Quantified Flu concept

- Goal setting or discussion
  - "Personally, I see the purpose of this project not so much as epidemiological, but about expanding the personal value of our data. Doing that as a group helps us learn from context, as well as individually"
  - "This project might also be a starting point for prospective tracking for people that get sick, going forward. Still thinking about if/how that would work."
- Protocol or tool design
  - "So i think it may make sense to give numerical values for each symptom, from 1-5 or 1-10 in terms of intensity, and also timestamp them to allow for multiple logging within a day"
  - "but my vicks smart temp thermometer arrived...the associated app allows me to record...medication, symptoms (cough, sore throat, chills, body ache, ear ache, nausea, stomach ache, fatigue, short breath, headache, diarrhea, runny nose), a free text 'notes'"
- Feature suggestion

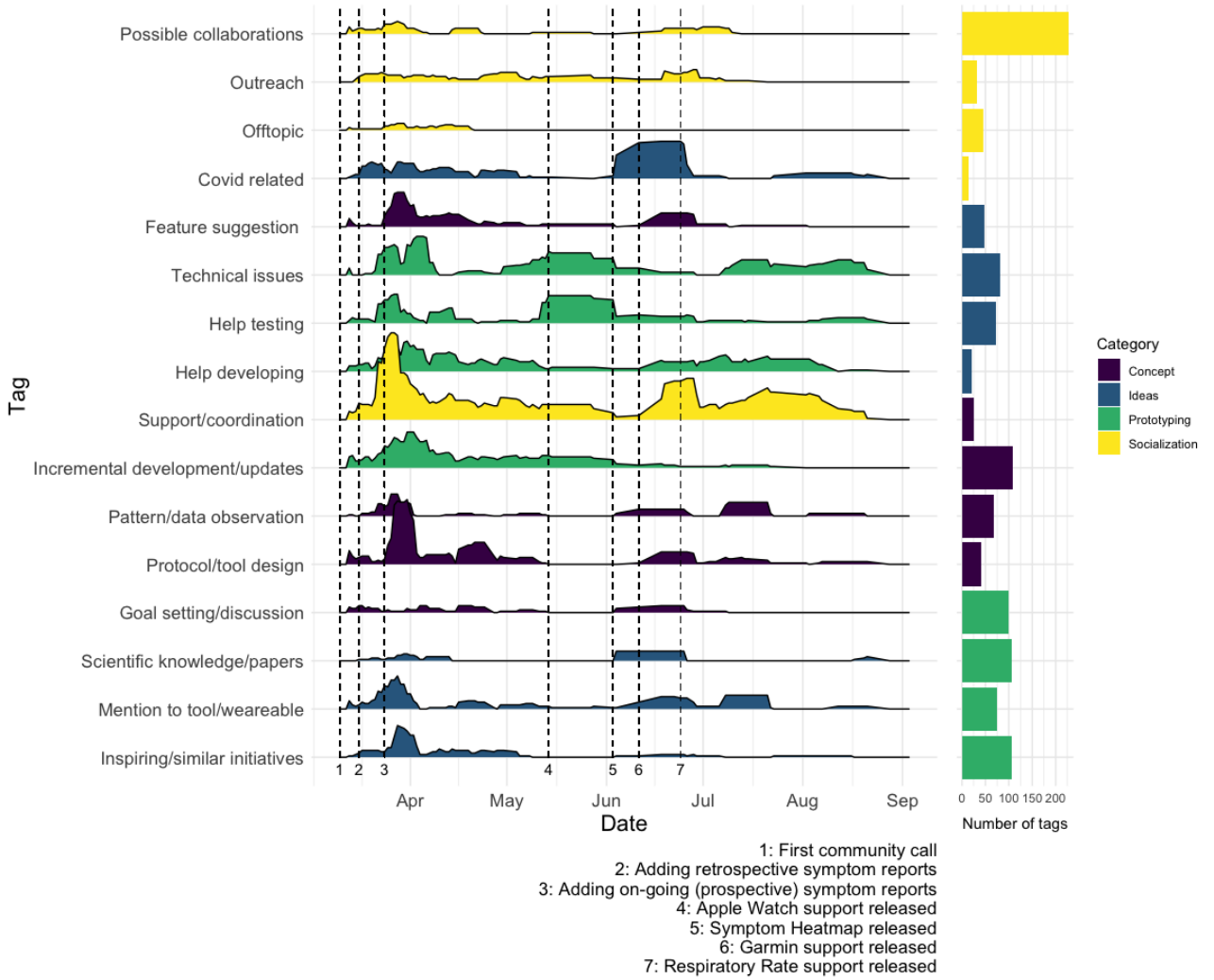
- "She wears an Apple Watch and has resting heart rate data. Should I invite her join quantified flu even though she does not have an Oura or Fitbit? Is adding support for Apple Watch too much work at this stage? Manual entry wouldn't be very challenging probably"
- "I'm realizing the public list could at least give event IDs so you have some sort of identifier for each one"

- Pattern or data observation
  - "Already contributed two sick events of mine from 2019, that are very obvious in the data but also quite different"
  - "So I don't think it's necessarily measurement noise. For my own data my gut feeling is that all variations ≤ ±0.3 °C are probably just daily fluctuations for a myriad reasons"

The Quantified Flu prototype

- Incremental development or updates
  - "Some publicly available data now – you can explore on the site, and there's JSON endpoints to get raw data"
  - "Hey <!channel>, we have another nice visualization update thanks to [community member]! The retrospective events now have the same display that can be found for the ongoing symptom reports, check out [QF link] for an example!"
- Technical issues
  - "Oh, not sure if that's true though! I think if the oura dies while doing the recording it doesn't deliver any data (happened to me 3-4 times with my broken oura where it would not record anything for the night)"
  - "Also, I found a strange inconsistency in the data. For one of the users, the JSON file states that they are sick on July 11th, but the interactive display on the website does not (the JSON says that the person had a sore throat, but the web display does not). I attach the examples"
- Help testing
  - "My daily symptom checkins have stopped, is this happening for anybody else? I thought it might be an email issue on my side"
  - "Does anyone of you have an android watch/wearable that would track heart rate to test whether it works?"
- Help developing
  - "Hi everyone! I am a programmer and would be happy to help. I have lots of experience with python"
  - "I thought a cool starting visualization could be a heatmap similar to the github activity view, but with time only on the x-axis and the different symptoms on the y-axis and colored by symptom severity. If you have other cool ideas for appealing and insightful visualizations feel free to let us know!"

Figure 1. Distribution of message types over time; the frequency of tags are given as 7-day rolling averages. Events 1-7 around Quantified Flu development are given as vertical lines. Bar plots show the total number of tags per category.

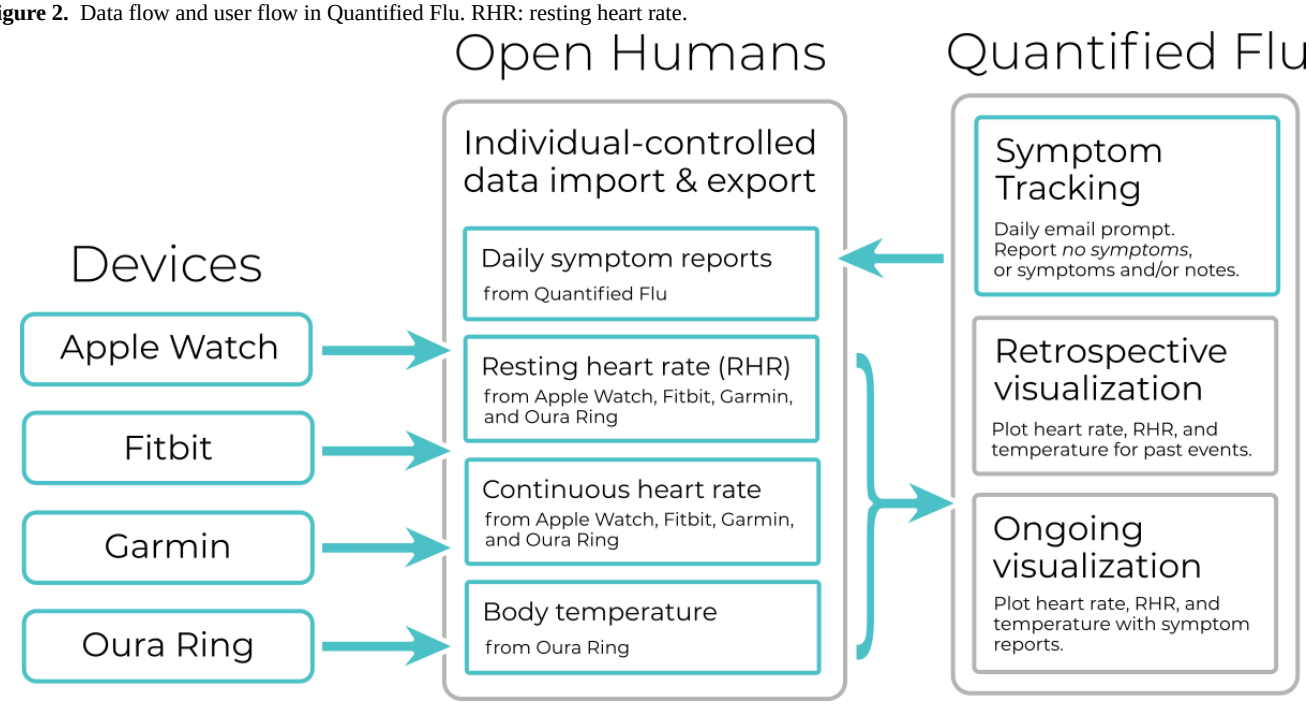


QF: Technical Platform Implementation

Overview

As a result of this community-based development process, QF evolved into a responsive web application that can connect to a wide variety of devices, implemented in Python or Django programming language. Users must be registered on OH, having the option of linking a range of available wearable devices from which physiological data (heart rate, body temperature, and respiratory rate) can be imported into the OH platform; visualize past sickness or infection events (retrospectively) on it (present since the first prototype, launched on March 16, 2020); and engage in daily (prospective) symptom tracking (added in the second prototype, released on March 24, 2020; Figure 2).

Figure 2. Data flow and user flow in Quantified Flu. RHR: resting heart rate.



User Accounts, Data Storing, and Anonymization

To enable rapid prototyping, QF is connected to the OH platform [33] as a back end to manage user permissions and store user data. OH provides OAuth2-based application programming interfaces (APIs) to authenticate users while keeping each user pseudonymous to the QF platform, as no personally identifiable information is transmitted. Instead, only a random 8-digit user identifier specific to the QF project is provided. Furthermore, OH provides APIs to access and store user data in their system through those identifiers and provides methods for users to consent to share data from the OH platform with third parties such as QF.

Wearables

To further bootstrap the creation of the prototype, QF made use of the existing wearable integrations that OH already offered (Fitbit daily summaries, Fitbit intraday data resolution, and Oura Ring). To facilitate usability, QF also integrated these data import methods directly into the prototype, using OH as the data store for the wearable data.

Furthermore, following community suggestions and ideation discussions (Textbox 2; *Mention of tool or wearable*), QF also added Google Fit (May 6, 2020), Garmin (June 11, 2020), and Apple Health (May 14, 2020) as additional supported wearable devices. Depending on the wearables, users can import and use their heart rate throughout the day, daily resting heart rate, body temperature, and respiratory rate in QF (Table 1).

Table 1. Wearables supported by Quantified Flu.

Wearable	Development	Resting heart rate	Heart rate throughout day	Body temperature	Respiratory rate
Fitbit	Existing	✓[a]			
Fitbit Intraday	Existing	✓	✓		
Oura Ring	Existing	✓		✓	✓
Google Fit	Extended (added heart rate data)	✓			
Apple Health	Added	✓	✓		
Garmin	Added	✓	✓		

[a]The feature is measured by the wearable.

Unlike other wearables integrated into QF, Apple Watch does not provide a web-based API to access and export data. Thus, following another community suggestion (Textbox 2; *Feature suggestion*), a mobile iOS app was created to provide a link to QF. This specific app enables users to export their heart rate data collected by Apple Watch. The source code for this mobile app is also available under an open license [43].

Symptom Tracking

Users can report symptoms using a QF website. On the basis of previous works [24-26] and early community discussion and feedback (Textbox 2; *Protocol or tool design*), QF implemented a list of 12 symptoms that were classified as respiratory, gastrointestinal, and systemic symptoms (Textbox 3), allowing users to score those on a 5-point scale (1=light; 5=worst). In addition, users can report fever measurements and use free-text

fields for the suspected origin of their symptoms, further symptoms, or notes to put their symptoms into context (Textbox 2; *Protocol or tool design*).

Textbox 3. Symptoms of sickness that users can monitor in Quantified Flu.

Respiratory

- Cough
- Cough with mucus or phlegm
- Reduced sense of smell or anosmia
- Runny or stuffy nose
- Sore throat
- Shortness of breath

Gastrointestinal

- Diarrhea
- Nausea or vomiting

Systemic

- Chills and sweats
- Fatigue and malaise
- Headache
- Muscle pains and body aches

Users can opt in to receive daily symptom report reminders that are sent through the anonymous OH email system at a user-selected time, as another tool feature that was discussed and regularly tested by participants (Textbox 2; *Help testing*). Each email contains the following two links: (1) the *reporting no symptoms* link, a single-click link that requires no further interaction of the user, and (2) the *reporting symptoms* link, which takes users to the symptom report form.

Data Visualization

Overview

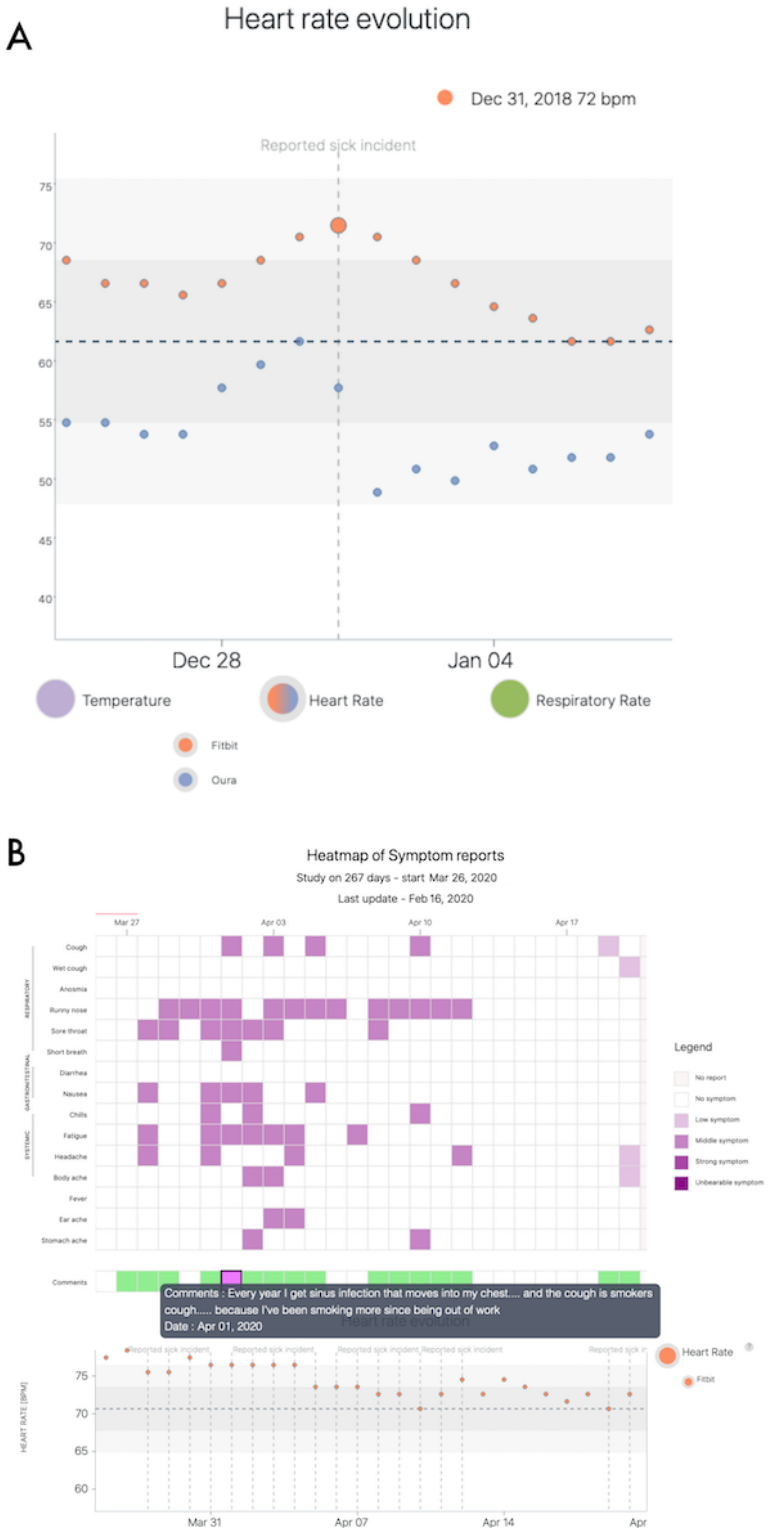
To provide users with easy ways to facilitate understanding of their own physiological data and potentially explore it in relation to their own symptom reports, QF used *D3.js* to create interactive visualizations. These visualizations present the evolution of the various physiological data points and put them into the context of their symptom reports where available.

The QF platform provides personal science practitioners 2 main ways to explore their physiological wearable data in relation to infections—the retrospective analysis of prior events and an ongoing (prospective) analysis of symptom reporting.

Retrospective Analysis

Users can select a given, historic date on which they fell sick or had specific symptoms, and QF will, if available, extract wearable data for the 3 weeks before that date and 2 weeks after the incident. This allows users to visualize sickness incidents that occurred before the launch of QF. Depending on the wearable (Table 1), users are given the option to display different physiological variables over that 5-week time period and explore how they change over time. To facilitate the interpretation of changes and outliers in the graphs, both the first and second SDs are presented as well (Figure 3). Although users can add comments to retrospective events, detailed symptom reports are absent in this mode, as most users do not have detailed records of the historic sickness events. The retrospective analyses were part of the first prototype of the QF, launched on March 16, 2020.

Figure 3. Screenshots of the Quantified Flu prototype showing typical visualizations generated by users. (A) An example data visualization of an individual, retrospective sickness incident that happened on December 31, 2018. The data plotted are the resting heart rate recordings as measured by the Fitbit and Oura Ring. (B) An example of an ongoing symptom report visualization. The top half shows a heat map of the symptoms that were present along with their strength, and green boxes display user-provided free-text comments. The bottom half shows physiological data from wearables. bpm: beats per minute.



Ongoing Symptom Reporting

Users can also report currently experienced symptoms through QF at any moment in time by selecting symptoms and their experienced strengths from a list (Textbox 3). This self-report is likely triggered by email, as explained earlier. Following symptom reports, users are automatically taken for their data visualization (Figure 3). On a wearable device data level, this visualization provides the same details as that of the retrospective analyses (*Retrospective Analysis* section). The ongoing symptom reports (Figure 3) were launched as a new feature in the second iteration of the prototype on March 24,

2020, also following discussion and contributions from the community (Textbox 2; *Incremental development or updates*).

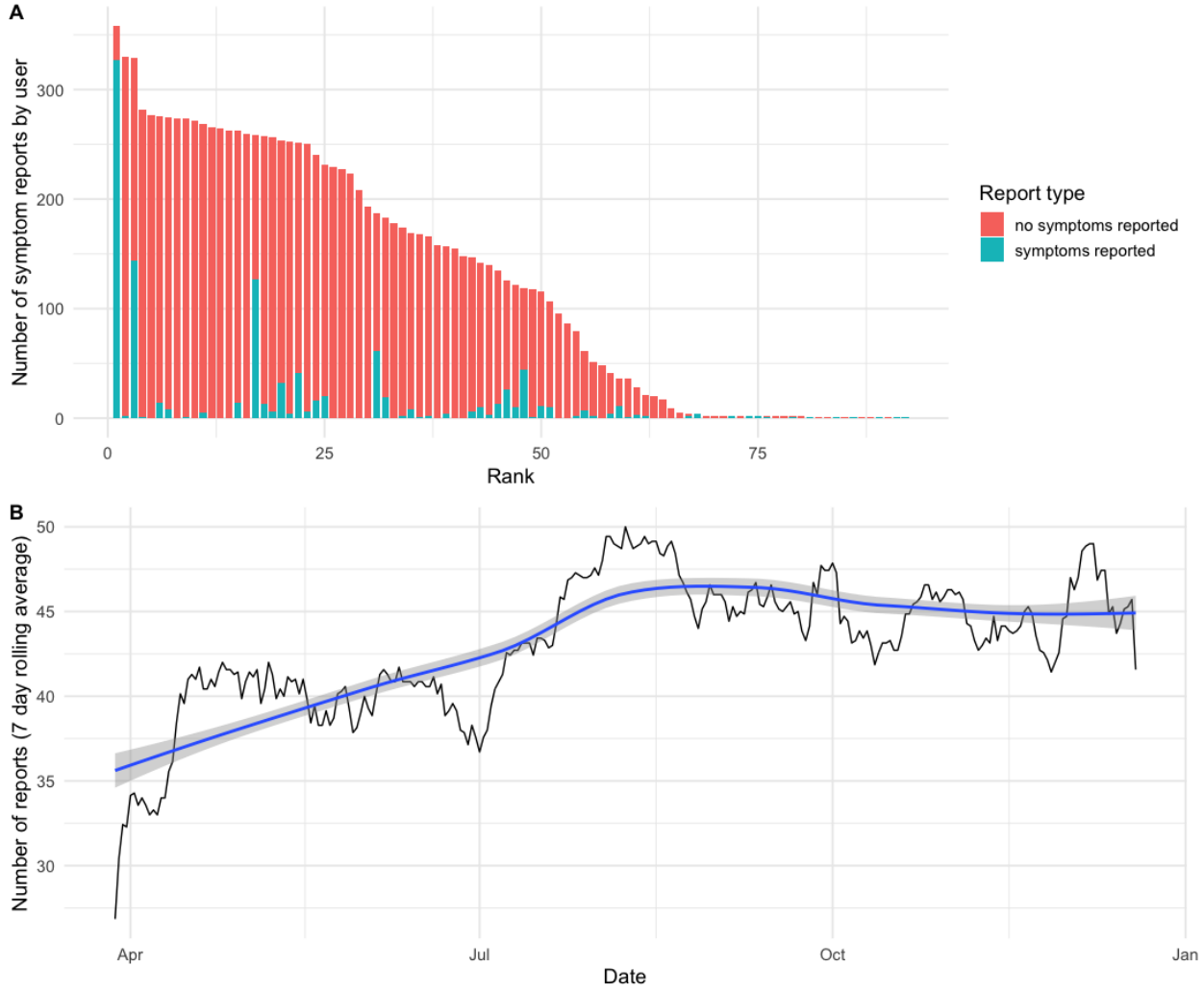
In addition, this latter view aligns a heat map of each daily symptom report to the wearable data timeline, allowing the identification of patterns within the reported symptoms themselves and for visual cross-comparisons between the physiological data and the symptom reports. Furthermore, users can also access their comments for each symptom report from this visualization, allowing them to understand the contexts in which they made those reports.

Community Use

A total of 190 personal science practitioners engaged with QF between its launch on March 16, 2020, and December 22, 2020. The initial prototype of QF (in place until March 24, 2020) only offered the possibility of analyzing retrospective sickness events. This feature was rarely used: only 24 users tried the feature, creating a total of 47 retrospective analyses. In total, 34 individual wearables were linked by these 24 users. The prospective ongoing symptom report feature was launched on March 24, 2020. In total, 92 users made use of this feature at least once, covering a range from a single symptom report being done up to over 300 reports for some members. Overall, 11,658 symptom reports were filed and 112 wearables were linked to it, between the launch of the feature and December 22, 2020.

The distribution of user engagement for the entire period (Figure 4), as measured by the number of reports, shows an approximately linear relationship between the number of reports done and the user's rank of activity. The reports with symptoms are also not equally distributed across all 92 users, with a sizable fraction of users having no or only a few reports that include symptoms, whereas for some users, symptom reports comprise half or nearly all of the reports. Overall, the vast majority (10,594/11,658, 90.87% reports) were reports that included no symptoms. Of the 1064 reports with symptoms, 176 (16.54%) included explanatory notes or comments, in addition to the standardized symptom reports.

Figure 4. The use of Quantified Flu as measured by ongoing symptom reports filed by users. (A) Users were ranked by the number of symptom reports they have filed; data were broken down into whether symptoms were reported (blue) or not reported (red). (B) The number of symptom reports filed per day. Values were averaged into a weekly rolling average. The blue line represents the local regression–smoothed data along with SE (gray background).



Looking at the number of symptom reports filed per day, we can observe a rapid rise in daily reports at the beginning of April 2020, reflecting the launch of the first prototype with the ongoing symptom reports. The second rise in daily reports started in July 2020, leading to the numbers starting to stabilize

at around 45 symptom reports filed per day (by an average 45 users per day, SD 5; Figure 4).

Discussion

Principal Findings

In this paper, we present QF, a cocreated web-based project that enables personal science practitioners to engage with their own wearable data and visualize it in the context of when they are experiencing symptoms of potential infection. The spark that led to this community deciding to cocreate a symptom tracking tool was the beginning of the global COVID-19 pandemic, along with population-wide studies that made individuals wonder how useful their own wearable data might be for them in such a pandemic context. With its focus on individual learning, QF stands in contrast to various population-level studies performed to evaluate the usefulness of wearable technology for the prediction of illness [20,24-27]. At this individual scale, symptom tracking and health data more generally can offer support for individual sense-making on health experiences and conditions [44], which can be idiosyncratic and complex [45]. QF was also distinguished by a cocreation approach that targeted the individual learning and research interests of a web-based community and involved the iterative development of a digital tool in response to feedback, resulting in a format that attracted increased and sustained participation from early users.

One of the main aims of our work is to investigate the consequences of a cocreation approach that focuses on personal science: We observe that the initial QF prototype, which focused solely on retrospective symptom tracking, was rarely used. Only 24 users were engaged in this prototype. However, importantly, this initial version facilitated additional discussions about designing both the data collection protocol and extending the prototype (Figure 1), leading to the creation of the ongoing symptom reports as a feature launched in the next QF iteration. This feature received much more attention from the participants, with a total of 92 people using QF for their own regular symptom tracking, delivering some first insight into the importance and potential benefits of early engaging potential users in a health research design cocreation approach.

Furthermore, we observe that the level of engagement across these 92 QF users seems to drop linearly when ordered from most to least engaged users (Figure 4). This distribution is atypical for user engagement in web-based communities, where one typically observes power-law distributions for engagement [46]. Related to this, digital or mobile health apps in particular typically struggle with achieving continued use, as a large fraction of users drop out after a few interactions [47,48]. In previous studies, only 2% of initial users showed sustained use in the most extreme cases, with observational studies having an average dropout rate of 49% [49]. In contrast, around half of the QF users who engaged with ongoing symptom tracking did so on a regular basis, leading to 45 (SD 5) symptom reports per day on average (Figure 4), and over 50 users reported more than 3 months of symptom reports, highlighting continued longitudinal use. We argue that these uncharacteristically high numbers of user engagement, which is sustained over time, is a result of the community cocreation process that led to the final prototype of QF. Previous studies have found that users are more likely to continue using mobile health apps if there is a good fit between users and applications [50], which means that a cocreation process among future users could be a key way of achieving this fit.

For some users, this continued engagement might also be an indication that they experience regular or recurring symptoms, making them particularly interested in learning empirically about them through this specific kind of self-tracking. This is supported by the number of reports that include symptoms, where a subset of users reported having symptoms frequently, with some users reporting symptoms in 40%, or extreme cases even 90%, of the time (Figure 4). Further evidence for this comes from the notes or annotations that users can submit to the QF website along with their symptoms when filing their daily reports. Looking at the publicly shared notes in these reports, we find examples like “the cough is smokers cough...because I’ve been smoking more since being out of work” and “I was deep cleaning the house...all the dust got my allergies going again” highlighting possible reasons for recurring symptoms. Furthermore, these annotations help to provide context to individuals and others that aim to reuse publicly shared data. Although a severe case of coughing or nasal congestion might hint at acute infection, they might also be unrelated, as the annotations highlight. These contextual descriptions can be difficult to formalize, potentially explaining why symptom-based diagnoses are difficult to achieve in many cases [51,52].

Our second main goal, in parallel to the development of the QF prototype itself, is to explore how a community-driven initiative can contribute to collectively creating the tools needed to build self-knowledge by conducting a netnographic analysis of the main QF communication channel. Reflecting on the use of this qualitative and interpretative methodology for the study of web-based communities [53], we find that it adapts well to user-led prototypes, with some particular strengths and limitations. In the case of QF, we found that the netnographic approach was well suited to allow a posthoc study of the participatory design process after the prototype creation. This approach could be valuable in obtaining a better understanding of cocreation dynamics in similar health-related projects and studies [54], as it can be applied to existing text corpora of community interactions on digital text tools such as Slack, mailing lists, or forums. Its reliance on text communication is also one of the main limitations, as synchronous meetings—remote and physical—are less accessible as archival data, requiring recordings and transcriptions. Given this, it might be advisable to organize cocreation processes with Netnography techniques in mind to ensure adequately sized text corpora.

Applying such a netnographic approach to QF, we found a marked overlap of the various phases of ideation, conceptualization, and prototyping over time. Although a greater number of interactions can be found in the initial phases, there is a sustained regularity later on, particularly in areas such as feature suggestions or the design of the tool and protocol. In this sense, messages and interactions related to helping with development throughout the whole process reflect a typology of continuous and iterative cocreation, which is typical of

collaboration processes in the development of open-source tools [55].

This iterative cocreation process is also highlighted in the burst-like appearance of feature suggestions and protocol or tool design discussions, which frequently appear following the release of new features, suggesting that new releases spark further protocol refinements and feature ideas, which in turn lead to the QF prototype refinement. Importantly, this means that the protocol itself, along with the concrete implementation, remains in a stage of flux over a longer period of time, compared with more traditional research design approaches. As a result, this type of collaborative approach is at odds with standard ethical oversight procedures for human subject research that require a precise predefinition of the protocol and the role of the individuals, whereas the main feature of cocreation is that it is emergent and adaptive, making detailed prespecifications impossible [56]. To fully take the advantage of the benefits of cocreation in the participant-led research, it might be necessary to develop different models of ethical oversight that recognize the autonomy of participants [57,58], to not discourage or stifle valuable forms of participant-led research [2].

Finally, it is also important to highlight how the other types of messages associated with communication in a community of practice context, which favor both web-based empathy and effective coordination, were produced in a prominent, constant, and sustained manner from the beginning of the cocreation process (*Support or coordination*; Figure 1). This mode of cocreation can be understood as an example of uninvited citizen science that relies on a shared set of values, self-stabilizing communication infrastructure, and a loosely defined coproduced knowledge object [59] (eg, the QF prototype itself). This way, the development of the data collection platform itself is framed in a dynamic, bottom-up, and adaptive way, similar to other open source and peer production experiences.

Conclusions

Although QF is a project that is still at the prototype stage and with a correspondingly small user base, the cocreation processes of the platform prototype described here represent an example of how the codevelopment of digital research objects, within the relatively new participatory paradigm of extreme citizen science [60], can be implemented following bottom-up, dialogic approaches and a high level of participant engagement. This aligns with the still scarce literature on what has been called do-it-yourself science or peer-to-peer science [61,62], in which similar participatory approaches can offer an opportunity for early and sustained engagement from personal science practitioners in the collaborative definition of concepts, features, and protocols for health-related digital platforms.

Acknowledgments

The authors would like to thank all participants of the Open Humans community calls and the Quantified Self and OpenCovid19 communities for their input and support. The authors would also like to thank all of the users of Quantified Flu for their support and feedback. The authors are also grateful for the suggestions and recommendations of the two anonymous reviewers who helped to substantially improve our manuscript. Thanks to the Bettencourt Schueller Foundation's long-term partnership, this work was partly supported by the Center for Research and Interdisciplinarity Research Fellowship (awarded to BGT). This work was also supported by a microgrant from the OpenCovid19 initiative. KW was supported by the H2020 WellCo Project (769765).

Authors' Contributions

BGT, KW, GW, and MB initiated the study. All the authors contributed to the study design. BGT, KA, LB, BM, and MB performed software development. BGT, ESH, IB, and MF performed data analyses and visualizations. BGT, ESH, and MB prepared the original draft of the manuscript. All authors reviewed and edited the manuscript before submission.

Conflicts of Interest

MB is the executive director of the Open Humans Foundation. BGT is the director of research for the Open Humans Foundation.

References

1. Chiauzzi E, Rodarte C, DasMahapatra P. Patient-centered activity monitoring in the self-management of chronic health conditions. BMC Med 2015 Apr 09;13:77 [FREE Full text] [doi: 10.1186/s12916-015-0319-2] [Medline: 25889598]
2. Vayena E, Brownsword R, Edwards SJ, Greshake B, Kahn JP, Ladher N, et al. Research led by participants: a new social contract for a new kind of research. J Med Ethics 2016 Apr;42(4):216-219 [FREE Full text] [doi: 10.1136/medethics-2015-102663] [Medline: 25825527]
3. Allen BL, Ferrier Y, Cohen AK. Through a maze of studies: health questions and 'undone science' in a French industrial region. Environ Sociol 2016 Oct 08;3(2):134-144. [doi: 10.1080/23251042.2016.1220850]
4. Wolf GI, De Groot M. A conceptual framework for personal science. Front Comput Sci 2020 Jun 30;2:00021. [doi: 10.3389/fcomp.2020.00021]
5. White paper: design and implementation of participant-led research in the quantified self community. Quantified Self. URL: https://quantifiedself.com/white-paper-design-and-implementation-of-participant-led-research/ [accessed 2020-09-02]
6. De Groot M, Drangsholt M, Martin-Sanchez F, Wolf G. Single subject (N-of-1) research design, data processing, and personal science. Methods Inf Med 2017;56(6):416-418. [doi: 10.3414/ME17-03-0001] [Medline: 29582912]

7. Daza EJ, Wac K, Oppezzo M. Effects of sleep deprivation on blood glucose, food cravings, and affect in a non-diabetic: an N-of-1 randomized pilot study. Healthcare (Basel) 2019 Dec 25;8(1):6 [FREE Full text] [doi: 10.3390/healthcare8010006] [Medline: 31881721]
8. Lupton D. 'It's made me a lot more aware': a new materialist analysis of health self-tracking. Media Int Aust 2019 Apr 26;171(1):66-79. [doi: 10.1177/1329878x19844042]
9. Heyen NB. From self-tracking to self-expertise: the production of self-related knowledge by doing personal science. Public Underst Sci 2020 Feb;29(2):124-138 [FREE Full text] [doi: 10.1177/0963662519888757] [Medline: 31778095]
10. Mück JE, Ünal B, Butt H, Yetisen AK. Market and patent analyses of wearables in medicine. Trends Biotechnol 2019 Jun;37(6):563-566. [doi: 10.1016/j.tibtech.2019.02.001] [Medline: 30851983]
11. Karjalainen J, Viitasalo M. Fever and cardiac rhythm. Arch Intern Med 1986 Jun;146(6):1169-1171. [Medline: 2424378]
12. Kamišalić A, Fister I, Turkanović M, Karakatič S. Sensors and functionalities of non-invasive wrist-wearable devices: a review. Sensors (Basel) 2018 May 25;18(6):1714 [FREE Full text] [doi: 10.3390/s18061714] [Medline: 29799504]
13. Wac K. From quantified self to quality of life. In: Digital Health. Basel, Switzerland: Springer International Publishing; Jan 2018.
14. Clim A, Zota RD, Tinica G. Big Data in home healthcare: a new frontier in personalized medicine. Medical emergency services and prediction of hypertension risks. Int J Healthc Manag 2018 Nov 28;12(3):241-249. [doi: 10.1080/20479700.2018.1548158]
15. Zheng J, Shen Y, Zhang Z, Wu T, Zhang G, Lu H. Emerging wearable medical devices towards personalized healthcare. In: Proceedings of the 8th International Conference on Body Area Networks. 2013 Presented at: 8th International Conference on Body Area Networks; Sep 30-Oct 2, 2013; Boston, United States. [doi: 10.4108/icst.bodynets.2013.253725]
16. Mizuno A, Changolkar S, Patel MS. Wearable devices to monitor and reduce the risk of cardiovascular disease: evidence and opportunities. Annu Rev Med 2021 Jan 27;72:459-471. [doi: 10.1146/annurev-med-050919-031534] [Medline: 32886543]
17. Sana F, Isselbacher EM, Singh JP, Heist EK, Pathik B, Armoundas AA. Wearable devices for ambulatory cardiac monitoring: JACC state-of-the-art review. J Am Coll Cardiol 2020 Apr 07;75(13):1582-1592 [FREE Full text] [doi: 10.1016/j.jacc.2020.01.046] [Medline: 32241375]
18. Kourtis LC, Regele OB, Wright JM, Jones GB. Digital biomarkers for Alzheimer's disease: the mobile/ wearable devices opportunity. NPJ Digit Med 2019;2:9 [FREE Full text] [doi: 10.1038/s41746-019-0084-2] [Medline: 31119198]
19. Tyler J, Choi SW, Tewari M. Real-time, personalized medicine through wearable sensors and dynamic predictive modeling: a new paradigm for clinical medicine. Curr Opin Syst Biol 2020 Apr;20:17-25 [FREE Full text] [doi: 10.1016/j.coisb.2020.07.001] [Medline: 32984661]
20. Sen-Crowe B, McKenney M, Elkbuli A. Utilizing technology as a method of contact tracing and surveillance to minimize the risk of contracting COVID-19 infection. Am J Emerg Med 2021 Jul;45:519 [FREE Full text] [doi: 10.1016/j.ajem.2020.07.003] [Medline: 32690222]
21. Seshadri DR, Davies EV, Harlow ER, Hsu JJ, Knighton SC, Walker TA, et al. Wearable sensors for COVID-19: a call to action to harness our digital infrastructure for remote patient monitoring and virtual assessments. Front Digit Health 2020 Jun 23;2:8. [doi: 10.3389/fdgth.2020.00008]
22. Jeong H, Rogers JA, Xu S. Continuous on-body sensing for the COVID-19 pandemic: gaps and opportunities. Sci Adv 2020 Sep 2;6(36):eabd4794 [FREE Full text] [doi: 10.1126/sciadv.abd4794] [Medline: 32917604]
23. Bruggeman L. Researchers investigate whether wearable apps could unveil hidden coronavirus cases. ABC News. URL: https://abcnews.go.com/Health/researchers-investigate-wearable-apps-unveil-hidden-coronavirus-cases/story?id=69925541 [accessed 2021-02-06]
24. Quer G, Radin JM, Gadaleta M, Baca-Motes K, Ariniello L, Ramos E, et al. Wearable sensor data and self-reported symptoms for COVID-19 detection. Nat Med 2021 Jan;27(1):73-77. [doi: 10.1038/s41591-020-1123-x] [Medline: 33122860]
25. Mishra T, Wang M, Metwally A, Bogu G, Brooks A, Bahmani A, et al. Early detection of COVID-19 using a smartwatch. medRxiv 2020:20147512 (forthcoming). [doi: 10.1101/2020.07.06.20147512]
26. Natarajan A, Su H, Heneghan C. Assessment of physiological signs associated with COVID-19 measured using wearable devices. NPJ Digit Med 2020 Nov 30;3(1):156 [FREE Full text] [doi: 10.1038/s41746-020-00363-7] [Medline: 33299095]
27. Smarr BL, Aschbacher K, Fisher SM, Chowdhary A, Dilchert S, Puldon K, et al. Feasibility of continuous fever monitoring using wearable devices. Sci Rep 2020 Dec 14;10(1):21640 [FREE Full text] [doi: 10.1038/s41598-020-78355-6] [Medline: 33318528]
28. Spinuzzi C. The methodology of participatory design. Tech Commun 2005 May;52(2):163-174 [FREE Full text]
29. Sanders EB, Stappers PJ. Probes, toolkits and prototypes: three approaches to making in codesigning. CoDesign 2014 Mar 06;10(1):5-14. [doi: 10.1080/15710882.2014.888183]
30. Foth M, Axup J. Participatory design and action research: identical twins or synergetic pair? In: Expanding Boundaries in Design: Proceedings Ninth Participatory Design Conference. Canada: Computer Professionals for Social Responsibility; 2006.

31. Madey G, Freeh V, Tynan R. The open source software development phenomenon: an analysis based on social network theory. In: Proceedings of the Eighth Americas Conference on Information Systems. 2002 Presented at: Eighth Americas Conference on Information System; 2002; Dallas URL: http://melody.syr.edu/hci/amcis02_minitrack/CR/Featherman.pdf
32. Kozinets R. Netnography. In: The International Encyclopedia of Digital Communication and Society. Hoboken, New Jersey: Wiley; Feb 11, 2015.
33. Tzovaras BG, Angrist M, Arvai K, Dulaney M, Estrada-Galiñanes V, Gunderson B, et al. Open Humans: a platform for participant-centered research and personal data exploration. Gigascience 2019 Jun 01;8(6):giz076 [FREE Full text] [doi: 10.1093/gigascience/giz076] [Medline: 31241153]
34. Open Humans Community. Slack. URL: https://slackin.openhumans.org [accessed 2021-02-22]
35. Quantified Self. URL: https://quantifiedself.com [accessed 2021-02-22]
36. Open COVID19 initiative. Just One Giant Lab. URL: https://app.jogl.io/program/opencovid19 [accessed 2021-02-22]
37. OpenHumans / quantified-flu. GitHub. URL: https://github.com/openhumans/quantified-flu [accessed 2021-02-22]
38. Zhao J, Wang T, Fan X. Patient value co-creation in online health communities: Social identity effects on customer knowledge contributions and membership continuance intentions in online health communities. J Serv Manag 2015 Mar 16;26(1):72-96. [doi: 10.1108/josm-12-2013-0344]
39. Costello L, McDermott M, Wallace R. Netnography: range of practices, misperceptions, and missed opportunities. Int J Qual Methods 2017 Apr 04;16(1):160940691770064. [doi: 10.1177/1609406917700647]
40. Zainal Z. Case study as a research method. J Kemanusiaan 2007;5(1):1 [FREE Full text]
41. Preece J. Etiquette, empathy and trust in communities of practice: stepping-stones to social capital. J Univers Comput Sci 2004;10(3):294-302 [FREE Full text] [doi: 10.3217/jucs-010-03-0294]
42. Quantified Flu. URL: https://quantifiedflu.org [accessed 2021-02-22]
43. OpenHumans / qf-heartrate-apple-health. GitHub. URL: https://github.com/OpenHumans/qf-heartrate-apple-health [accessed 2021-02-22]
44. Mamykina L, Smaldone AM, Bakken SR. Adopting the sensemaking perspective for chronic disease self-management. J Biomed Inform 2015 Aug;56:406-417 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.006] [Medline: 26071681]
45. Hekler EB, Klasnja P, Chevance G, Golaszewski NM, Lewis D, Sim I. Why we need a small data paradigm. BMC Med 2019 Jul 17;17(1):133 [FREE Full text] [doi: 10.1186/s12916-019-1366-x] [Medline: 31311528]
46. Kraut R, Resnick P. Building Successful Online Communities. Cambridge, MA: The MIT Press; 2016.
47. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 2016 Mar 03;3:160011 [FREE Full text] [doi: 10.1038/sdata.2016.11] [Medline: 26938265]
48. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
49. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
50. Vaghefi I, Tulu B. The continued use of mobile health apps: insights from a longitudinal study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e12983 [FREE Full text] [doi: 10.2196/12983] [Medline: 31469081]
51. Call SA, Vollenweider MA, Hornung CA, Simel DL, McKinney WP. Does this patient have influenza? J Am Med Assoc 2005 Feb 23;293(8):987-997. [doi: 10.1001/jama.293.8.987] [Medline: 15728170]
52. Dugas AF, Valsamakis A, Atreya MR, Thind K, Manchego PA, Faisal A, et al. Clinical diagnosis of influenza in the ED. Am J Emerg Med 2015 Jun;33(6):770-775 [FREE Full text] [doi: 10.1016/j.ajem.2015.03.008] [Medline: 25827595]
53. Addeo F, Paoli A, Esposito M, Bolcato M. Doing social research on online communities: The benefits of netnography. Athens J Soc Sci 2019 Dec 05;7(1):9-38. [doi: 10.30958/ajss.7-1-1] [Medline: 22489612]
54. Jong S. Netnography: researching online populations. In: Handbook of Research Methods in Health Social Sciences. Singapore: Springer; 2019.
55. Sack W, Détienne F, Ducheneaut N, Burkhardt J, Mahendran D, Barcellini F. A methodological framework for socio-cognitive analyses of collaborative design of open source software. Comput Supported Coop Work 2006 Jun 10;15(2-3):229-250. [doi: 10.1007/s10606-006-9020-5]
56. Goodyear-Smith F, Jackson C, Greenhalgh T. Co-design and implementation research: challenges and solutions for ethics committees. BMC Med Ethics 2015 Nov 16;16:78 [FREE Full text] [doi: 10.1186/s12910-015-0072-2] [Medline: 26573410]
57. Wiggins A, Wilbanks J. The rise of citizen science in health and biomedical research. Am J Bioeth 2019 Aug 24;19(8):3-14. [doi: 10.1080/15265161.2019.1619859] [Medline: 31339831]
58. Grant AD, Wolf GI, Nebeker C. Approaches to governance of participant-led research: a qualitative case study. BMJ Open 2019 Apr 02;9(4):e025633 [FREE Full text] [doi: 10.1136/bmjopen-2018-025633] [Medline: 30944134]
59. Mahr D, Dickel S. Citizen science beyond invited participation: nineteenth century amateur naturalists, epistemic autonomy, and big data approaches avant la lettre. Hist Philos Life Sci 2019 Oct 07;41(4):41. [doi: 10.1007/s40656-019-0280-z] [Medline: 31591649]
60. Haklay M. Citizen science and volunteered geographic information: overview and typology of participation. In: Crowdsourcing Geographic Knowledge. Dordrecht, Netherlands: Springer; 2013.

61. Ferretti F. Mapping do-it-yourself science. Life Sci Soc Policy 2019 Jan 14;15(1):1 [FREE Full text] [doi: 10.1186/s40504-018-0090-1] [Medline: 30741364]
62. Delfanti A. Users and peers. From citizen science to P2P science. J Sci Commun 2010;09(01):09010501. [doi: 10.22323/2.09010501]

Abbreviations

API: application programming interface
OH: Open Humans
QF: Quantified Flu

Edited by R Kukafka; submitted 22.02.21; peer-reviewed by B Smarr, R Cruz Martínez; comments to author 27.04.21; revised version received 16.06.21; accepted 05.07.21; published 10.09.21.

Please cite as:
Greshake Tzovaras B, Senabre Hidalgo E, Alexiou K, Baldy L, Morane B, Bussod I, Fribourg M, Wac K, Wolf G, Ball M
Using an Individual-Centered Approach to Gain Insights From Wearable Data in the Quantified Flu Platform: Netnography Study
J Med Internet Res 2021;23(9):e28116
URL: https://www.jmir.org/2021/9/e28116
doi:10.2196/28116
PMID:34505836

Original Paper

Videoconferencing for Home Care Delivery in Japan: Observational Study

Hirotomo Miyatake[1*], MD; Makoto Kosaka[1,2*]; Satoshi Arita[1], MSc; Chie Tsunetoshi[1], BA; Hidehisa Masunaga[1], MSc; Yasuhiro Kotera[3], PhD; Yoshitaka Nishikawa[1], MD, PhD; Akihiko Ozaki[1,4,5], MD, PhD; Hiroyuki Beniya[1], MD

[1]Orange Home-Care Clinic, Fukui, Japan

[2]Faculty of Medicine, The University of Tokyo, Tokyo, Japan

[3]College of Health, Psychology and Social Care, University of Derby, Derby, United Kingdom

[4]Department of Breast Surgery, Jyoban Hospital of Tokiwa Foundation, Iwaki, Japan

[5]Medical Governance Research Institute, Tokyo, Japan

[*]these authors contributed equally

Corresponding Author:
Hirotomo Miyatake, MD
Orange Home-Care Clinic
Tawara 1-2-20, Fukui City
Fukui
Japan
Phone: 81 776 21 3333
Fax: 81 776 21 3353
Email: hmiyatake@orangeclinic.jp

Abstract

Background: Telemedicine has been increasingly used in many health care fields, including home care, where patients receive medical care at home. Owing to the current COVID-19 crisis, the value of telemedicine via videoconferencing is more recognized, particularly in allowing immobile patients to continue receiving care. However, the efficacy of telemedicine in home care settings in Japan remains to be fully appraised.

Objective: This study aims to identify the use and impact of telemedicine in a singular home care delivery setting in Japan.

Methods: A retrospective observational study was conducted using patient and other administrative records from a home care clinic. We considered patients who were involved in videoconferencing with home care physicians and telepresenters serving patients during 2018 and 2019. We extracted sociodemographic data of the patients and details of the videoconferencing and descriptively illustrated some specific cases.

Results: In a home care clinic in Japan, videoconferencing was conducted in 17 cases (involving 14 patients) over a 2-year period. Of all the cases, 12% (2/17) required emergency transfers and were hospitalized. A total of 88% (15/17) of cases remained; 71% (12/17) of cases were found to need extra medication or to go to a medical facility for consultation, whereas 18% (3/17) of cases were found not to be in need of urgent attention and were asked to rest. Problematic symptoms subsequently improved in 82% (14/17) of cases, and only 6% (1/17) of cases were later hospitalized.

Conclusions: Telemedicine was deemed effective for assessing patients' conditions in the home care setting in situations where home visits by a physician cannot be carried out. Our findings indicate that consultations via videoconferencing are safe and effective, suggesting more active use of videoconferencing in other clinical contexts.

(J Med Internet Res 2021;23(9):e23539) doi:10.2196/23539

KEYWORDS

telehome care; videoconference; home care; caregiver; telepresenter; mobile phone

Introduction

Background

Telemedicine, the use of information and communication technology to deliver health care at a distance, has been increasingly used to deliver health care covering a wide range of specialties, for numerous conditions, and through a variety of channels and systems [1]. An ever-increasing body of evidence suggests that care delivered via telemedicine is safe and can be as, if not more, effective and economical as face-to-face care [1]. However, one review article pointed out that the effectiveness of telemedicine could depend on conditions and specialties, and evidence on its cost-effectiveness was limited [2]. In addition, it has been suggested that applications of telemedicine in home care should be cautiously conducted, holistically weighing various individual characteristics and external factors surrounding patients [3].

Moreover, significant barriers to wider use have been identified: health care providers and patients face regulatory ambiguities and financing uncertainties. Inequalities in both health and digital literacy are another concern, as patients who could benefit the most are often those who are the least able to have access to or make use of telemedicine [4]. Telemedicine has evolved to help improve effectiveness, efficiency, and equity in health care. However, it can also create risks, amplify existing inequalities, and worsen health care delivery.

Japan has been actively promoting telemedicine as a means to help meet the evolving health care needs of the population. Key areas of focus have been (1) access to health care for rural and remote communities, (2) home telemedicine, and (3) illness prevention and lifestyle modification. Evidence on the safety and efficacy of telemedicine in various situations has been presented at conferences [5]. The Japanese government has been encouraging telemedicine, especially as part of a strategy focusing on an effective home-based primary care approach, in an attempt to move away from the nation's famed lengthy hospitalization practice [6]. In 2015, Japan's Ministry of Health, Labor and Welfare (MHLW) permitted the use of telemedicine in areas other than rural and remote areas [7] and, in 2018, allowed medical treatment fees for remote examinations to be partially covered under the national health insurance scheme. However, provider-to-patient telemedicine services are only permitted after an initial face-to-face meeting has taken place between the physician and the patient [8]. It is the sole responsibility of the physician to decide whether remote health care is safe and appropriate.

Objectives

Only a few telemedicine studies in Japan have been published, mostly in the Japanese language, limiting the global reach of the findings [9]. An in-depth appraisal of the efficacy of telemedicine is particularly important in home care settings, as there are many factors that could prohibit physicians from physically visiting patients' homes immediately, beyond planned visits and in emergency situations. Telemedicine via videoconferencing is a possible means to help overcome this problem. However, its effectiveness and safety in assessing patients' conditions, including identifying potential acute illnesses and potential health emergencies, has not been evaluated in the home care setting in depth. Accordingly, the main goal of this study is to demonstrate some cases using videoconferencing and discuss its efficacy and safety and to identify some key elements in the home care setting in Japan.

Methods

Overview of Japanese Health Insurance System, Home Care, and Remote Examination

Home care in Japan is provided and paid for within the Japanese health insurance system. The Japanese government has promoted home care as a part of the Integrated Community Care System, which requires cooperation between home care providers [10] to reduce medical costs and overcome constraints, such as the number of physicians in Japan (2.5 per 1000 inhabitants), which is the lowest among the Organization for Economic Cooperation and Development countries [11].

Integrated Community Care System

The Integrated Community Care System, which is designed to provide medicine, nursing, and life support in an integrated format, has been developed to enable some patients, particularly older adults, to continue living at home in the manner to which they have become accustomed [10]. Home care is indicated only when the physician determines that it is applicable. There are no clear criteria, but *patients who are at least unable to walk to the clinic on their own without the help of family members or caregivers* is a requirement, and there are no age restrictions. Therefore, the patients involved in this research were those living in specialized housing facilities for older adults where a formal caregiver was available and those who were regularly visited by their physician, living in their own residences with the assistance of informal family caregivers.

Cooperation Between Home Care Physicians and Nurses and Pharmacists

In Japanese home care, patients receive nursing services from institutional nurses or visiting nurses under the medical insurance system or the long-term care insurance system. Some specialized caregiving facilities use full-time institutional nurses. Visiting nurses belong to clinics, hospitals, or nurse stations and provide physical and mental health services, including medical care and basic life support services. However, any physical intervention, such as management of intravenous injections or blood sampling, can only be administered under the supervision of the patient's physician. The clinic provides a visiting nurse service; however, it is up to patients whether they make use of the service.

In addition, pharmacies are counted as facilities that provide medical services in the Integrated Community Care System. Some pharmacies or clinics dispatch community pharmacists and provide indications for drug management to home care patients [11].

Remote Examination in the Japanese Home Care Setting

The MHLW has created strict definitions for what constitutes a *remote examination*, in which physicians deal with patients

via a digital audiovisual link, such as via a videophone or the internet. Remote examinations were conducted in the same manner as face-to-face examinations. Until recently, the first examination must have been face-to-face; however, because of COVID-19, remote examination can now be used as the first examination session under certain conditions. As of April 2020, all physicians who intend to use videoconferencing are required to undertake training regulated by the MHLW.

Doctor to patient with nurse has been introduced as a common form of implementing remote examinations, in which a physician can see patients receiving home care via a digital audiovisual connection but only with a visiting nurse attending the patient. In this situation, visiting nurses are supposed to provide medical assistance within the range expected in the treatment plan and visiting nursing instruction. In this study, all remote examinations were conducted as a substitute for unplanned physician visits. Although doctor to patient without nurse is also permitted as a form of remote examination, most of the informal caregivers or patients involved in this study requested the presence of nurses.

In Japan, physicians are primarily responsible for any intervention, either directly or in an oversight capacity, conducted as part of any remote examination or intervention. If a physician determines that they are unable to perform an adequate remote examination, the physician must immediately discontinue the remote examination and arrange a face-to-face examination. Physicians are also required to take adequate security measures to ensure that patients' medical information is not compromised.

Setting and Participants

The Orange Home Care Clinic specializes in providing medical care to patients based on their own homes or residences in Fukui City, Japan. This clinic has five working teams, consisting of 1 physician and 1 clerk per team, treating an average of 330 patients annually. Several nurses have been employed to support the smooth operation of the clinic, such as coordination with other visiting nurse stations. Therefore, regular nursing care is provided by other visiting nurse stations. Each team visits 8-10 patients a day on weekdays. The visiting area covers a radius of 16 km from the clinic. Owing to various unforeseen circumstances (eg, natural disasters), the clinic used videoconferencing in 17 instances from January 2018 to December 2019.

Data Collection

Basic information on patients, such as age, sex, primary diseases, and availability of caregivers, was collected from medical records at the clinic. Data on cases where videoconferencing was used, such as the reason for the consultation, the telepresenter, any intervention, and the outcome, were collected from medical records and other documentation.

The availability of a caregiver was categorized as (1) formal caregiver (a patient living in a specialized housing facility where an official caregiver is present), (2) informal caregiver (patient living with a family member), and (3) no caregiver (a patient living alone). We used this classification because the feasibility of videoconferencing is supposed to depend on the living environment and the demographics in Japan have been changing rapidly, and the number of single-person households, especially among older adults, has been increasing significantly [12].

In this study, videoconferencing was performed when a caregiver or a visiting nurse decided that it was necessary to consult with a physician when visiting the patient. Videoconferencing was performed as an alternative to an emergency home visit because the physician could not visit immediately. The reason for videoconferencing consultation involved the chief complaint when a consultation was requested by the patient or caregiver. The telepresenter was the person with the patient who orchestrated the remote videoconferencing connection with the physician. Telepresenters need to have adequate knowledge and skill with respect to information technology devices and a proper understanding of the patient's status. They can be informal caregivers, formal caregivers, and visiting nurses; however, we defined the telepresenter as *no one* when the patients facilitated videoconferencing by themselves. The intervention referred to any medical procedure, which was determined to be necessary during videoconferencing and before the next planned visit could be undertaken. Although the options of interventions depend on the presence of nurses with patients, interventions such as applying drugs to patients are applicable even if patients contact the physician by themselves. The outcome was categorized as (1) symptom improvement when the patients did not need additional treatment by physicians other than intervention decided through videoconferencing, (2) patients required another unplanned visit by a home care physician before the next planned visit, or (3) patients required hospitalization (when the patient was hospitalized after videoconferencing).

Ethical Review

This study adheres to the Ethical Guidelines for Medical and Health Research Involving Human Subjects and meets Japan's local legal requirements. Study approval was granted by the Ethics Committee of the Medical Governance Research Institute on June 4, 2020 (MG2018-18-20200604).

Results

Participant Statistics

Table 1 shows a summary of all the cases included in this study.

Table 1. Sociodemographic details of the patients and videoconference information.

ID	Age (years)	Sex	Primary disease	Availability of caregiver	Reasons for remote consultation	Telepresenter	Diagnosis	Intervention	Outcome
1-1	97	Female	Stomach cancer	Formal caregiver	Fall	Visiting nurse	Suspect of bone fracture	Prescription of pain reliever	Symptom improvement
1-2	97	Female	Stomach cancer	Formal caregiver	Phlegm	Visiting nurse	Bacterial infection	Prescription of antibiotics	Symptom improvement
2	92	Female	Dementia	Formal caregiver	Vomiting	Visiting nurse	Gastroenteritis	Prescription of antiemetic drug	Symptom improvement
3	93	Male	Dementia	Informal caregiver	Scratch	Visiting nurse	Eczema	Applying ointment	Symptom improvement
4	70	Female	Last stage of stomach cancer	Informal caregiver	Headache	Visiting nurse	Muscle contraction headache	Treatment in acupuncture clinic	Symptom improvement
5	66	Female	Last stage of lung cancer	Informal caregiver	Uncomfortable feeling	Visiting nurse	Dehydration	Infusion	Symptom improvement
6	35	Male	Cerebral contusion	Informal caregiver	Suspect of drug eruption	Visiting nurse	Drug eruption	Stopping antibiotics	Symptom improvement
7	33	Male	Acquired cerebral palsy	Informal caregiver	Suspect of low-temperature burn	Informal caregiver	Low-temperature burn	Applying ointment	Symptom improvement
8	16	Male	Neurofibromatosis type 1	Informal caregiver	Fever	Visiting nurse	Acute bronchitis	Prescription of antibiotics	Symptom improvement
9	11	Male	West syndrome	Informal caregiver	Epileptic seizures	Institutional nurse	Epileptic seizures	Watch and wait	Symptom improvement
10	7	Male	Cerebral palsy	Informal caregiver	Swelling of the eyelid	Institutional nurse	Hordeolum	Prescription of eye-drops	Symptom improvement
11-1	4	Female	Double outlet right ventricle	Informal caregiver	Suspect of CV[a] block	Visiting nurse	CV block	Emergency transfer	Hospitalization
11-2	4	Female	Double outlet right ventricle	Informal caregiver	Vomiting and diarrhea	Visiting nurse	Dehydration	Emergency transfer	Hospitalization
12	3	Female	Lissencephaly	Informal caregiver	Tachycardia	Visiting nurse	None	Watch and wait	Symptom improvement
13-1	2	Male	Cerebral palsy	Informal caregiver	Increase of phlegm	Informal caregiver	Upper respiratory inflammation	Watch and wait	Symptom improvement
13-2	2	Male	Cerebral palsy	Informal caregiver	Fever	Visiting nurse	Middle ear inflammation	Continuing antibiotics	Hospitalization
14	72	Female	Rheumatoid arthritis	No caregiver	Eczema	Visiting nurse	Miliaria	Prescription of ointment	Symptom improvement

[a]CV: cardiovascular.

Remote videoconferencing was conducted in 17 instances involving 14 patients (3 patients were involved in videoconferencing twice). Of the 14 patients, 5 (36%) were children and 5 (36%) were older adults (aged >65 years). Approximately 43% (6/14) of patients were male. According to the International Classification of Diseases 11th categorization, the most common primary disease was that of the nervous system (5/14, 36%), followed by neoplasms (3/14, 21%) and mental, behavioral, or neurodevelopmental disorders (2/14, 14%). About 79% (11/14) of patients lived with family members, 14% (2/14) lived in assisted-living facilities, and 7% (1/14) lived alone.

The reasons for videoconferencing consultation and diagnoses are shown in Table 1. In 77% (13/17) of instances, visiting nurses coordinated the videoconferencing, whereas, in 12% (2/17) of instances, institutional nurses performed that task. In 12% (2/17) of instances, an informal caregiver handled the videoconferencing.

As for subsequent interventions, a new prescription for medicine followed in 35% (6/17) of videoconferences; the physician faxed the prescription to the community pharmacist, and the community pharmacist delivered the new medicine to the patients' residences. In 18% (3/17) of instances, *watch and wait* until the next planned visit instructions and the guidance on the application of medicine were given. In 12% (2/17) of instances, the physician instructed patients to take the medicine in hand. In contrast, in 12% (2/17) of instances, emergency transfer to hospital was advised; the physician arranged the emergency transfers and passed the patient information to the attending physicians in the hospital who provided emergency treatment.

With regard to outcomes, 82% (14/17) of cases with the use of videoconferencing resulted in improvement of the problematic symptoms before the next planned visit was conducted within the following 2 weeks, and the remaining 18% (3/17) of cases resulted in hospitalization. One hospitalization was because of pneumonia from an upper respiratory infection that occurred independently after the videoconferencing had taken place.

In ID 1-2, although a formal caregiver was available, albeit a part-time worker, the clinic was called, and the physician asked a visiting nurse to visit the patient's residence as the physician was not able to visit the residence immediately. The visiting nurse subsequently organized and conducted the videoconferencing. In IDs 6 and 8, an informal caregiver was available, and he or she summoned a visiting nurse who organized and conducted the videoconferencing. In ID 7, an informal caregiver was available and organized and conducted the videoconferencing.

Indicative Videoconferencing Examples

ID Number 1-2

The patient was a 97-year-old female with early-stage gastric cancer and chronic heart failure. She lived in a long-term facility with no full-time nurses in attendance but with part-time nurses, with nonmedical informal caregiver helpers providing various care services to the residents. The patient was able to walk using a walker and take a bath with the support of helpers. As she was unable to visit the clinic by herself because of lower-extremity muscle weakness and as her gastric cancer was relatively stable, her medical home care started in February 2015, and the team visited her twice monthly. She wished to spend the rest of her life in the facility. One night in October 2018, the facility's part-time nurse contacted us as the patient had an approximate 2-week history of excess phlegm and breathlessness. The last regular visit by a home care physician was 12 days previously when she had a hoarse voice and was taking regular herbal medicine, as she presumed that she had caught a cold. She was able to eat and did not feel seriously ill; therefore, a home care physician decided to follow up and told her to contact the clinic if her condition worsened. Her body temperature was 37.3 °C, blood pressure was 155/92 mm Hg, pulse rate was 80 bpm, saturation of percutaneous oxygen was 94%, and the state of consciousness was clear. Although a home care physician attempted to visit the facility, the physician was not able to do so immediately. Therefore, the physician ordered a visiting nurse from the clinic to visit the patient's facility and instead carried out a videoconferencing. The nurse told the physician that there were no obvious lung noises, which suggested that pneumonia was not present, and her body movement, such as standing and sitting, was smooth. The physician carefully observed the patient's complexion and respiratory status through videoconferencing and reconciled the findings with those reported by the nurse. As a result of the videoconferencing examination, the physician ordered the nurse to take a blood test to judge whether there was an acute bacterial respiratory infection and deterioration of chronic heart failure. The nurse returned the blood sample to the clinic and submitted it for testing. The blood test indicated that the white blood cell (WBC) count was 5180 per μL and C-reactive protein (CRP) was 0.12 mg/dL, and the physician diagnosed an acute viral respiratory infection based on the results of the blood test. The physician prescribed a traditional Chinese medicine to reduce her symptoms and then faxed the prescription to a specific pharmacy. The community pharmacist delivered the drug to the patient's assisted-living facility and started the treatment with the traditional Chinese medicine (syoseiryuutou 9 g 3 times a day for 5 days). Two days later, the physician confirmed an increase in N-terminal probrain natriuretic peptide from the blood test taken by the visiting nurse, and the deterioration of chronic heart failure was diagnosed, and the treatment with a diuretic drug (furosemide 20 mg once per day for 7 days) was started**.** The community pharmacist contacted the patient, delivered the drug to the facility, and administered the first dose. In this case, the nurse being physically present with the patient during the remote examination allowed blood tests to be completed, which allowed the diagnosis of an acute viral respiratory infection and deterioration of the patient's cardiac problem, thereby allowing early and customized treatment.

ID Number 8

The patient was a 16-year-old male with neurofibromatosis type 1. He had a gastrostomy to be fed and underwent noninvasive positive pressure ventilation every night. He spent most of the daytime using a wheelchair and lived in his house with the support of his parents. His twice-monthly home care started in February 2011 to help reduce the burden on his family, who previously had to frequently visit emergency medical departments because of his regular fevers. One morning in September 2018, the patient's family reported by telephone that the patient had a fever of 37.8 °C with excess phlegm; however, his respiratory condition was not troublesome. As the condition was not serious but needed careful monitoring, the physician told the family to observe with an antipyretic drug and water intake. In addition, the physician ordered a visiting nurse to visit the patient's residence to check the patient's condition, as he had acute viral bronchitis with bacterial secondary infection and was treated with an antibiotic 1 month before this consultation. When the visiting nurse visited the patient's residence that afternoon, his temperature was 37.8 °C, although an antipyretic drug had been administered by his informal caregivers approximately 3 hours previously, and thus, the nurse decided to conduct a videoconferencing. The patient's condition was similar to that of acute bronchitis that he had experienced 1 month before, which had been resolved following a visit by the team 2 weeks earlier. Therefore, the physician diagnosed acute viral bronchitis with bacterial secondary infection based

on the clinical symptoms and the patient's past history and prescribed a course of antibiotics—sultamicillin tosylate dihydrate 9 g after every meal for 7 days. The prescription was faxed to a specialized pharmacy, and the patient's mother went to the pharmacy to get the medication and administered it as per instructions. In the meantime, the visiting nurse kept an eye on the patient. The physician performed an emergency home visit 2 days later because the patient's mother called the clinic as the fever continued. At that visit, as the physician suspected pneumonia, the physician performed a sputum test and a blood test to decide whether his antibiotic drug had to be changed. The results of the laboratory test showed a WBC count of 9410 per μL and CRP of 3.14 mg/dL. The physician judged that his infection was getting past the acute stage as the CRP level was mildly elevated; however, the WBC and other data were within normal limits. The physician decided not to change the antibiotic and finish the 7-day treatment because adolescents with special care needs are prone to viral infections and secondary bacterial infections, and the infections are often severe. The patient's condition improved 1 week later, and a new bacterial culture of sputum did not reveal any infection. In this case, remote medical examination was useful to diagnose possible acute viral bronchitis with bacterial secondary infection based on the patient's past history, and action could be taken to respond to the new infection.

ID Number 6

The patient was a 35-year-old bedridden male patient who had a postoperative left acute subdural hematoma caused by a traffic accident. He had undergone a number of medical procedures, including a tracheostomy, 2 gastrointestinal stomata, a gastronomy, and insertion of a ventral venous catheter. The two gastrointestinal stomata were used to drain intestinal fluids because he developed nonocclusive mesenteric ischemia and had most of his intestinal tract removed in 2017. For that reason, he underwent a gastrostomy and used a central venous catheter to obtain nutrition. His parents provided his daily care, and home-visiting nurses made two regular visits and provided regular medical attention daily. The clinic also provided weekly medical home care from March 2018. In April 2018, the visiting nurse telephoned the clinic to report that the patient had had fever for a few days, and the physician told the nurse to take blood for testing. From the results—increased WBC (10,210 per μL with a left shift) and CRP (5.65 mg/dL)—the patient was assessed to have an acute bacterial infection. The clinic staff brought sulbactam sodium and ampicillin sodium from the clinic's stocks to the patient's residence, and the physician told the nurse to administer it during regular morning and evening visits because there was no allergic information about the patient from the referral hospital. A few minutes after the intravenous injection of the antibiotic, eczema appeared on the patient's chest, indicating an allergic reaction. Therefore, a remote visual medical examination was conducted to check the patient's condition. No other abnormal findings were observed, such as respiratory distress or decrease in saturation of percutaneous oxygen, and the eczema disappeared 30 minutes after the intravenous injection. The suspected allergic reaction reminded the patient's father of the patient's previous history of suspected allergic reactions, and he reported that the same reaction was observed when the patient was hospitalized and that most of the antibiotics gave the patient redness. As the referral from the hospital stated that the patient had no allergies, the home care physician contacted the hospital; however, there was no information about redness in the hospital records. Thus, they discontinued sulbactam sodium and ampicillin sodium because these drugs were suspected as allergens and instead initiated ceftriaxone sodium hydrate. As such, the remote medical examination was useful for diagnosing possible allergic reactions and helped to change the antibiotics being administered.

ID Number 7

A 33-year-old male patient was bedridden because of cerebral palsy and had undergone a gastrostomy. The twice-monthly home care service started in December 2014 because the patient's family had to regularly take him to a hospital because he was regularly developing a fever. In December 2018, the family called the clinic to report that there was redness in the joints of his left hand, which seemed to be a low-temperature burn. The physician made a remote visual examination using the family's mobile phone to ascertain whether a home visit was required to deal with the situation. In discussion with the family, the physician confirmed that the red lesion appeared after using a hot-water bottle to keep the body warm. His left hand had been in prolonged contact with the hot-water bottle; however, there was no pain, only redness. Therefore, the physician decided to treat the scald using steroid ointment, which was originally prescribed to be applied to the granulation around the gastrostomy and so was available at the patient's home, and told the family to contact the clinic if there was any worsening of the patient's condition. After 2 weeks, the scald healed at the time of the next regular visit. Remote medical treatment was useful to diagnose a low-temperature burn and start treatment immediately without the need for a visiting nurse.

Discussion

Principal Findings

Treatments following videoconferencing in the clinic were generally safe and effective. As shown in Table 1, for 77% (13/17) of cases, the problematic symptom causing the need for a videoconferencing improved without emergency transfer or treatment in a medical facility. No intervention by medical professionals was required in 29% (5/17) of instances. The use of videoconferencing was deemed effective for physicians to reduce unnecessary visits, whereas in some cases, more nurse visits were required.

One key element for the successful use of videoconferencing was collaboration between medical staff in the prevailing situation. In Japan, specialized home medical care is overseen by physicians in specialized home care clinics. Other medical staff, such as visiting nurses, community pharmacists, and rehabilitation staff, work in a system whereby the corps of local visiting medical specialists provide more regular home care. A shared notebook stored in the patient's residence is used to update all clinical information for use by medical professionals involved in the provision of care to coordinate their work and actions. In addition, when medical staff detect any significant adverse signs in the patient's physical condition while visiting

residences, they arrange a teleconference with other medical staff, including physicians, immediately. The videoconferencing enabled visiting nurses to implement prompt blood tests necessary for diagnosis and treatment and subsequently facilitate a physician's judgment in any necessity for intervention to deal with the patients' condition. In fact, Funderskov et al [13] showed that the use of video consultation allows community-based nurses to discuss their observations, enabling them to take part in active home care, including palliative care. A comparison of our findings with those in other countries is difficult because of the Japanese health context, where physicians predominantly have governing control in patients' care. However, in home care and remote examination, physicians have to rely on the judgments and actions of other medical staff to implement timely and appropriate medical interventions.

Videoconferencing allows the analysis of outward appearance of signs and symptoms. Interviews with patients and clinicians in previous studies showed that video consulting was better than telephone consulting, especially in consultations involving psychological assessment where visual cues are important, as illustrated by the cases of IDs 1-2 in this study [14]. With regard to the feasibility and accuracy of diagnosing through videoconferencing, according to a review by Trettel et al [15], most of the articles on teledermatology reported that telemedicine was feasible, reliable, and effective under various conditions. Furthermore, the review noted that in 90.3% of the cases, clinicians rated that the experience of conducting a telehealth session was equivalent to or better than a home visit and that the communication environment (eg, internet reception) was the important factor in telehealth sessions [16]. Japan has a reliable digital communication system: in our cases, all the videoconferencing with the patient's homes was done using 4G connections, and the communication channel was never interrupted or lost.

However, there are several critical points that need to be considered with respect to incorporating telemedicine in the home care setting. First, the telepresenter, the person who organizes the videoconferencing in the patient's location, is critical. It is ideal that the main caregivers, including informal caregivers, are able to orchestrate videoconferencing, although this might be a heavy burden for informal caregivers. Clinically, telepresenters are chosen considering their availability, ability, and digital literacy. Future research is expected to clarify whether informal caregivers can play the role of telepresenters appropriately and what kind of support or intervention is required to overcome this problem. Furthermore, in addition to medical literacy, the availability of suitable equipment, technological literacy of patients, their caregivers, attendant nursing staff, and attending physicians, as well as costs, have to be taken into account. A study of telemedicine between nurses and patients at home in Norway highlighted the importance of the levels of training and experience with respect to videoconferencing [17]. In particular, older people are not accustomed to using digital devices, and many do not even own or have access to computers or sophisticated mobile phones. According to Japan's Ministry of Internal Affairs and Communications, the proportion of internet users in the country is approximately 20% in those aged >80 years, whereas it is >90% in people aged 20-40 years [18]. It might be helpful to instruct all home care caregivers or nurses on how to use communication devices in telemedicine. A previous study recommended supporting digital literacy and confidence of older people to strengthen their intention to use technology [19]. Considering the increase in the number of older patients who live alone in Japan, developing digital skills among older adults is also beneficial to create more space in the limited caregiving facilities. Therefore, policies may have to be put in place to accommodate the provision of medical home care to older patients who live alone.

Second, the cost of communication devices should be considered. Some patients do not have any equipment enabling videoconferencing, which poses significant constraints on developing a videoconferencing telemedicine system. This endangers one of the primary aims of videoconferencing implementation—equity. In fact, the Ministry of Internal Affairs and Communications reported that the internet use rate in Japan was >54.2% in households, with an annual income of <¥2 million (<US $19,000), whereas it was >85% in households with >¥4 million (>US $38,000) [20]. The cost of internet use depends on the type of service used. For example, in cases of our study, because the smartphones of patients, family caregivers, or clinics were used in videoconferencing, no additional cost other than communication costs paid by patients was needed. Commonly in Japan, medical institutions use remote examination platforms or phone calls when conducting remote examinations [21]. However, it is uncommon that the Japanese national insurance system covers the cost of provision of adequate devices, training in their use, the videoconferencing connection, and the administration of setting up videoconferencing. It is important to encourage patients or caregivers to, wherever and whenever possible, be able to use modern communication equipment, enabling them to use videoconferencing or telemedicine in times of emergency or whenever a physician is unable to make a physical visit. Another option is to change the system so that nurses and paramedics can conduct medical interventions without referring to physicians.

Third, telemedicine may miss some important clinical information that could be obtained in a face-to-face consultation. In addition, orchestrating sample collection, transport for testing, and analyses also takes longer than it would in a hospital or clinic environment, and the inability to use specialized equipment to aid diagnoses, which would often be available in medical facilities, makes home care less than ideal. Of note, we should be aware that the use of telemedicine in emergency cases in a home care setting is still in its infancy, and a careful and gradual evolution is imperative. It is essential for physicians and other health care professionals to perform a follow-up, in-person evaluation following any telemedicine event, as is currently done. It is also important to err on the side of caution when carrying out emergency telemedical consultations by calling for emergency transfers in cases possibly requiring rapid interventions or access to specialized physicians, facilities, or equipment.

Limitations

The sample size of this study was small and diverse, and the study findings should be interpreted with caution. We could not compare the difference in the feasibility of conducting videoconferencing, depending on the living environment; however, formal caregivers may facilitate videoconferencing better because of their knowledge and experience of using videoconferencing. Moreover, considering the increase in older patients who live alone in Japan, the costs involved, and the scarcity of suitable accommodation, it is difficult for them to move into caregiving facilities. Therefore, new strategies and policies should be developed to ensure the provision of a suitable level of care as Japan moves toward increased provision of home care for patients.

Owing to the notification from the MHLW concerning remote examination, it is expected for videoconferencing in home care to increase in Japan in the foreseeable future. Troubles may manifest when considerably more examinations are conducted at the same time. In particular, as we discussed, the age of patients and caregivers can affect the use of videoconferencing, as observed in this study. Future research should be conducted with larger sample sizes to examine the reliability of our findings through stratified analyses.

Conclusions

Consultations using videoconferencing in clinics are generally safe and effective**.** Telemedicine is a valuable substitute for face-to-face investigations where necessary or desirable. This is a particularly valuable option in emergency situations, especially when home care health personnel cannot access the patients under their care. Japan has recently introduced the concept of primary care and has introduced training and qualifications in primary care. This will have a major impact on home care options and practice, and the importance of and potential for videoconferencing will be further enhanced accordingly. Furthermore, the findings of our study have special implications in the COVID-19 era. Considering the vast and long-lasting impacts of the current global COVID-19 pandemic, where physical visits are restricted, the findings in this study can help health care workers, patients, and their families involved in home care to appraise the characteristics of this relatively new approach.

Acknowledgments

The authors express their sincere gratitude to all the clinic staff members, patients, and caregivers involved in this study. They would like to thank Professor Andy Crump for his expertise and guidance in the production of this paper. This research was supported by Yuumi Memorial Foundation for Home Health Care.

Authors' Contributions

HM and MK analyzed the data and wrote the manuscript. All authors conceptualized and designed the study and revised the manuscript.

Conflicts of Interest

AO received a personal fee from MNES Inc, outside the submitted work. YN received a personal fee from MRT Inc, outside the submitted work.

References

1. Hashiguchi TC. Bringing health care to the patient: an overview of the use of telemedicine in OECD countries. In: OECD Health Working Papers. Paris: OECD Publishing; 2020.
2. Eze ND, Mateus C, Hashiguchi TC. Telemedicine in the OECD: an umbrella review of clinical and cost-effectiveness, patient experience and implementation. PLoS One 2020 Aug 13;15(8):e0237585 [FREE Full text] [doi: 10.1371/journal.pone.0237585] [Medline: 32790752]
3. Shigekawa E, Fix M, Corbett G, Roby DH, Coffman J. The current state of telehealth evidence: a rapid review. Health Aff (Millwood) 2018 Dec;37(12):1975-1982. [doi: 10.1377/hlthaff.2018.05132] [Medline: 30633674]
4. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017 Apr 27;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
5. Japanese Journal of Telemedicine and Telecare Vol 15 Supplement (Japanese). Japanese Telemedicine and Telecare Association. 2020. URL: https://plaza.umin.ac.jp/~jtta/pdf/book/vol15no3.pdf [accessed 2021-08-23]
6. [Telemedicine in Japan 2013]. Japanese Telemedicine and Telecare Association. URL: http://jtta.umin.jp/pdf/telemedicine/telemedicine_in_japan_20131015-2_jp.pdf [accessed 2020-10-06]
7. [About examination using internet communication devices]. Ministry of Health, Labour and Welfare of Japan. URL: https://www.mhlw.go.jp/file/06-Seisakujouhou-10800000-Iseikyoku/0000094452.pdf [accessed 2020-08-06]
8. [Guidelines for appropriate online examination practice]. The Ministry of Health, Labour and Welfare, Japan. URL: https://www.mhlw.go.jp/content/000534254.pdf [accessed 2020-08-06]

9. Ito J, Edirippulige S, Aono T, Armfield NR. The use of telemedicine for delivering healthcare in Japan: systematic review of literature published in Japanese and English languages. J Telemed Telecare 2017 Dec;23(10):828-834. [doi: 10.1177/1357633X17732801] [Medline: 29081269]
10. [Integrated community care system]. The Ministry of Health, Labour and Welfare, Japan. URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/hukushi_kaigo/kaigo_koureisha/chiiki-houkatsu/ [accessed 2020-07-23]
11. Onda M, Imai H, Nanaumi Y, Hirano A, Fujii S, Arakawa Y. [Nationwide study on relations between workload and outcomes of home visiting service by community pharmacists]. Yakugaku Zasshi 2015;135(3):519-527 [FREE Full text] [doi: 10.1248/yakushi.14-00220] [Medline: 25759060]
12. [Housing for the elderly]. The Ministry of Health, Labour and Welfare, Japan. URL: https://www.mhlw.go.jp/seisaku/2009/03/01.html [accessed 2020-10-06]
13. Funderskov KF, Danbjørg DB, Jess M, Munk L, Zwisler AO, Dieperink KB. Telemedicine in specialised palliative care: healthcare professionals' and their perspectives on video consultations-a qualitative study. J Clin Nurs 2019 Nov;28(21-22):3966-3976. [doi: 10.1111/jocn.15004] [Medline: 31328336]
14. Hammersley V, Donaghy E, Parker R, McNeilly H, Atherton H, Bikker A, et al. Comparing the content and quality of video, telephone, and face-to-face consultations: a non-randomised, quasi-experimental, exploratory study in UK primary care. Br J Gen Pract 2019 Aug;69(686):595-604 [FREE Full text] [doi: 10.3399/bjgp19X704573] [Medline: 31262846]
15. Trettel A, Eissing L, Augustin M. Telemedicine in dermatology: findings and experiences worldwide - a systematic literature review. J Eur Acad Dermatol Venereol 2018 Feb;32(2):215-224. [doi: 10.1111/jdv.14341] [Medline: 28516492]
16. Taylor A, Morris G, Pech J, Rechter S, Carati C, Kidd M. Home telehealth video conferencing: perceptions and performance. JMIR Mhealth Uhealth 2015 Sep 17;3(3):e90-e9- [FREE Full text] [doi: 10.2196/mhealth.4666] [Medline: 26381104]
17. Rygg L, Brataas H, Nordtug B. Introducing videoconferencing on tablet computers in nurse-patient communication: technical and training challenges. Int J Telemed Appl 2018;2018:8943960 [FREE Full text] [doi: 10.1155/2018/8943960] [Medline: 30420881]
18. [White paper information and communications in Japan]. The Ministry of Internal Affairs and Communications, Japan. URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h30/html/nd142110.html [accessed 2020-07-23]
19. van Houwelingen CT, Ettema RG, Antonietti MG, Kort HS. Understanding older people's readiness for receiving telehealth: mixed-method study. J Med Internet Res 2018 Apr 06;20(4):e123 [FREE Full text] [doi: 10.2196/jmir.8407] [Medline: 29625950]
20. [White paper information and communications in Japan]. The Ministry of Internal Affairs and Communications, Japan. URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h30/html/nd252120.html [accessed 2020-09-18]
21. [Verification of the results of telephone and online medical services from April to June 2020]. The Ministry of Health, Labour and Welfare, Japan. URL: https://www.mhlw.go.jp/content/10803000/000657020.pdf [accessed 2020-09-27]

Abbreviations

CRP: C-reactive protein
MHLW: Ministry of Health, Labor and Welfare
WBC: white blood cell

Edited by R Kukafka; submitted 17.08.20; peer-reviewed by A Taylor, E Donaghy, I Mircheva, CC Lin; comments to author 17.09.20; revised version received 05.11.20; accepted 21.06.21; published 01.09.21.

Please cite as:
Miyatake H, Kosaka M, Arita S, Tsunetoshi C, Masunaga H, Kotera Y, Nishikawa Y, Ozaki A, Beniya H
Videoconferencing for Home Care Delivery in Japan: Observational Study
J Med Internet Res 2021;23(9):e23539
URL: https://www.jmir.org/2021/9/e23539
doi:10.2196/23539
PMID:34468333

Original Paper

Primary Care Psychiatry eConsults at a Rural Academic Medical Center: Descriptive Analysis

Jade Avery[1*], MD; Dennis Dwan[2*], MD; Gillian Sowden[3], MD; Matthew Duncan[3], MD

[1]Department of Psychiatry, New York Presbyterian, Columbia University, New York City, NY, United States

[2]Carney Hospital, Dorchester, MA, United States

[3]Department of Psychiatry, Dartmouth-Hitchock Medical Center, Lebanon, NH, United States

[*]these authors contributed equally

Corresponding Author:
Matthew Duncan, MD
Department of Psychiatry
Dartmouth-Hitchock Medical Center
1 Medical Center Drive
Lebanon, NH, 03756
United States
Phone: 1 6036504725
Email: matthew.s.duncan@hitchcock.org

Abstract

Background: Primary care providers serve a crucial role in addressing the mental health needs of many patients. However, there are times when input from a psychiatric specialist may be helpful in supporting the mental health care provided in primary care. Psychiatry eConsults can serve as a valuable tool in providing specialist advice for primary care physicians when direct referral to specialty care is not readily available.

Objective: The goal of this study is to evaluate the content and implementation of psychiatric eConsults by primary care providers in a rural academic medical center.

Methods: This is a retrospective review of 343 eConsults placed between May 2016 and February 2019 by primary care providers at a single academic medical center. The content of eConsult requests, including patient diagnosis, consult question type, specialist recommendations, patient demographics, the distance of patient and primary care providers from the consulting provider, rate of implementation of the recommendation, and response time, were analyzed.

Results: The most common diagnoses associated with eConsults were depression (162/450, 36%) and anxiety (118/450, 26%). The most commonly asked eConsult question was regarding medication management, including medication choice, side effects, interactions, and medication taper (288/343, 84%). More than one recommendation was included in 76% (259/343) of eConsults, and at least one recommendation was implemented by the primary care provider in 94% (282/300) of eConsults. The average time to respond to an eConsult was 26 hours.

Conclusions: This study demonstrates that psychiatry eConsults can be conducted in a timely manner and that primary care providers implement the recommendations at a high rate.

(J Med Internet Res 2021;23(9):e24650) doi:10.2196/24650

KEYWORDS

electronic consultation; psychiatry; eConsult; telehealth; telemedicine; rural

Introduction

Mental illness is especially prevalent, with nearly 1 in 5 adults meeting criteria for a psychiatric disorder [1]. Untreated mental illness has been associated with decreased functionality, reduced quality of life, heightened physical health complications, and premature mortality [2]. Although the demand for mental health treatment has been increasing, studies have demonstrated a shortage of mental health care clinicians, limited access to mental health care [3], and extensive wait times [4].

Patients in rural settings (in particular) face additional challenges in accessing mental health services. Rural populations are also faced with geographic isolation and greater stigmatization of mental health conditions [5]. Residents of small towns that are

geographically far from larger metropolitan areas report significantly less treatment of mental health conditions than residents of cities, suburban areas, or rural areas adjacent to larger metropolitan areas [3]. Although primary care providers (PCPs) serve a crucial role in addressing the mental health needs of patients and providing access to care [6], referrals to specialists may be helpful in the treatment of many psychiatric conditions. Increasing rates of specialist referrals and a shortage of specialist availability represent a growing need for timely specialist advice and increased collaboration between PCPs and psychiatrists to reduce gaps in care.

The opportunity for collaboration between a PCP and psychiatrist can be achieved through eConsults [7]. eConsults are provider-to-provider communications within a shared electronic medical record or web-based platform. The use of eConsults has steadily expanded over the years, with eConsults serving as an opportunity to improve health care quality and reduce specialty care costs [7-10]. Although the use of eConsults has been established in several specialties—included but not limited to dermatology, cardiology, and nephrology—its use within the field of psychiatry has recently increased and may prove beneficial [11-13]. With extensive wait times for mental health prescribers in many areas, especially among rural populations, these consultations provide access to psychiatric specialists, may reduce delay in starting psychiatric treatment, and reach patients who may otherwise go without psychiatric care [7,14]. Although there are possible advantages for the use of psychiatric eConsults, little research has been conducted on the model of a psychiatric eConsult system in an academic rural health care setting.

The aim of this study is to explore the patterns, content, and provider implementation of recommendations of psychiatric eConsults in a large academic health care system. By conducting a descriptive analysis of 343 psychiatric eConsults, we outlined the practice of psychiatric eConsults, with the goal of portraying the acceptability and efficiency of this practice. We hope that our findings will add to the growing literature on e-consultation and show that a modest amount of specialist time can support PCPs in their delivery of mental health care for patients seeking help in primary care.

Methods

Implementation, Execution, and Structure of a Psychiatric Electronic Consultation Service Within a Large Academic System

Dartmouth-Hitchcock Health is a rural academic medical center with five primary care practice sites spread across New Hampshire. PCPs include attending physicians, resident physicians, physician assistants, and nurse practitioners. All clinics use Epic Systems software as a shared electronic health record (EHR). General electronic consultations were first launched at Dartmouth-Hitchcock Health in 2014 as part of the Association of American Medical Colleges (AAMC) Project CORE (Coordinating Optimal Referral Experiences: Implementing eConsults and Enhanced Referrals). The Department of Psychiatry joined the eConsult program at Dartmouth-Hitchcock Health in 2016. By 2018, Dartmouth-Hitchcock Health had 39 participating specialties. Psychiatric eConsult procedures and protocols were developed based on established guidelines set forth by the AAMC and used by all specialties within Dartmouth-Hitchcock Health. These included referral template guidelines, quality metrics for response content and time frames, and the expectations of referring providers and responding specialists. eConsult templates were created for questions regarding depression, anxiety, substance use, attention-deficit/hyperactivity disorder, bipolar disorder, and psychosis. Templates include a free-text field for the specific clinical question and short checklists for common comorbid conditions and prior medication trials. Patient-reported measures such as the Patient Health Questionnaire 9 and General Anxiety Disorder 7 are commonly available in the EHR, and relevant laboratory data are automatically pulled by the EHR into the referral template. eConsults are submitted to a shared pool within the EHR, which is monitored daily by two psychiatrists. The responding psychiatrist reviews the specific consultation question and information in the eConsult referral template and may also choose to review additional clinical information available in the EHR, such as the PCP's most recent progress note. An example eConsult is shown in Textbox 1, with a standard disclaimer included at the end of every eConsult in Textbox 2. Response time is expected within 3 business days (72 hours) and both the request and response become part of the patient's EHR. Typically, the specialist response to the PCP completes the eConsult exchange. Further clinical management is conducted by the PCP. However, in a small minority of cases, the consulting psychiatrist may recommend an additional referral for in-person consultation. Dartmouth-Hitchcock Health provides an internal accounting of a nominal work relative value unit credit for both the PCP and consulting psychiatrist, based on the time taken to complete the e-consultation. e-Consultations serve as no additional cost to the patient, and the insurance company of the patient is not billed.

Textbox 1. Example of a psychiatric eConsult.

1. eConsult question(s): Patient currently takes citalopram 40 mg daily. She has been on this dose since June of 2012. She shares that it may not be as effective for her mood as it used to be. She suffers from chronic neck pain with radiculopathy. She desires to switch to duloxetine, with the hope of gaining better pain control. Can you please advise on a cross taper from 40 mg citalopram to duloxetine?

2. Recommendation(s): Cross tapering an anti-depressant is often an art more than a science. I usually take into account a patient's sensitivity to medication side effects when doing a cross taper.

Assuming that this patient is not particularly sensitive to medication side effects, I would suggest the following:

Week 1: citalopram 20mg po daily and duloxetine 20mg po daily

Week 2: citalopram 10mg po daily and duloxetine 40mg po daily

Week 3: Discontinue citalopram and increase to duloxetine 60mg po daily

You can speed up or slow down this taper depending on how the patient tends to respond to these sorts of medication changes. I would slow down the taper if she is having any side effects. I would also educate about the risk for serotonin syndrome, as a cross taper like this puts the patient at a slightly increased risk for serotonin syndrome compared to tapering and discontinuing the citalopram first and then starting duloxetine.

Textbox 2. Standard disclaimer enclosed at the end of each eConsult.

This eConsult is focused on the specific clinical question(s) asked by the referring clinician, is based on the clinical data available to me, the consulting physician, at the time of the request, and is furnished without benefit of a comprehensive evaluation or physical examination of the patient by me. The guidance set forth in the eConsult note will need to be interpreted in light of any clinical issues not known to me or any changes in patient status that I may not be aware of at the time of filing this eConsult. If further consultation is necessary, an in-person visit with me or another member of our group is an option.

Data Collection Plan

We performed content analyses of all eConsults to psychiatry that occurred between May 2016 and February 2019 at Dartmouth-Hitchcock Medical Center (DHMC), located in Lebanon, New Hampshire. We included all eConsults for patients18 years or older, totaling 343 eConsults.

We reviewed and categorized eConsults according to the following: consult diagnosis, question type, recommendations, outcome classification, implementation status, time frame, and distance.

- "Consult diagnosis" was the listed billable diagnosis placed by the PCP at the time of the consult. This diagnosis did not always meet the Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition; DSM-5) criteria, given the limited information in the medical chart required to fully meet criteria for specific diagnoses.
- "Question type" assessed the specific type of question asked by the PCP. This aspect of the eConsult was further categorized to distinguish and group the various types of questions. All eConsult "question types" were independently assessed by two raters and categorized as confirming or making a diagnosis, medication recommendation, medication side effects or interactions, referral for psychotherapy, navigation of the health care system, need for psychiatric in-person evaluation, or other. Navigation of the health care system question types were categorized as all questions where a PCP requested assistance from the eConsult psychiatrist to determine the appropriate level of care or specific services that a patient might require based on their presentation. Medication recommendation was further categorized as pharmacological recommendation, assistance with a medication taper, or validation of current medication treatment plan.
- "Recommendations" outlined the specific recommendation type made by a psychiatrist to the PCP. These were categorized as one or more of the following: psychotherapy, medication choice, need for diagnostic clarity or additional testing, face-to-face consultation, further workup needed, no changes to plan, or other.
- "Implementation status" categorized whether or not the primary care clinician implemented the recommendations. All records were reviewed to evaluate the consulting psychiatrist recommendation and whether that specific recommendation was implemented by the PCP. If the PCP documented within the EHR that they followed the specific recommendation by the psychiatrist (eg, ordered recommended medication, referred to recommended provider, or gathered additional information for diagnostic clarification), they were documented to have "implemented recommendation." If the recommendations made by the psychiatrist were not implemented by the PCP, the reason for lack of implementation was explored, and they were documented to "have not implemented the recommendation." If there was no information in the EHR as to whether or not a recommendation was implemented either due to a loss of follow-up or lack of documentation in the chart by the PCP, the eConsult was excluded from this category.
- "Time frame" calculated the time it took from placement of the eConsult until completion of the eConsult by the consulting psychiatrist. We documented the number of exchanges between a consulting psychiatrist and primary care physician. A standard exchange was documented as zero if the only exchange was the initial consult question and a response back to the consult question.
- "Distance" was the distance in miles between the PCP office from whence the eConsult originated and DHMC, where the eConsult psychiatrist was located. The distance between

the patient's hometown and the location of the eConsult psychiatrist was also calculated.

Data Analysis

All information from charts were obtained via chart abstraction. We had three independent reviews for all consultations (two medical students and one attending psychiatrist). We used descriptive statistics to report patient demographics, consult diagnosis, question type, psychiatrist recommendation, recommendation implementation, time frame, and distance of primary care physician and patient from consulting provider.

Results

Patient Characteristics

Patient age ranged from 18 to 97 years. Most of the 343 patients were female (n=241, 70.3%), were non-Hispanic White (n=325, 94.8%), were employed full-time (n=157, 45.8%), were married (n=114, 42%), and had commercial insurance (n=212, 61.8%). All patients preferred English (Table 1).

Table 1. Patient characteristics of psychiatry eConsults from 2016 to 2019.

Characteristics	Patients (N=343)
Age (years), median (range)	47 (18-97)
Gender, n (%)	
Female	241 (70.3)
Male	100 (29.2)
Female (transgender)	1 (0.3)
Male (transgender)	1 (0.3)
Race, n (%)	
White (non-Hispanic)	325 (94.8)
White (Hispanic)	4 (1.2)
Asian	5 (1.5)
African American	4 (1.2)
Other	5 (1.5)
Employment status, n (%)	
Employed (full-time)	157 (45.8)
Employed (part-time)	15 (4.4)
Not employed	71 (20.7)
Retired	70 (20.4)
Disabled	12 (3.5)
Other (not listed)	18 (5.2)
Insurance, n (%)	
Commercial insurance	212 (61.8)
Dual eligibility (Medicare and Medicaid)	7 (2.0)
Medicaid	31 (9.0)
Medicare	78 (22.7)
No insurance	15 (4.4)
Marital status, n (%)	
Divorced	45 (13.1)
Married	144 (42.0)
Other	23 (6.7)
Separated	7 (2.0)
Single	124 (36.2)
Preferred language, n (%)	
English	340 (99.1)
Other	3 (0.9)

eConsult Content and Characteristics

Consult Diagnosis

eConsults often included multiple diagnoses. Therefore, among 343 eConsults, 450 diagnoses were abstracted. The most common diagnosis was depression (n=162, 36%), followed by anxiety (n=118, 26.2%), psychosis (n=47, 10.4%), other (n=47, 10.4%), bipolar disorder (n=28, 6.2%), attention-deficit/hyperactivity disorder (n=24, 5.3%), substance use (n=12, 2.7%), and posttraumatic stress disorder (n=12, 2.7%). “Other” diagnoses included eating disorders (n=11), personality disorders (n=5), obsessive-compulsive disorder (n=5), neurocognitive disease (n=8), insomnia (n=7), unspecified mood disorder (n=9), and adverse drug reactions (n=2). DSM-5–specific diagnoses were often not used, as there often was not sufficient information in the medical chart to make a DSM-5 diagnosis.

Question Type

A majority (n=288, 84%) of the 343 eConsults pertained specifically to medication management, which included medication choice, side effects or interactions, or medication taper. eConsults seeking help in securing a psychiatry referral represented 6.4% (n=22) of eConsults, and eConsults dealing with navigating the health care system, including questions on appropriate referrals or available services, represented 4.6% (n=16) of eConsults. Only 1.7% (n=6) of eConsults asked for assistance in making a clinical diagnosis, and 1.2% (n=4) were a request for referral for psychotherapy. Finally, 2% (n=7) of eConsult questions were classified as "other" and included management strategies (eg, how to approach a patient in a clinic who does not recognize their own delusions) and verification of a management plan.

eConsult Recommendations

Each eConsult response typically included multiple recommendations, with a total of 602 recommendations collected across 343 eConsults. Most eConsults recommended medication management, psychotherapy, or a combination of both (419/602, 69.6%). The consulting psychiatrist recommended to either reassess the diagnosis or obtain additional history 15% (90/602) of the time; face-to-face consultations or referrals to outpatient therapy were recommended 7% (43/602) of the time. No changes were recommended in 1.7% (10/602) of the eConsult responses. The category "other" included referral to specialized resources (eg, aging resource center), direct phone call to PCP, and other treatment modalities (eg, light box), and was recommended 1.7% (10/602) of times.

eConsult Implementation Status

A total of 43 out of 343 of eConsult cases could not be evaluated for implementation status due to a loss of follow-up or lack of documentation in the chart by the PCP. At least one of the psychiatrist's recommendations was implemented by the PCP in 282 out of 300 of cases. The recommendation was specifically not implemented in 6% (18/300) of the cases due to patient refusal (n=14), PCP refusal (n=2), and the patient improving without intervention (n=2). Therefore, of the 300 eConsults available for analysis, at least one recommendation was implemented 94% (282/300) of the time.

eConsult Time Frame

The average time for a psychiatrist to respond to an eConsult was 25.8 (IQR 5.4-28.5) hours. Nearly all eConsults (330/343, 96.2%) were sufficiently answered with one response.

Distance of Primary Care Physician and Patient From Consulting Provider

The average distance of the referring PCP office was 46 miles away from the psychiatrist who answered the eConsult. Of the 330 patients with listed PCPs, 41% (n=142) of eConsults were from PCPs with offices in the same hospital as the answering psychiatrist, and 35.8% (n=118) of eConsults were placed by PCPs in offices greater than 50 miles away. The average distance of the patient's hometown from the answering psychiatrist was also analyzed to measure the distance a patient would have to travel to attend an in-person evaluation by one of our psychiatrists. A total of 8 patients were excluded from this analysis due to lack of hometown data, resulting in an average distance of 43.7 miles from patient hometown to consulting psychiatrist. We found that 29% (97/335) of patients lived within 15 miles of the consulting psychiatrist, and 42% (142/335) lived more than 50 miles away.

Discussion

Our study demonstrates that psychiatric e-consultation can be a helpful tool to support PCPs in delivering mental health care to their patients. Psychiatric e-consultation serves as a bridge to improve timely access to specialist opinion, aiding both PCPs and patients. The asynchronous nature of eConsults permits the PCP to ask and the consulting psychiatrist to answer eConsults at a time convenient to both providers, rather than having to respond to a page or phone call from the PCP who is requesting the consultation. At our institution, the eConsult psychiatrists are outpatient psychiatrists who are not embedded in primary care practices and are often not available to discuss a case with a PCP by phone. Though eConsults do not replace the benefits of having a psychiatrist embedded in a primary care practice and available to do curbside consultations, they offer access to psychiatric care for primary care clinicians who do not have easy access to a psychiatrist.

e-Consultation may be of particular benefit in rural settings, where access to psychiatric specialists is often limited. Our analysis shows that eConsults can be answered promptly (on average <26 hours), and the vast majority of recommendations (94%) are implemented by primary care physicians. Considering in-person or telehealth psychiatry referrals for initial evaluations can take months to schedule, the timeliness of e-consultation can be helpful in reducing delays initiating treatment for common mental health disorders that frequently present in primary care. With additional constraints on in-person care caused by the COVID-19 pandemic, e-consultation provides yet another tool that may be used to meet the growing mental health needs in the United States.

Over 80% of eConsults in our study were questions about the management of psychiatric medications, which is similar to the findings of other studies evaluating the content of psychiatry eConsults [8]. Specifically, most PCPs in our study submitted an eConsult for medication recommendation or a second opinion on their choice of medication for an established diagnosis. Less commonly, questions were asked about side effect profile or advice on tapering a certain medication. Our finding that most eConsult questions were regarding pharmacotherapy may be explained by the therapeutic complexity of managing patients with psychiatric conditions.

Our response time was similar to those of other studies evaluating psychiatry eConsults, with an average response time of 1.1 days compared to 1.4 days in a study by Wren et al [9] and 2.6 days in a study by Lowenstein et al [10]. Furthermore, our response time to eConsults may have been slightly increased due to eConsults placed outside of normal work hours where response was delayed until the next workday. For reference, eConsults were sometimes returned as early as 12 minutes, and

31.2% (107/343) were returned within 12 hours. Overall, the quick turnaround time both in our study and in the literature confirms the viability of eConsults as a means to get more timely psychiatric recommendations than an in-person evaluation.

We believe that our study's 94% implementation rate of at least one eConsult recommendation and the fact that 96% of eConsults did not prompt further questions or follow-up from the PCP demonstrates that the quality of recommendations still remain high despite the lack of in-person evaluation, which is congruent to the 96% implementation rate reported by Lowenstein et al [10]. However, satisfaction of our eConsults would be better evaluated with a survey to PCPs, which was not performed. A follow-up study should be conducted to measure PCP satisfaction and patient outcomes following implementation of eConsult recommendations.

The utility of psychiatric eConsults particularly in rural settings is substantial. In our study, both patients and PCP physician offices were on average nearly an hour away from the center where the consulting psychiatry providers were located. Therefore, providing PCPs with the tool to receive timely specialist advice and potentially obviate the need for in-person evaluation has the potential to provide more immediate access to mental health care, reduce additional costs or need for separate outpatient appointments, and potentially reduce additional costs either for transit or time off from work.

Limitations to this study include the fact that this is a single-center study in a rural academic medical center, which is susceptible to a lack of external generalizability when compared to medical centers in other settings. eConsults were also performed by two psychiatrists, so certain elements of the consult such as the time to respond may differ if more consulting psychiatrists were included. Finally, this was a retrospective study evaluating only patients 18 years or older and from a 3-year period.

Psychiatric e-consultations are effective in delivering specialist recommendations to PCPs and patients in a timely manner. eConsults may be of particular benefit in rural settings, bridging both time and distance, which are common barriers that delay care. eConsults allow PCPs to quickly verify or make changes to their management of psychiatric disorders and may avoid the need for in-person appointments with a psychiatric specialist in a majority of cases. Demonstrating the value and effectiveness of psychiatric e-consultation is especially important given the barriers patients and PCPs may experience when trying to refer for in-person psychiatry evaluation.

Acknowledgments

We would like to acknowledge Megan Colgan for obtaining the list of eConsults at our center from 2016 to 2019. There was no funding for this study.

Authors' Contributions

JA and DD contributed to data collection and data analysis, and contributed drafts and revisions to the manuscript. MD and GS contributed to study design and data analysis, and contributed to revisions of the manuscript.

Conflicts of Interest

None declared.

References

1. Steel Z, Marnane C, Iranpour C, Chey T, Jackson JW, Patel V, et al. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. Int J Epidemiol 2014 Apr;43(2):476-493 [FREE Full text] [doi: 10.1093/ije/dyu038] [Medline: 24648481]
2. Fagiolini A, Goracci A. The effects of undertreated chronic medical illnesses in patients with severe mental disorders. J Clin Psychiatry 2009;70 Suppl 3:22-29. [doi: 10.4088/JCP.7075su1c.04] [Medline: 19570498]
3. Thomas KC, Ellis AR, Konrad TR, Holzer CE, Morrissey JP. County-level estimates of mental health professional shortage in the United States. Psychiatr Serv 2009 Oct;60(10):1323-1328. [doi: 10.1176/ps.2009.60.10.1323] [Medline: 19797371]
4. Gleason N, Prasad PA, Ackerman S, Ho C, Monacelli J, Wang M, et al. Adoption and impact of an eConsult system in a fee-for-service setting. Healthc (Amst) 2017 Mar;5(1-2):40-45. [doi: 10.1016/j.hjdsi.2016.05.005] [Medline: 27469441]
5. Ngui EM, Khasakhala L, Ndetei D, Roberts LW. Mental disorders, health inequalities and ethics: a global perspective. Int Rev Psychiatry 2010;22(3):235-244 [FREE Full text] [doi: 10.3109/09540261.2010.485273] [Medline: 20528652]
6. Cook NL, Hicks LS, O'Malley AJ, Keegan T, Guadagnoli E, Landon BE. Access to specialty care and medical services in community health centers. Health Aff (Millwood) 2007;26(5):1459-1468. [doi: 10.1377/hlthaff.26.5.1459] [Medline: 17848458]
7. Liddy C, Afkham A, Drosinis P, Joschko J, Keely E. Impact of and satisfaction with a new eConsult service: a mixed methods study of primary care providers. J Am Board Fam Med 2015;28(3):394-403 [FREE Full text] [doi: 10.3122/jabfm.2015.03.140255] [Medline: 25957372]

8. Archibald D, Stratton J, Liddy C, Grant RE, Green D, Keely EJ. Evaluation of an electronic consultation service in psychiatry for primary care providers. BMC Psychiatry 2018 May 02;18(1):119 [FREE Full text] [doi: 10.1186/s12888-018-1701-3] [Medline: 29720133]
9. Wrenn K, Catschegn S, Cruz M, Gleason N, Gonzales R. Analysis of an electronic consultation program at an academic medical centre: primary care provider questions, specialist responses, and primary care provider actions. J Telemed Telecare 2017 Feb;23(2):217-224. [doi: 10.1177/1357633X16633553] [Medline: 26940797]
10. Lowenstein M, Bamgbose O, Gleason N, Feldman MD. Psychiatric consultation at your fingertips: descriptive analysis of electronic consultation from primary care to psychiatry. J Med Internet Res 2017 Aug 04;19(8):e279 [FREE Full text] [doi: 10.2196/jmir.7921] [Medline: 28778852]
11. Oseran AS, Wasfy JH. Early experiences with cardiology electronic consults: a systematic review. Am Heart J 2019 Sep;215:139-146. [doi: 10.1016/j.ahj.2019.06.013] [Medline: 31325772]
12. Wang RF, Trinidad J, Lawrence J, Pootrakul L, Forrest LA, Goist K, et al. Improved patient access and outcomes with the integration of an eConsult program (teledermatology) within a large academic medical center. J Am Acad Dermatol 2020 Dec;83(6):1633-1638. [doi: 10.1016/j.jaad.2019.10.053] [Medline: 31678336]
13. Keely E, Li J, Magner P, Afkham A, Liddy C. Nephrology eConsults for primary care providers: original investigation. Can J Kidney Health Dis 2018;5:2054358117753619 [FREE Full text] [doi: 10.1177/2054358117753619] [Medline: 29399366]
14. Liddy C, Moroz I, Afkham A, Keely E. Sustainability of a primary care-driven eConsult service. Ann Fam Med 2018 Mar;16(2):120-126 [FREE Full text] [doi: 10.1370/afm.2177] [Medline: 29531102]

Abbreviations

AAMC: Association of American Medical Colleges
CORE: Coordinating Optimal Referral Experiences: Implementing eConsults and Enhanced Referrals
DHMC: Dartmouth-Hitchcock Medical Center
DSM-5: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
EHR: electronic health record
PCP: primary care provider

Edited by R Kukafka; submitted 16.11.20; peer-reviewed by E van der Velde, J Pecina; comments to author 12.01.21; revised version received 21.04.21; accepted 19.07.21; published 01.09.21.

Please cite as:
Avery J, Dwan D, Sowden G, Duncan M
Primary Care Psychiatry eConsults at a Rural Academic Medical Center: Descriptive Analysis
J Med Internet Res 2021;23(9):e24650
URL: https://www.jmir.org/2021/9/e24650
doi:10.2196/24650
PMID:34468329

Original Paper

Disparities in Outpatient and Telehealth Visits During the COVID-19 Pandemic in a Large Integrated Health Care Organization: Retrospective Cohort Study

Lei Qian[1], PhD; Lina S Sy[1], MPH; Vennis Hong[1], MPH; Sungching C Glenn[1], MS; Denison S Ryan[1], MPH; Kerresa Morrissette[1], MPH; Steven J Jacobsen[1], MD, PhD; Stanley Xu[1], PhD

Department of Research & Evaluation, Kaiser Permanente Southern California, Pasadena, CA, United States

Corresponding Author:
Lei Qian, PhD
Department of Research & Evaluation
Kaiser Permanente Southern California
100 S. Los Robles Ave, 5th Floor
Pasadena, CA, 91101
United States
Phone: 1 6266586128
Email: lei.x.qian@kp.org

Abstract

Background: Dramatic decreases in outpatient visits and sudden increases in telehealth visits were observed during the COVID-19 pandemic, but it was unclear whether these changes differed by patient demographics and socioeconomic status.

Objective: This study aimed to assess the impact of the pandemic on in-person outpatient and telehealth visits (telephone and video) by demographic characteristics and household income in a diverse population.

Methods: We calculated weekly rates of outpatient and telehealth visits by age, sex, race/ethnicity, and neighborhood-level median household income among members of Kaiser Permanente Southern California (KPSC) from January 5, 2020, to October 31, 2020, and the corresponding period in 2019. We estimated the percentage change in visit rates during the early pandemic period (March 22 to April 25, 2020) and the late pandemic period (October 4 to October 31, 2020) from the prepandemic period (January 5 to March 7, 2020) in Poisson regression models for each subgroup while adjusting for seasonality using 2019 data. We examined if the changes in visit rates differed by subgroups statistically by comparing their 95% CIs.

Results: Among 4.56 million KPSC members enrolled in January 2020, 15.0% (n=682,947) were ≥65 years old, 51.5% (n=2,345,020) were female, 39.4% (n=1,795,994) were Hispanic, and 7.7% (n=350,721) lived in an area of median household income <US $40,000. Increases in telehealth visits during the pandemic varied across subgroups, while decreases in outpatient visits were similar, except by age. Among age groups, the ≥65 years population had the least increase in telehealth visits (236.6%, 95% CI 228.8%-244.5%), with 4.9 per one person-year during the early pandemic period versus 1.5 per one person-year during the prepandemic period. During the same periods, across racial/ethnic groups, Hispanic individuals had the largest increase in telehealth visits (295.5%, 95% CI 275.5%-316.5%). Across income levels, telehealth visits in the low-income group increased the most (313.5%, 95% CI 294.8%-333.1%). The rate of combined outpatient and telehealth visits in the Hispanic, non-Hispanic Black, and low-income groups returned to prepandemic levels by October 2020.

Conclusions: The Hispanic group and low-income group had the largest percentage increase in telehealth utilization in response to the COVID-19 pandemic. The use of virtual care potentially mitigated the impact of the pandemic on health care utilization in these vulnerable populations.

(J Med Internet Res 2021;23(9):e29959) doi:10.2196/29959

KEYWORDS

COVID-19; pandemic; health care utilization; telehealth; disparity

Introduction

The COVID-19 pandemic has impacted every aspect of life since it started in late December 2019, including health care delivery in the United States. After the declaration of a national state of emergency concerning the pandemic on March 13, 2020, the Centers for Disease Control and Prevention (CDC) and Centers for Medicare and Medicaid Services recommended delaying elective and nonemergent care on March 18, 2020 [1]. In addition, the CDC encouraged the use of telehealth services to deliver care [2]. Due to these policy changes and patients' concerns about COVID-19 infection during health care visits, dramatic decreases in outpatient visits and sudden increases in telehealth visits (telephone and video) were observed during the COVID-19 pandemic [3-5].

However, the changes in outpatient and telehealth visits caused by the pandemic may differ by demographic and socioeconomic characteristics such as age, sex, race/ethnicity, language preference, income, and education [6,7]. Although telehealth visits increased prior to the pandemic primarily due to increased internet availability, insurance coverage, and cost savings [8], access to telehealth has inherent barriers [9,10]. Individuals with limited digital literacy or access to digital technology, limited health literacy, or limited English proficiency may not have the means to pursue telehealth technologies [11]. During the COVID-19 pandemic, these barriers may influence the existing disparities in accessing telehealth visits among individuals of low socioeconomic status, individuals of racial and ethnic minorities, and women [12].

Knowledge about utilization disparities during the pandemic is limited, and there remains great interest in understanding utilization disparities in large and diverse health care organizations. The objective of this study was to assess the impact of the COVID-19 pandemic on in-person outpatient and telehealth visits (telephone and video) in subgroups defined by age, sex, race/ethnicity, and neighborhood-level median household income in a large diverse population through October 31, 2020.

Methods

Study Setting, Population, and Period

This retrospective cohort study was conducted at Kaiser Permanente Southern California (KPSC), an integrated health care system that provides comprehensive health care to over 4.7 million racially, ethnically, and socioeconomically diverse members at its 15 hospitals and 234 medical offices [13]. The prepaid health plan provides strong motivation for members to use services at KPSC facilities. KPSC's electronic health record system stores information about all aspects of the care provided to members, including demographics, medical encounters, vaccinations, pharmacy utilization, membership history, and claims. The study population consisted of individuals of all ages who were KPSC members from January 5, 2020, to October 31, 2020 (week 1 to week 43 in 2020) and the corresponding period in 2019.

Outpatient and Telehealth Visits

We identified outpatient visits and telehealth visits using electronic health record data and claims data. The outpatient visits included all in-person outpatient visits, while telehealth visits included telephone appointment visits and video visits that were conducted synchronously using real-time telephone or live video-audio interaction. These telehealth visits were billable and had a diagnosis or procedure code. Less than 6% of outpatient and telehealth encounters in KPSC were from claims [14]. The study was reviewed and approved by the KPSC Institutional Review Board.

Statistical Analysis

We first described the characteristics of KPSC members enrolled in January 2020. The weekly (Sunday to Saturday) rates of outpatient visits, telehealth visits, and combined outpatient and telehealth visits were calculated during the study period by age (0-5 years, 6-17 years, 18-44 years, 45-64 years, and ≥65 years), sex, self-reported race/ethnicity (non-Hispanic White, non-Hispanic Black, Hispanic, non-Hispanic Asian, and other/unknown), and neighborhood-level median household income (<US $40,000, US $40,000-$79,999, and ≥US $80,000). In calculating a weekly rate, the numerator was the visit count of each visit type and the denominator was person-years of membership during a given week. The trend of visit rates during the pandemic year by subgroups was plotted separately for outpatient, telehealth, and combined outpatient and telehealth visits.

To assess the impact of the COVID-19 pandemic on the visit rate, we selected the following three periods in 2020: week 1-9 (January 5 to March 7, 2020) for the prepandemic period, week 12-16 (March 22 to April 25, 2020) for the early pandemic period immediately after California's stay-at-home order, and week 40-43 (October 4 to October 31, 2020) for the later pandemic period after reopening. Three corresponding periods in 2019 were also identified and used to control for seasonality. We estimated the ratio of the visit rate during the pandemic period to the visit rate during the prepandemic period in Poisson regression models with the visit count as the dependent variable and natural logarithm of person-years of membership as an offset for each subgroup, adjusting for other subgroup variables and seasonality. The robust variance estimator was used to obtain standard errors of parameter estimates in Poisson regression models. After the adjusted rate ratio was estimated from Poisson regression models, the percentage change was calculated as follows: (adjusted rate ratio − 1) × 100%. We examined if the changes in visit rates differed by age, sex, race/ethnicity, and household income by comparing their 95% CIs.

Results

Study Population

Among the 4.56 million KPSC members enrolled in January 2020, 15.0% (n=682,947) were ≥65 years old, 51.5% (n=2,345,020) were female, 39.4% (n=1,795,994) were Hispanic, and 7.7% (n=350,721) lived in census tract areas with median household income <US $40,000 (Table 1).

Table 1. Demographic characteristics and neighborhood household income levels of members in Kaiser Permanente Southern California in January 2020 (N=4,556,646).

Characteristic	Value, n (%)
Age (years)	
0-5	284,413 (6.2)
6-17	653,649 (14.3)
18-44	1,743,358 (38.3)
45-64	1,192,279 (26.2)
≥65	682,947 (15.0)
Sex	
Female	2,345,020 (51.5)
Male	2,211,304 (48.5)
Unknown	322 (0.0)
Race/ethnicity	
Hispanic	1,795,994 (39.4)
Non-Hispanic White	1,403,840 (30.8)
Non-Hispanic Black	347,019 (7.6)
Non-Hispanic Asian	448,270 (9.8)
Other and unknown	561,523 (12.3)
Neighborhood median household income (US $)	
<40,000	350,721 (7.7)
40,000-79,999	2,178,471 (47.8)
≥80,000	1,913,377 (42.0)
Unknown	114,077 (2.5)

Outpatient Visits

Across all subgroups, outpatient visit rates had a steep drop-off after the start of the pandemic, reaching their lowest levels by early April 2020 (week 14), and then, they gradually increased but remained lower than prepandemic levels (Figure 1). Among age groups, the ≥65 years population had the smallest decrease in outpatient visits (adjusted percentage change: −29.6%, 95% CI −30.5% to −28.6%), with 6.25 visits per person-year during the later pandemic period versus 8.92 visits per person-year during the prepandemic period in 2020 (Table 2). Children aged 6-17 years had the largest decrease (−50.2%, 95% CI −51.6% to −48.7%) in outpatient visits, with 1.92 visits per person-year during the later pandemic period versus 3.86 visits per person-year during the prepandemic period in 2020. The decreases in outpatient visits did not differ statistically by sex, race/ethnicity, and household income groups, ranging from −83.8% to −80.7% when comparing the early pandemic period to the prepandemic period in 2020 and ranging from −38.8% to −35.9% when comparing the later pandemic period to the prepandemic period in 2020.

Figure 1. Outpatient visit rates over time by age, sex, race/ethnicity, and neighborhood median household income in 2020.

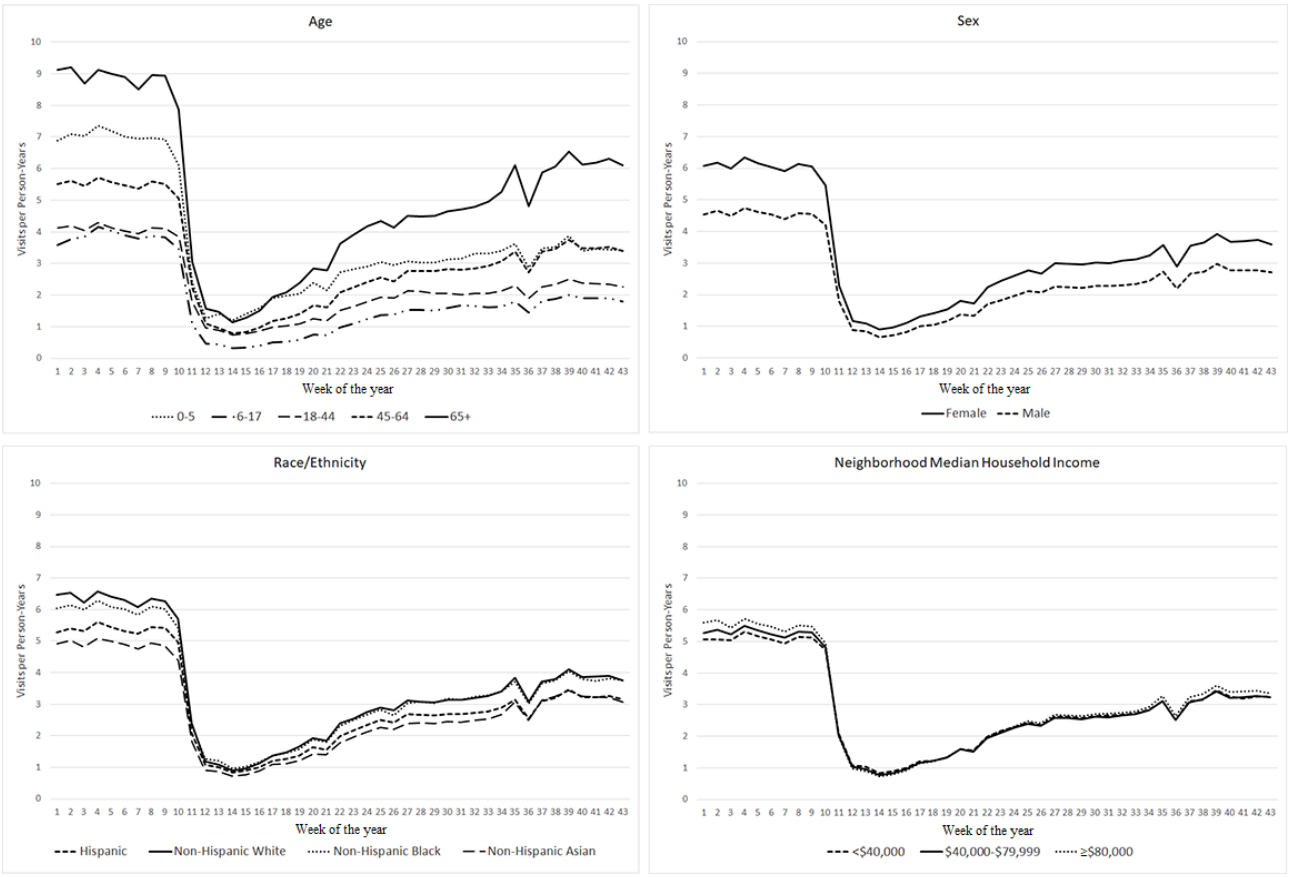


Table 2. Outpatient visit rates per one person-year before and during the COVID-19 pandemic in 2020.

Characteristic	Prepandemic	During the pandemic		Percentage change from prepandemic (%)		Adjusted percentage change from prepandemic (%) (95% CI)[a]	
	Week 1-9	Week 12-16 (early)	Week 40-43 (later)	Week 12-16 (early)	Week 40-43 (later)	Week 12-16 (early)	Week 40-43 (later)
Age (years)							
0-5	7.03	1.35	3.53	−80.8	−49.7	−79.6 (−80.7 to −78.5)	−47.2 (−48.9 to −45.4)
6-17	3.86	0.38	1.92	−90.2	−50.3	−90.0 (−90.4 to −89.5)	−50.2 (−51.6 to −48.7)
18-44	4.10	0.83	2.39	−79.6	−41.7	−78.9 (−79.5 to −78.3)	−40.8 (−42.0 to −39.4)
45-64	5.52	0.93	3.52	−83.2	−36.2	−82.6 (−83.0 to −82.2)	−35.4 (−36.5 to −34.2)
≥65	8.92	1.39	6.25	−84.4	−29.9	−84.1 (−84.5 to −83.8)	−29.6 (−30.5 to −28.6)
Sex							
Female	6.08	1.03	3.73	−83.0	−38.7	−82.5 (−82.9 to −82.1)	−38.3 (−39.2 to −37.4)
Male	4.55	0.78	2.80	−83.0	−38.4	−82.3 (−82.7 to −81.9)	−37.4 (−38.5 to −36.2)
Race/ethnicity							
Hispanic	5.37	0.95	3.26	−82.2	−39.3	−81.7 (−82.8 to −80.5)	−38.8 (−41.6 to −35.8)
Non-Hispanic White	6.34	1.03	3.90	−83.7	−38.4	−83.1 (−83.9 to −82.2)	−38.0 (−40.3 to −35.7)
Non-Hispanic Black	6.04	1.11	3.82	−81.6	−36.8	−81.1 (−82.2 to −80.0)	−36.3 (−39.0 to −33.4)
Non-Hispanic Asian	4.90	0.82	3.23	−83.2	−34.0	−83.0 (−84.0 to −81.9)	−35.9 (−38.5 to −33.1)
Neighborhood median household income (US $)							
<40,000	5.09	0.97	3.23	−81.0	−36.5	−80.7 (−81.7 to −79.5)	−36.8 (−39.6 to −33.8)
40,000-79,999	5.28	0.92	3.23	−82.7	−38.9	−82.2 (−83.2 to −81.2)	−38.7 (−41.2 to −36.1)
≥80,000	5.51	0.87	3.40	−84.3	−38.3	−83.8 (−84.6 to −82.9)	−38.1 (−40.3 to −35.7)

[a]Analyses for age groups were adjusted for sex and race/ethnicity; analyses for sex were adjusted for age and race/ethnicity; analyses for race/ethnicity were adjusted for age and sex; analyses for neighborhood median household income were adjusted for age, sex, and race/ethnicity; all analyses were adjusted for seasonality using 2019 data.

Telehealth Visits

Across all subgroups, except children aged 6-17 years, telehealth visit rates increased sharply after the start of the pandemic and then slowly decreased (Figure 2). The telehealth visit rates in children aged 6-17 years remained stable after an initial increase during the pandemic. Comparing the early pandemic period to the prepandemic period in 2020, across age groups, the telehealth visits increased the most among children aged 0-5 years (475.5%, 95% CI 446.2%-506.4%) and the least among adults aged ≥65 years (236.6%, 95% CI 228.8%-244.5%) (Table 3). Telehealth visits increased significantly more among males (293.9%, 95% CI 283.4%-304.7%) than females (260.0%, 95% CI 252.6%-267.6%). Across racial/ethnic groups, Hispanic individuals had the largest increase in telehealth visits (295.5%, 95% CI 275.5%-316.5%). Across income levels, telehealth visits in the low-income group increased the most (313.5%, 95% CI 294.8%-333.1%). The patterns were similar when comparing the later pandemic period to the prepandemic period in 2020, but with smaller differences.

Figure 2. Telehealth visit rates over time by age, sex, race/ethnicity, and neighborhood median household income in 2020.

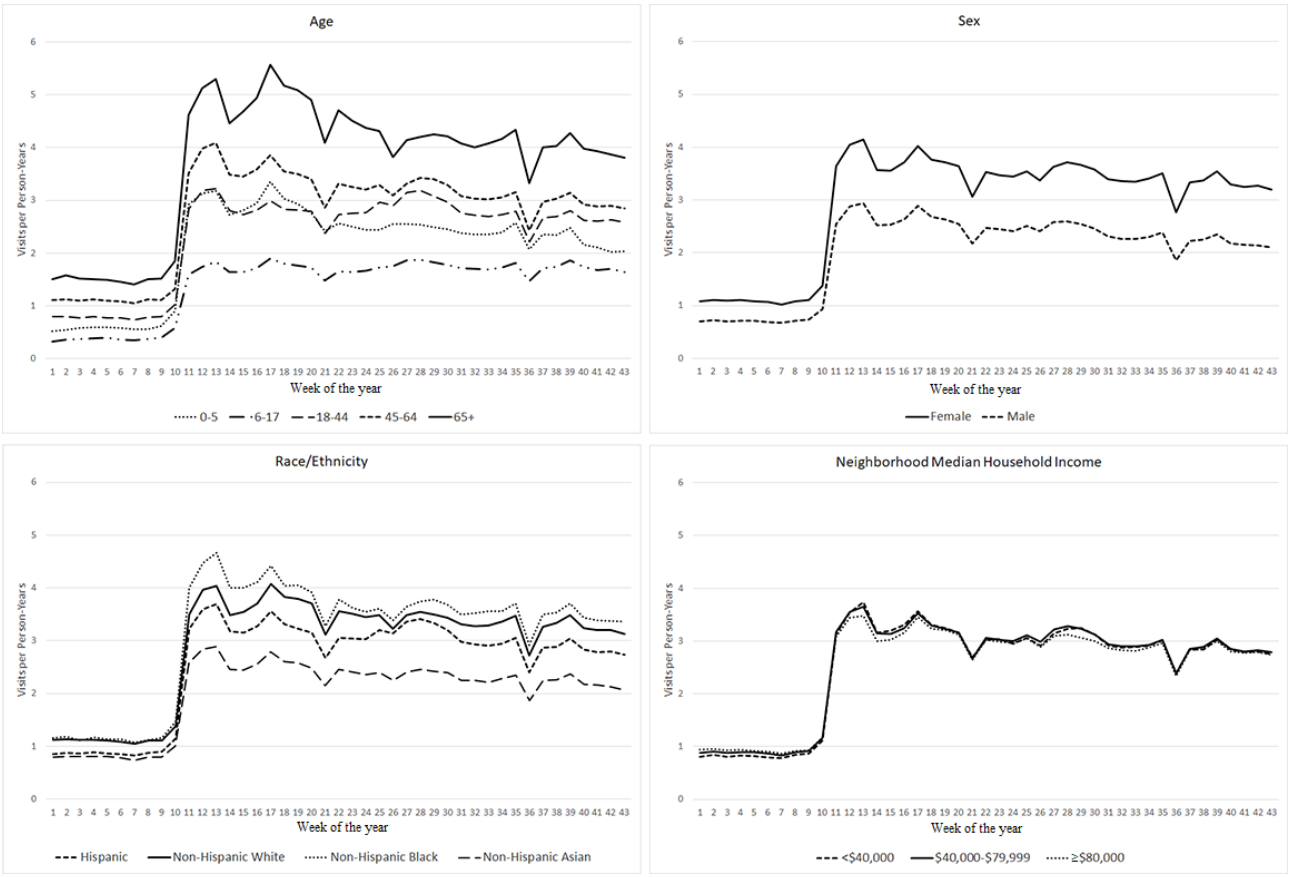


Table 3. Telehealth visit rates per one person-year before and during the COVID-19 pandemic in 2020.

Characteristic	Prepandemic	During the pandemic		Percentage change from prepandemic (%)		Adjusted percentage change from prepandemic (%) (95% CI)[a]	
	Week 1-9	Week 12-16 (early)	Week 40-43 (later)	Week 12-16 (early)	Week 40-43 (later)	Week 12-16 (early)	Week 40-43 (later)
Age (years)							
0-5	0.57	2.97	2.24	422.2	293.4	475.5 (446.2-506.4)	354.2 (331.4-378.2)
6-17	0.37	1.72	1.77	370.4	383.6	383.5 (362.9-405.1)	394.8 (372.9-417.7)
18-44	0.78	2.95	2.65	278.5	239.6	282.6 (271.2-294.3)	237.6 (228.3-247.2)
45-64	1.10	3.72	2.91	238.5	165.2	247.7 (237.8-257.8)	164.0 (157.0-171.2)
≥65	1.49	4.90	3.91	227.7	161.5	236.6 (228.8-244.5)	161.1 (156.2-166.2)
Sex							
Female	1.08	3.81	3.30	252.2	204.9	260.0 (252.6-267.6)	204.0 (197.1-211.0)
Male	0.70	2.70	2.19	283.7	211.0	293.9 (283.4-304.7)	212.4 (203.1-222.0)
Race/ethnicity							
Hispanic	0.86	3.37	2.83	291.3	227.6	295.5 (275.5-316.5)	220.3 (203.0-238.5)
Non-Hispanic White	1.10	3.75	3.23	240.4	193.3	255.2 (240.7-270.3)	198.9 (186.0-212.4)
Non-Hispanic Black	1.13	4.25	3.43	274.1	202.5	278.6 (260.1-298.0)	188.3 (173.7-203.6)
Non-Hispanic Asian	0.79	2.63	2.18	233.7	176.1	233.2 (215.7-251.7)	169.0 (153.7-185.3)
Neighborhood median household income (US $)							
<40,000	0.82	3.38	2.79	311.7	239.5	313.5 (294.8-333.1)	220.4 (204.7-236.9)
40,000-79,999	0.88	3.34	2.81	277.6	217.8	283.3 (267.5-299.7)	211.4 (197.6-225.8)
≥80,000	0.92	3.22	2.77	249.0	200.7	259.3 (245.4-273.8)	202.4 (189.9-215.5)

[a]Analyses for age groups were adjusted for sex and race/ethnicity; analyses for sex were adjusted for age and race/ethnicity; analyses for race/ethnicity were adjusted for age and sex; analyses for neighborhood median household income were adjusted for age, sex, and race/ethnicity; all analyses were adjusted for seasonality using 2019 data.

Combined Outpatient and Telehealth Visits

Despite increases in telehealth visits during the pandemic, across all subgroups, the rate of combined telehealth and outpatient visits still decreased during the early pandemic period and only returned to prepandemic levels for some subgroups (Figure 3). Comparing the later pandemic period to the prepandemic period in 2020, across age groups, the rate of combined outpatient and telehealth visits decreased the most among children aged 0-5 years (−19.7%, 95% CI −21.9% to −17.5%) (Table 4). The combined visit rate among adults aged 18-44 years during the later pandemic period exceeded the prepandemic level by 4.7% (95% CI 2.5%-6.8%). The combined visit rates returned to the prepandemic levels by October 2020 for the following subgroups: Hispanics (percentage change: −1.8%, 95% CI −6.1% to 2.6%), Non-Hispanic Black individuals (percentage change: 1.1%, 95% CI −3.0% to 5.3%), those with income <US $40,000 (percentage change: 0.7%, 95% CI −3.4% to 5.0%), and those with income US $40,000-$79,999 (percentage change: −2.1%, 95% CI −5.8% to 1.8%).

Figure 3. Combined outpatient and telehealth visit rates over time by age, sex, race/ethnicity, and neighborhood median household income in 2020.

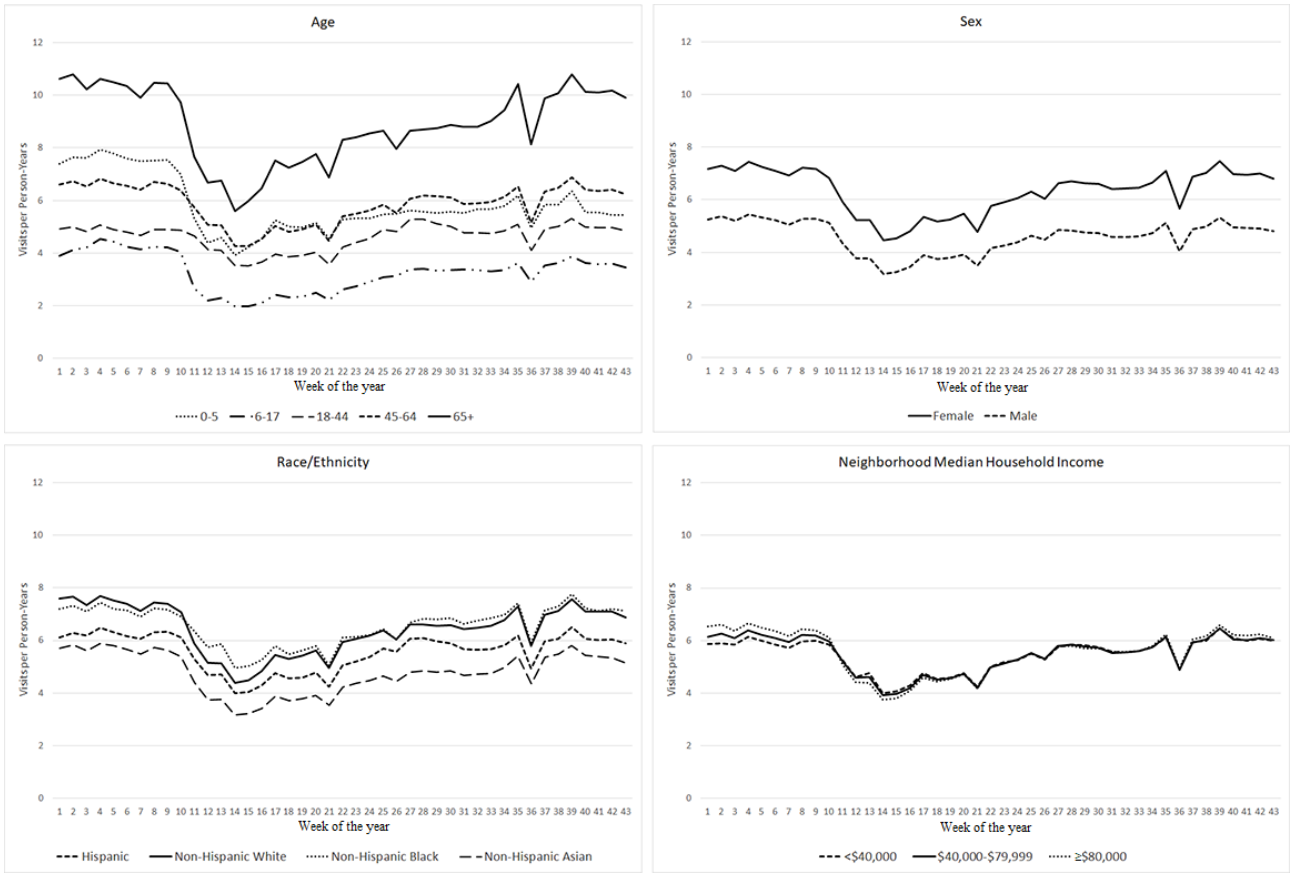


Table 4. Combined outpatient and telehealth visit rates per one person-year before and during the COVID-19 pandemic in 2020.

Characteristic	Prepandemic	During the pandemic		Percentage change from prepandemic (%)		Adjusted percentage change from prepandemic (%) (95% CI)[a]	
	Week 1-9	Week 12-16 (early)	Week 40-43 (later)	Week 12-16 (early)	Week 40-43 (later)	Week 12-16 (early)	Week 40-43 (later)
Age (years)							
0-5	7.60	4.32	5.77	−43.1	−24.0	−39.5 (−41.5 to −37.4)	−19.7 (−21.9 to −17.5)
6-17	4.22	2.10	3.69	−50.2	−12.7	−49.2 (−50.6 to −47.7)	−12.3 (−14.6 to −9.9)
18-44	4.88	3.79	5.04	−22.3	3.3	−19.8 (−21.4 to −18.2)	4.7 (2.5 to 6.8)
45-64	6.62	4.64	6.43	−29.9	−2.8	−27.4 (−28.7 to −26.0)	−1.8 (−3.4 to −0.1)
≥65	10.41	6.28	10.16	−39.6	−2.5	−38.5 (−39.4 to −37.5)	−2.1 (−3.3 to −0.9)
Sex							
Female	7.16	4.84	7.02	−32.4	−2.0	−30.6 (−31.7 to −29.5)	−1.4 (−2.7 to −0.1)
Male	5.26	3.48	4.99	−33.9	−5.0	−31.4 (−32.6 to −30.2)	−3.5 (−5.1 to −2.0)
Race/ethnicity							
Hispanic	6.23	4.33	6.08	−30.5	−2.4	−28.7 (−31.8 to −25.5)	−1.8 (−6.1 to 2.6)
Non-Hispanic White	7.44	4.78	7.13	−35.8	−4.1	−33.3 (−35.6 to −30.9)	−3.4 (−6.6 to 0.0)
Non-Hispanic Black	7.18	5.36	7.25	−25.3	1.0	−23.5 (−26.7 to −20.3)	1.1 (−3.0 to 5.3)
Non-Hispanic Asian	5.69	3.45	5.41	−39.3	−4.8	−38.4 (−41.0 to −35.7)	−7.5 (−11.1 to −3.7)
Neighborhood median household income (US $)							
<40,000	5.91	4.35	6.02	−26.5	1.9	−25.2 (−28.3 to −22.1)	0.7 (−3.4 to 5.0)
40,000-79,999	6.17	4.26	6.04	−31.0	−2.1	−29.2 (−31.9 to −26.4)	−2.1 (−5.8 to 1.8)
≥80,000	6.43	4.08	6.17	−36.5	−4.1	−34.4 (−36.7 to −32.1)	−3.7 (−6.9 to −0.3)

[a]Analyses for age groups were adjusted for sex and race/ethnicity; analyses for sex were adjusted for age and race/ethnicity; analyses for race/ethnicity were adjusted for age and sex; analyses for neighborhood median household income were adjusted for age, sex, and race/ethnicity; all analyses were adjusted for seasonality using 2019 data.

Discussion

Principal Findings

We examined the impact of the COVID-19 pandemic on outpatient and telehealth visits by demographic characteristics and household income levels in a large integrated health care system. We found that increases in telehealth visits during the COVID-19 pandemic varied across age, sex, race/ethnicity, and household income groups, while decreases in outpatient visits during the COVID-19 pandemic were similar across all subgroups, except age groups. The rates of combined outpatient and telehealth visits in the Hispanic, non-Hispanic Black, and low-income groups returned to prepandemic levels by October 2020.

While adults aged ≥65 years had the highest telehealth visit rates before and during the pandemic, this age group had the smallest percentage increase in telehealth visits during the pandemic among all age groups, suggesting that greater barriers existed for the senior population to further expand their use of telehealth services during the pandemic. For example, seniors may have limited access to digital technologies and may have difficulty in using technologies required for telehealth visits. Recent studies among Medicare beneficiaries showed that 26.3% of individuals lacked digital access at home, making it unlikely for them to have video visits with clinicians [15]. Furthermore, a recent study showed that difficulty hearing and problems of speaking or making oneself understood were the top two reasons for hesitancy to use telehealth [16].

Members living in an area with median household income <US $40,000 increased their telehealth visits per person-year from 0.82 during week 1-9 to 3.38 during week 12-16, a higher percentage increase than that for those living in an area with median household income ≥US $80,000 (from 0.92 during week 1-9 to 3.22 during week 12-16). The KPSC service area covers mostly nonrural residents whose access to a telephone is not a barrier for using telehealth services. Following the CDC's recommendation to delay elective and nonemergent care on March 18, 2020, KPSC offered zero co-payments for telehealth visits, which might have been especially helpful for the low-income population. In addition, telehealth visits provide flexibility in care location, and save time and cost associated with commuting for in-person care.

Hispanic individuals had the greatest percentage increase in telehealth visits across racial/ethnic groups, although their visit rates during the prepandemic and pandemic periods were lower than those of non-Hispanic Black and non-Hispanic White individuals. A study conducted in New York City during the pandemic showed that Hispanic and Black patients had lower odds of using telehealth visits versus office visits than either White or Asian patients, but the authors did not consider the trend of telehealth utilization [17]. A recent study based on survey data showed that Black respondents were more likely than White respondents to report using telehealth because of the pandemic [18]. The fact that a larger proportion of racial/ethnic minorities live in low-income areas, along with KPSC's zero co-payment policy for telehealth visits, might support telehealth care among Hispanic individuals. In addition, telehealth is accessible for members with limited English proficiency. Spanish is a common language in southern California, spoken by many providers. The removal of Spanish language barriers could have encouraged Hispanic individuals to continue using telehealth care and contributed to larger increases of telehealth care in this population.

The impact of the COVID-19 pandemic on telephone and video visits may differ because the barrier to technology use is higher for video than telephone [12]. We assessed the rate of telehealth visits within KPSC separately by telephone and video visits (Multimedia Appendix 1). The trend of telephone visits was similar to the trend of overall telehealth visits, with the rates increasing sharply after the start of the pandemic and then gradually decreasing. The use of video visits was very low before the pandemic, with only 2.1% of telehealth visits occurring within KPSC being video visits. The video visit rate increased gradually over time after the start of the pandemic, contributing to 24% of telehealth visits by the end of October 2020. The slower and more gradual increase in video visits was likely due to technology barriers, for example, need for a web camera and digital literacy to conduct a video visit. These barriers were more prominent in those aged ≥65 years (Multimedia Appendix 2), with video visits having a slower start compared to that in some other age groups. In addition, it took time for both members and providers to adopt this new approach, with gradual increases in video visits occurring across all subgroups. Given the growing familiarity with new telehealth technologies, even as the COVID-19 pandemic wanes, patients likely will continue to pursue telehealth services beyond pre-COVID-19 levels.

Limitations

There are some potential limitations in this study. First, while we studied the impact of the pandemic on health care utilization among subpopulations, we did not address the quality of care and population health. Second, we assessed members' neighborhood-level income based on their address, which is less accurate than individual-level income. Third, because KPSC's health plan, hospitals, and medical groups are integrated to create a system for coordinated and comprehensive patient care, our patients may have better access to care than the general population. Therefore, these results may not be generalizable to populations in less integrated health care settings. Finally, we focused on relative change (ie, percentage change from baseline) in utilization as part of this study, but we acknowledge that absolute change (ie, difference from baseline) in utilization may be of interest as well. As such, we have presented absolute visit rates in each time period by subgroup to allow for such comparisons to be made.

Conclusion

While the ≥65 years population had the least percentage increase in telehealth visits, the Hispanic and low-income groups had the greatest percentage increase in telehealth utilization, possibly due to the low cost and flexibility of location and time. The use of virtual care in these subgroups reduced exposure to COVID-19, reinforced social distancing protocols, increased the number of patients who could be treated, and potentially mitigated the impact of COVID-19 in vulnerable populations. Furthermore, the use of virtual care might be a promising way to reduce health care disparities even after the COVID-19 pandemic ends.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Telephone and video visit rates within Kaiser Permanente Southern California over time in 2020.
[PNG File , 30 KB - jmir_v23i9e29959_app1.png]

Multimedia Appendix 2
Video visit rates within Kaiser Permanente Southern California over time by age, sex, race/ethnicity, and neighborhood median household income in 2020.
[PNG File , 162 KB - jmir_v23i9e29959_app2.png]

References

1. CMS Releases Recommendations on Adult Elective Surgeries, Non-Essential Medical, Surgical, and Dental Procedures During COVID-19 Response. Centers for Medicare and Medicaid Services. URL: https://www.cms.gov/newsroom/press-releases/cms-releases-recommendations-adult-elective-surgeries-non-essential-medical-surgical-and-dental [accessed 2021-04-15]

2. Using Telehealth to Expand Access to Essential Health Services during the COVID-19 Pandemic. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/telehealth.html [accessed 2021-04-15]
3. Baum A, Kaboli PJ, Schwartz MD. Reduced In-Person and Increased Telehealth Outpatient Visits During the COVID-19 Pandemic. Ann Intern Med 2021 Jan;174(1):129-131 [FREE Full text] [doi: 10.7326/M20-3026] [Medline: 32776780]
4. Patel SY, Mehrotra A, Huskamp HA, Uscher-Pines L, Ganguli I, Barnett ML. Trends in Outpatient Care Delivery and Telemedicine During the COVID-19 Pandemic in the US. JAMA Intern Med 2021 Mar 01;181(3):388-391. [doi: 10.1001/jamainternmed.2020.5928] [Medline: 33196765]
5. Xu S, Glenn S, Sy L, Qian L, Hong V, Ryan DS, et al. Impact of the COVID-19 Pandemic on Health Care Utilization in a Large Integrated Health Care System: Retrospective Cohort Study. J Med Internet Res 2021 Apr 29;23(4):e26558 [FREE Full text] [doi: 10.2196/26558] [Medline: 33882020]
6. Chunara R, Zhao Y, Chen J, Lawrence K, Testa P, Nov O, et al. Telemedicine and healthcare disparities: a cohort study in a large healthcare system in New York City during COVID-19. J Am Med Inform Assoc 2021 Jan 15;28(1):33-41 [FREE Full text] [doi: 10.1093/jamia/ocaa217] [Medline: 32866264]
7. Eberly LA, Kallan MJ, Julien HM, Haynes N, Khatana SAM, Nathan AS, et al. Patient Characteristics Associated With Telemedicine Access for Primary and Specialty Ambulatory Care During the COVID-19 Pandemic. JAMA Netw Open 2020 Dec 01;3(12):e2031640 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.31640] [Medline: 33372974]
8. A Multilayered Analysis of Telehealth: How This Emerging Venue of Care Is Affecting the Healthcare Landscape. Fair Health. 2019. URL: https://s3.amazonaws.com/media2.fairhealth.org/whitepaper/asset/A%20Multilayered%20Analysis%20of%20Telehealth%20-%20A%20FAIR%20Health%20White%20Paper.pdf [accessed 2021-04-15]
9. Call KT, McAlpine DD, Garcia CM, Shippee N, Beebe T, Adeniyi TC, et al. Barriers to care in an ethnically diverse publicly insured population: is health care reform enough? Med Care 2014 Aug;52(8):720-727. [doi: 10.1097/MLR.0000000000000172] [Medline: 25023917]
10. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and Facilitators to the Implementation of eHealth Services: Systematic Literature Analysis. J Med Internet Res 2019 Nov 22;21(11):e14197 [FREE Full text] [doi: 10.2196/14197] [Medline: 31755869]
11. Nouri S, Khoong E, Lyles C, Karliner L. Addressing Equity in Telemedicine for Chronic Disease Management During the Covid-19 Pandemic. NEJM Catalyst: Innovations in Care Delivery 2020 May [FREE Full text] [doi: 10.1056/CAT.20.0123]
12. Yuan N, Pevnick JM, Botting PG, Elad Y, Miller SJ, Cheng S, et al. Patient Use and Clinical Practice Patterns of Remote Cardiology Clinic Visits in the Era of COVID-19. JAMA Netw Open 2021 Apr 01;4(4):e214157 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.4157] [Medline: 33818619]
13. Koebnick C, Langer-Gould AM, Gould MK, Chao CR, Iyer RL, Smith N, et al. Sociodemographic characteristics of members of a large, integrated health care system: comparison with US Census Bureau data. Perm J 2012;16(3):37-41 [FREE Full text] [doi: 10.7812/tpp/12-031] [Medline: 23012597]
14. Hechter RC, Qian L, Sy LS, Greene SK, Weintraub ES, Naleway AL, et al. Secular trends in diagnostic code density in electronic healthcare data from health care systems in the Vaccine Safety Datalink project. Vaccine 2013 Feb 04;31(7):1080-1085. [doi: 10.1016/j.vaccine.2012.12.030] [Medline: 23267842]
15. Roberts ET, Mehrotra A. Assessment of Disparities in Digital Access Among Medicare Beneficiaries and Implications for Telemedicine. JAMA Intern Med 2020 Oct 01;180(10):1386-1389 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2666] [Medline: 32744601]
16. Lam K, Lu AD, Shi Y, Covinsky KE. Assessing Telemedicine Unreadiness Among Older Adults in the United States During the COVID-19 Pandemic. JAMA Intern Med 2020 Oct 01;180(10):1389-1391 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2671] [Medline: 32744593]
17. Weber E, Miller SJ, Astha V, Janevic T, Benn E. Characteristics of telehealth users in NYC for COVID-related care during the coronavirus pandemic. J Am Med Inform Assoc 2020 Dec 09;27(12):1949-1954 [FREE Full text] [doi: 10.1093/jamia/ocaa216] [Medline: 32866249]
18. Campos-Castillo C, Anthony D. Racial and ethnic differences in self-reported telehealth use during the COVID-19 pandemic: a secondary analysis of a US survey of internet users from late March. J Am Med Inform Assoc 2021 Jan 15;28(1):119-125 [FREE Full text] [doi: 10.1093/jamia/ocaa221] [Medline: 32894772]

Abbreviations

CDC: Centers for Disease Control and Prevention
KPSC: Kaiser Permanente Southern California

Edited by G Eysenbach; submitted 26.04.21; peer-reviewed by N Yuan, T Jamieson; comments to author 18.05.21; revised version received 18.06.21; accepted 02.08.21; published 01.09.21.

Please cite as:
Qian L, Sy LS, Hong V, Glenn SC, Ryan DS, Morrissette K, Jacobsen SJ, Xu S
Disparities in Outpatient and Telehealth Visits During the COVID-19 Pandemic in a Large Integrated Health Care Organization: Retrospective Cohort Study
J Med Internet Res 2021;23(9):e29959
URL: https://www.jmir.org/2021/9/e29959
doi:10.2196/29959
PMID:34351865

Original Paper

Adherence to Telemonitoring Therapy for Medicaid Patients With Hypertension: Case Study

Sulki Park[1,2], MS; Hye-Chung Kum[1,2,3], MS, MSW, PhD; Michael A Morrisey[3], PhD; Qi Zheng[4], PhD; Mark A Lawley[1,2,4], MS, PhD

[1]Population Informatics Lab, Texas A&M University, College Station, TX, United States

[2]Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX, United States

[3]Department of Health Policy and Management, Texas A&M University, College Station, TX, United States

[4]Department of Epidemiology and Biostatistics, Texas A&M University, College Station, TX, United States

Corresponding Author:
Hye-Chung Kum, MS, MSW, PhD
Department of Health Policy and Management
Texas A&M University
212 Adriance Lab Rd
College Station, TX, 77843
United States
Phone: 1 979 436 9439
Email: kum@tamu.edu

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2022/6/e39666

Abstract

Background: Almost 50% of the adults in the United States have hypertension. Although clinical trials indicate that home blood pressure monitoring can be effective in managing hypertension, the reported results might not materialize in practice because of patient adherence problems.

Objective: The aims of this study are to characterize the adherence of Medicaid patients with hypertension to daily telemonitoring, identify the impacts of adherence reminder calls, and investigate associations with blood pressure control.

Methods: This study targeted Medicaid patients with hypertension from the state of Texas. A total of 180 days of blood pressure and pulse data in 2016-2018 from a telemonitoring company were analyzed for mean transmission rate and mean blood pressure change. The first 30 days of data were excluded because of startup effects. The protocols required the patients to transmit readings by a specified time daily. Patients not transmitting their readings received an adherence reminder call to troubleshoot problems and encourage transmission. The patients were classified into adherent and nonadherent cohorts; adherent patients were those who transmitted data on at least 80% of the days.

Results: The mean patient age was 73.2 (SD 11.7) years. Of the 823 patients, 536 (65.1%) were women, and 660 (80.2%) were urban residents. The adherent cohort (475/823, 57.7%) had mean transmission rates of 74.9% before the adherence reminder call and 91.3% after the call, whereas the nonadherent cohort (348/823, 42.3%) had mean transmission rates of 39% and 58% before and after the call, respectively. From month 1 to month 5, the transmission rates dropped by 1.9% and 10.2% for the adherent and nonadherent cohorts, respectively. The systolic and diastolic blood pressure values improved by an average of 2.2 and 0.7 mm Hg (P<.001 and P=.004), respectively, for the adherent cohort during the study period, whereas only the systolic blood pressure value improved by an average of 1.6 mm Hg (P=.02) for the nonadherent cohort.

Conclusions: Although we found that patients can achieve high levels of adherence, many experience adherence problems. Although adherence reminder calls help, they may not be sufficient. Telemonitoring lowered blood pressure, as has been observed in clinical trials. Furthermore, blood pressure control was positively associated with adherence.

(J Med Internet Res 2021;23(9):e29018) doi:10.2196/29018

KEYWORDS

telemedicine; hypertension; Medicaid; blood pressure; pulse; telemonitoring; mobile phone

Introduction

Background

Hypertension affects nearly half of the adults in the United States, costs approximately US $131 billion annually, and is a major risk factor for cardiovascular disease and stroke [1-3]. Researchers have estimated the hazard ratios of cardiovascular events, stroke, and all-cause mortality to be 1.11-1.42, 1.28-1.40, and 1.02-1.13, respectively, for every 10 mm Hg increase in the ambulatory systolic blood pressure value [4]. Nonetheless, approximately only 1 in 4 adults with hypertension have their blood pressure under control [1].

Home blood pressure monitoring is an emerging strategy to help control hypertension, with many medical organizations recommending its use in diagnosis, distinguishing between blood pressure phenotypes, and ongoing hypertension management [5]. Several meta-analyses of published clinical trials have found evidence that home blood pressure monitoring can lead to clinically significant reductions in blood pressure values when accompanied by additional support services such as medication titration, education, and lifestyle counseling [6-9].

Although clinical trials are the gold standard, the reported results may not materialize in practice. Physicians have concerns about instrumentation quality, patient skills in taking readings, regular recording and transmission of results, and adherence to a regimen of routine measurement [7]. The 2010 and 2014 surveys of Canadian patients at a hypertension clinic where patients were encouraged to conduct home blood pressure monitoring found that only 39.2% and 40.6%, respectively, reported blood pressure more than 80% of the time [10]. Thus, poor patient adherence to daily monitoring and reporting could significantly undermine the positive effects observed in clinical trials.

Objectives

The aims of this study are to (1) investigate how well Texas Medicaid patients adhere to daily blood pressure and pulse rate monitoring when supported by a daily telemonitoring services company, (2) determine whether an adherence reminder call intervention improves the daily transmission rate, and (3) investigate any association between daily adherence and blood pressure control.

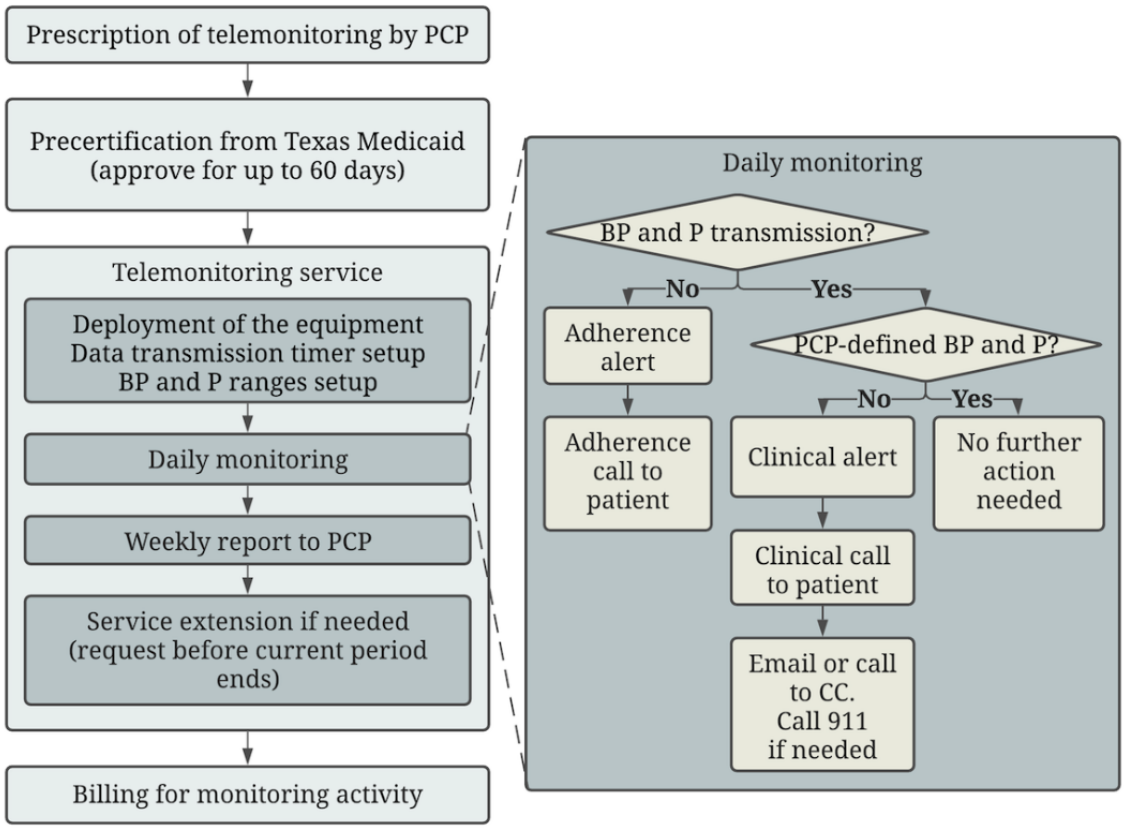
We note that insurance coverage is a requisite for daily telemonitoring. Medicare began paying for home monitoring in November 2018 [11], and Texas Medicaid reimburses physician-prescribed home telemonitoring for hypertension for 60 days, with reauthorization for additional monitoring at physician request [12,13]. More generally, reimbursement rates for home monitoring services vary significantly among states and insurers and have an uncertain future [5]. Although the temporary support for telehealth services by the Centers for Medicare & Medicaid Services and private insurers as a result of the coronavirus pandemic may lead to permanent changes in the delivery of routine care [14,15], the future of home telemonitoring coverage is unclear. In this uncertain environment, the analysis of real-world telemonitoring implementations is of great importance.

Methods

Monitoring Protocol

In this study, a telemonitoring company provided historical telemonitoring data (from January 2016 to December 2018) for Medicaid clients with hypertension in the state of Texas. The monitoring protocol (Figure 1) required patients to be referred by their physician. After Medicaid approval, a company technology deployer visited the patient's home to set up the equipment and provide training. The equipment—Food and Drug Administration–approved devices—consisted of a monitoring device with Bluetooth technology and a signal transmission unit that transferred the monitoring results to the company's cloud storage. No internet connection or smartphone was required. The training protocols and materials were developed based on American Medical Association guidelines [16]. The patients received education on how to use the equipment to take proper readings and were informed about the company's protocols for responding to the patients' technical or clinical needs. The patients were asked to select a daily time by which they would check and transmit their readings. If transmission did not occur by that time, an automated alert prompted a company staff member to make an adherence reminder call to the patient to troubleshoot any technical issues and to ask the patient to check and transmit the readings. Once the patient's data were received, if the blood pressure reading or pulse rate fell outside the physician-defined acceptable ranges, an automated clinical alert was transmitted to a company nurse. The nurse placed a clinical phone call to the patient, categorized the extent of concern, and contacted the provider by email for the lowest level of concern and by both email and phone call for more severe concerns. The company provided weekly summary reports to each provider for the enrolled patients. Under Texas Medicaid rules, a request for reauthorization of the telemonitoring therapy was made every 60 days when the physician prescribed additional monitoring. Texas Medicaid paid as much as US $1074.60 for 60 days of monitoring [12,13].

Figure 1. Workflow processes for telemonitoring service. BP: blood pressure; CC: clinical contact; P: pulse; PCP: primary care physician.



Data and Design

The telemonitoring company provided historical telemonitoring data for Texas Medicaid clients using their service. Only clients with 180 days or more of home telemonitoring were included in this study. The first 30 days were regarded as a startup period during which the patients learned to use the equipment to measure their vital signs. Data from the first 30 days were excluded from this study; thus, the study period was 150 days (months 1-5).

The number of transmissions before and after the adherence reminder calls was recorded each day, as was the number of adherence reminder calls made. We included all attempted adherence reminder calls, even those that the patients did not answer, because, in these cases, voice mail was left whenever possible. Daily systolic and diastolic blood pressure values were also collected to investigate improvements in blood pressure values during the study period. The mean systolic and diastolic blood pressure values at month 5 for each patient were calculated and compared with those at month 1. If the blood pressure values of the patient were missing for the entire month, that patient was excluded from this analysis.

The patients were separated into adherent and nonadherent cohorts; adherent patients were those who transmitted blood pressure and pulse values on at least 120 of the 150 days (at least 80% of the days, the same threshold used in the study by Milot et al [10]).

This study was approved by the institutional review board of Texas A&M University.

Statistical Analysis

To determine whether the patient baseline characteristics differed by population subgroups, we used chi-square tests for categorical variables and two-tailed *t* tests for continuous variables. In addition, z tests for the equity of the two proportions were performed to examine whether the rates of transmission differed by population subgroups. Paired *t* tests were performed to analyze the changes in blood pressure values at month 5 by comparing them with those at month 1 for each subgroup. Independent *t* tests were used to compare the changes in blood pressure values between the population subgroups. Analyses were conducted using SAS version 9.4 (SAS Institute).

Results

Patient Characteristics

The data of 2093 clients enrolled in hypertension telemonitoring were provided. Of the 2093 patients, 1325 (63.31%) transmitted data at least once, and 823 (39.32%) transmitted data throughout a continuous 180-day period. Table 1 summarizes their characteristics. The mean age of the participants was 73.2 (SD 11.7) years, and 65.1% (536/823) were women. All patients included in this study were diagnosed with hypertension and were on pharmaceutical therapy.

Table 1. Demographics and nonalert ranges for overall, adherent, and nonadherent cohorts (N=823).

Characteristics	Patients		
	Overall (N=823)	Adherent (n=475)	Nonadherent (n=348)
Age (years), mean (SD)	73.2 (11.7)	73.8 (10.9)	72.3 (12.6)
Women, n (%)	536 (65.1)	301 (63.4)	235 (67.5)
Area of residence[a] , n (%)			
Dallas	34 (4.1)	13 (2.7)	21 (6)
Houston	26 (3.2)	11 (2.3)	15 (4.3)
McAllen	648 (78.7)	368 (77.5)	280 (80.5)
San Antonio	115 (14)	83 (17.5)	32 (9.2)
Urban-rural classification[a] , n (%)			
Urban	660 (80.2)	360 (75.8)	300 (86.2)
Suburban or rural	163 (19.8)	115 (24.2)	48 (13.8)
Assigned nonalert range for systolic blood pressure, n (%)			
Default (90-160 mm Hg)	731 (88.8)	424 (89.3)	307 (88.2)
Personalized	92 (11.2)	51 (10.7)	41 (11.8)
Assigned nonalert range for diastolic blood pressure, n (%)			
Default (60-90 mmHg)	725 (88.1)	422 (88.8)	303 (87.1)
Personalized	98 (11.9)	53 (11.2)	45 (12.9)
Assigned nonalert range for pulse, n (%)			
Default (60-120 bpm)	724 (88)	421 (88.6)	303 (87.1)
Personalized	99 (12)	54 (11.4)	45 (12.9)

[a]*P*<.001. Patient characteristics differ between adherent and nonadherent cohorts.

Most of the participants (648/823, 78.7%) were from McAllen in south Texas near the Mexican border, and most of them (660/823, 80.2%) resided in urban areas. Of the 823 participants, 731 (88.8%), 725 (88.1%), and 724 (88%) participants had acceptable systolic blood pressure, diastolic blood pressure, and pulse ranges of 90-160 mm Hg, 60-90 mm Hg, and 60-120 bpm, respectively, which were defined by their primary care physician. The remaining 92 (11.2%), 98 (11.9%), and 99 (12%) participants had customized acceptable values above or below these ranges (55-200 mm Hg, 50-120 mmHg, and 50-120 bpm for systolic blood pressure, diastolic blood pressure, and pulse ranges, respectively). Table 1 also provides descriptive characteristics of the adherent and nonadherent cohorts. The characteristics across the two cohorts were similar, although the adherent cohort had a higher proportion of suburban or rural patients, with more of them living in south Texas (*P*<.001).

Adherence

Figure 2 shows the transmission rates (calculated using the following formula: transmission rate = 100 × total number of patients who transmitted readings / 823 patients) over the 5-month (150-day) period. The overall mean transmission rates across all 5 months were 59.7% before the adherence reminder call and 77.2% after the call (*P*<.001). The mean transmission rates for the first month were 61.6% and 79.1% before and after the call, respectively. These values declined until the fifth month when they reached 56.2% and 73.7% before and after the call, respectively. As indicated by the orange area in Figure 2, an average of 17.6% of the data transmissions were received after an adherence reminder call. However, the percentage of participants not transmitting after an adherence reminder call increased from 15.9% in the first month to 21.5% in the fifth month.

These aggregate findings mask large differences between the adherent and nonadherent cohorts (Figure 3). The adherent cohort was much more likely to transmit data without an adherence reminder call, with an overall mean transmission rate of 74.9% compared with only 39% for the nonadherent cohort (*P*<.001). After the adherence reminder call, these values increased to 91.3% and 58% (*P*<.001), respectively.

Figure 2. Monthly transmission rates for all patients over 150 days of telemonitoring. Mean transmission rate before (after) adherence call: 59.7% (77.2%).

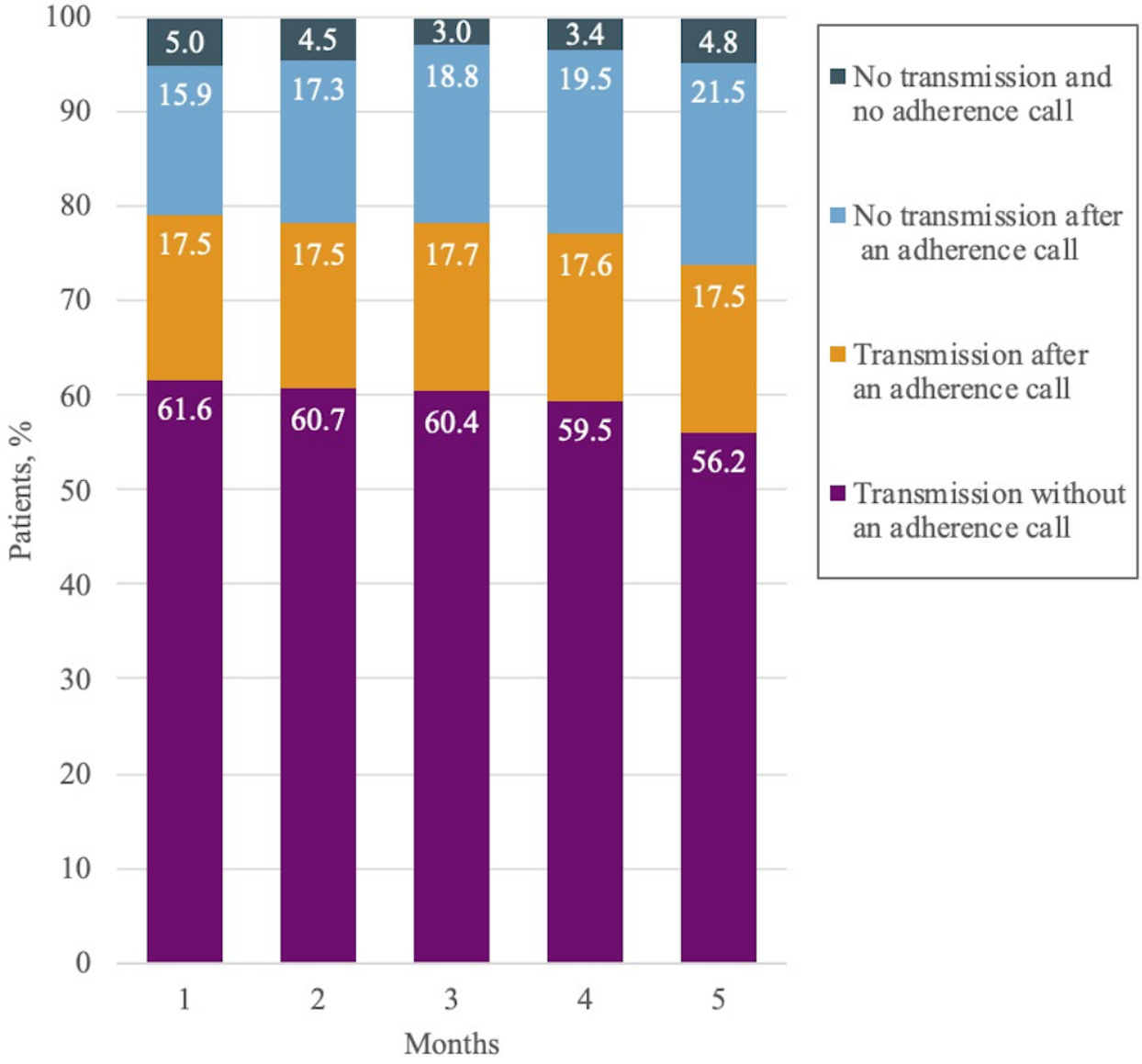
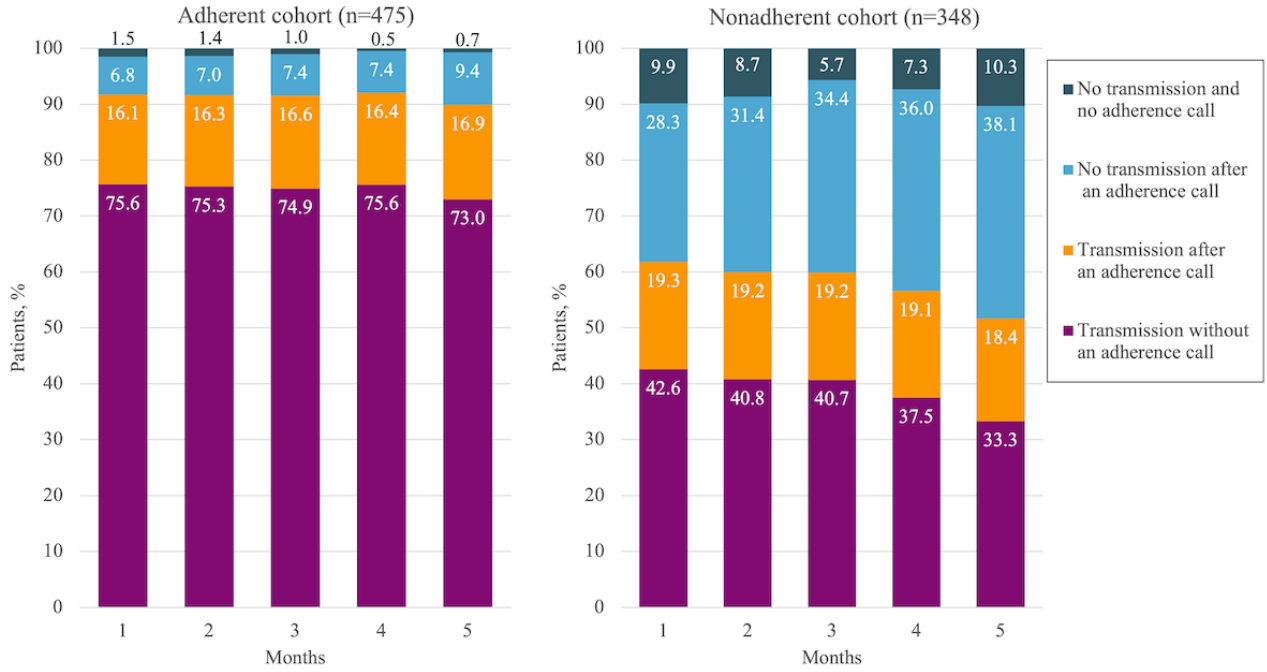


Figure 3. Monthly transmission rates for the adherent and nonadherent cohorts over 150 days of telemonitoring. Mean transmission rate before (after) adherence call: adherent cohort, 74.9% (91.3%); nonadherent cohort, 39% (58%).



The mean transmission rates for the first month were 75.6% before the adherence reminder call and 91.7% after the call for the adherent cohort and 42.6% and 61.9% before and after the call, respectively, for the nonadherent cohort. These values fluctuated and declined until the fifth month when they reached 73% and 89.9% before and after the call, respectively, for the adherent cohort and 33.3% and 51.7% before and after the call, respectively, for the nonadherent cohort. On average, an additional 16.5% and 19% transmissions were received after an adherence reminder call from the adherent and nonadherent cohorts, respectively (P<.001).

The percentage of participants not transmitting after an adherence reminder call was, on average, 7.6% for the adherent cohort and 33.6% for the nonadherent cohort. These values increased from 6.8% in the first month to 9.4% in the fifth month for the adherent cohort and from 28.3% in the first month to 38.1% in the fifth month for the nonadherent cohort (P<.001). We noted that, on average, 8.4% of the nonadherent participants who did not transmit data by the specified time failed to receive an adherence reminder call. This value increased to 10.3% in the fifth month of monitoring. In contrast, only 1.02% of the adherent cohort who did not transmit data failed to receive an adherence reminder call.

As might be expected, adherence was lowest on weekends (Tables 2 and 3), especially on Sundays, when the transmission rate (after the adherence reminder call) dropped to 88.4% and 46.3% for the adherent and nonadherent cohorts, respectively. The Sunday transmission rate was also observed to decrease over the 5-month period from 88.7% to 87.3% for the adherent cohort and from 49.6% to 41.8% for the nonadherent cohort.

Table 2. Weekday adherence by month for the adherent cohort (N=475).

	Adherence (%)						
	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Month							
Month 1	88.7	92.9	93.3	93.4	92.5	91.8	90
Month 2	87.8	92.5	92.8	92.1	92.6	92.7	90.8
Month 3	88.4	91.9	93.1	92.9	92.8	92.4	89.1
Month 4	89.6	92.9	91.8	93.4	93.1	93.1	90.5
Month 5	87.3	91.5	91	90.4	90.7	90.1	88.4
Value, mean (SD)	88.4 (0.8)	92.3 (0.6)	92.4 (0.9)	92.4 (1.1)	92.3 (0.8)	92 (1)	89.8 (0.9)

Table 3. Weekday adherence by month for the nonadherent cohort (N=348).

	Adherence (%)						
	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Month							
Month 1	49.6	66.9	67.1	68.6	65.6	63	52.8
Month 2	47.9	64.9	64.2	64.9	64.9	61.8	51.7
Month 3	46.7	64.4	64.1	65.2	63.8	62.1	52.1
Month 4	45.7	60.2	61	61.6	59.9	57.7	50.1
Month 5	41.8	55	55.5	56	55.6	54.2	44
Value, mean (SD)	46.3 (2.6)	62.3 (4.2)	62.4 (3.9)	63.3 (4.2)	62 (3.7)	59.7 (3.3)	50.1 (3.2)

Along with adherence to the daily protocol, the data also indicated whether the transmissions received were in or out of the physician-specified range. The average percentage of transmissions in range (calculated using the following formula: average percentage of transmissions in range = 100 × [number of transmissions in range / total number of transmissions]) was found to be 60.9% (SD 26%) for the adherent cohort and 53.9% (SD 24.9%) for the nonadherent cohort. The percentage in range increased for both cohorts over the 5-month period, indicating that telemonitoring was effective, from 59.2% in month 1 to 62.3% in month 5 for the adherent cohort and 49.8% in month 1 to 56.7% in month 5 for the nonadherent cohort.

Finally, the data indicated that the transmission results for 2 consecutive days were related. Note that for any given day, there were three possible outcomes: the patient did not transmit, the patient transmitted an out-of-range reading (blood pressure values, pulse rate, or both) or the patient transmitted an in-range reading. We refer to these as *transmission events*. Frequency analysis indicated an association between the transmission events observed on consecutive days. This is explored in the following sections. The percentages are listed in Table 4.

Table 4. Next day transition (N=823).

From	To (%)					
	Adherent cohort (n=475)			Nonadherent cohort (n=348)		
	NT[a]	ORT[b]	IRT[c]	NT	ORT	IRT
NT	32.9	29.3	37.8	61.7	18.5	19.8
ORT	5.6	41.6	52.7	28.9	34.8	36.3
IRT	6.3	32	61.7	29.6	28	42.5

[a]NT: no transmission.

[b]ORT: out-of-range transmission.

[c]IRT: in-range transmission.

Trends in Transmission Events Between 2 Consecutive Days

Adherence on the day after a missed transmission was far below the overall average for both adherent (67.1% vs an average of 91.3%) and nonadherent (38.3% vs an average of 58%) cohorts. Furthermore, the transmissions that were received the day after a missed transmission were less likely to be in range than the average for both cohorts. For the adherent cohort, 37.8% of the missed transmissions were followed by in-range transmissions, indicating that 56.3% (100 × [37.8 / 67.1]) of the transmissions received the day after a missed transmission were in range, whereas for the nonadherent cohort, 19.8% of the missed transmissions were followed by in-range transmissions, indicating that 51.6% (100 × [19.8 / 38.3]) of the transmissions received the day after a missed transmission were in range.

Adherence and in-range transmission after out-of-range transmission also showed similar patterns across the 2 cohorts. For the adherent cohort, out-of-range transmissions were followed by 41.6% of out-of-range transmissions, 52.7% of in-range transmissions, and only 5.6% of no transmissions the next day. For the nonadherent cohort, out-of-range transmissions were followed by 34.8% of out-of-range transmissions, 36.3% of in-range transmissions, and 28.9% of no transmissions the next day. Thus, adherence after an out-of-range day was greater than the overall average (94.6% vs an average of 91.3% for the adherent cohort and 71.1% vs an average of 58% for the nonadherent cohort). Furthermore, the transmissions that were received after an out-of-range transmission were less likely to be in range than the overall average (55.7%—100 × [52.7 / 94.6]—vs an average of 60.9% for the adherent cohort and 51%—100 × [36.3 / 71.1]—vs an average of 53.8% for the nonadherent cohort). It is worth noting that when an adherent patient transmitted an out-of-range reading, the next in-range transmission occurred within 2-3 days on average, that is, it took 2-3 days to resolve whatever problem was causing the out-of-range reading and for the patient to regain blood pressure and pulse rate control. However, when a nonadherent patient

transmitted an out-of-range reading, the next in-range transmission did not occur for 5-6 days on average, indicating that nonadherent patients were likely to experience elevated levels of blood pressure or pulse rate over a longer period.

Finally, adherence and in-range transmission after an in-range transmission also had similar patterns across the 2 cohorts, with better adherence and more in-range transmissions on the following day. For the adherent cohort, in-range transmissions were followed by 61.7% of in-range transmissions, 32% of out-of-range transmissions, and only 6.3% of no transmissions the next day. For the nonadherent cohort, in-range transmissions were followed by 42.5% of in-range transmissions, 28% of out-of-range transmissions, and 29.6% of no transmissions the next day. Thus, adherence after an in-range day was greater than the overall average (93.7% vs an average of 91.3% for the adherent cohort and 70.5% vs an average of 58% for the nonadherent cohort). Furthermore, the transmissions that were received after an in-range transmission were more likely to be in-range again the next day than the overall average (65.8%—100 × [61.7 / 93.7]—vs an average of 60.9% for the adherent cohort and 60.3%—100 × [42.5 / 70.5]—vs an average of 53.8% for the nonadherent cohort).

Relationship Between Daily Adherence and Blood Pressure Control

Overall, we found that the systolic blood pressure values of the adherent cohort improved by an average of 2.2 mm Hg (*P*<.001) over 5 months, whereas those of the nonadherent cohort improved by an average of 1.6 mm Hg (*P*=.02; Table 5). This improvement in the adherent cohort was significantly higher than that in the nonadherent cohort (*P*=.049).

Table 5. Systolic blood pressure changes between month 1 and month 5 (n=781).

	Adherent cohort (n=475; mm Hg)	Nonadherent cohort (n=306[a]; mm Hg)	*P* value[b]
Month 1			
Value, mean (SD)	133.7 (12.5)	137.9 (15.0)	N/A[c]
Month 5			
Value, mean (SD)	131.4 (12.2)	136.3 (14.4)	N/A
Comparison between month 1 and month 5			
Value, mean (SD)	2.2 (9.5)	1.6 (12.0)	.049
P value[d]	<.001	.02	N/A

[a]A total of 42 patients were excluded because of missing data.

[b]A two-tailed independent *t* test was performed to compare the systolic blood pressure changes between the adherent and nonadherent cohorts.

[c]N/A: not applicable.

[d]A two-tailed paired *t* test was performed to analyze the differences in systolic blood pressure values between month 1 and month 5 for each cohort.

Furthermore, of the 21 patients with an average systolic and diastolic blood pressure reading of more than 140 and 90 mm Hg for the first month, we found that the systolic blood pressure of the adherent patients (7/21, 33%) improved by an average of 14.8 mm Hg (*P*=.02) over 5 months, whereas that of the nonadherent patients (14/21, 67%) improved by an average of 10.6 mm Hg over 5 months, which was not significantly different (*P*=.11). The diastolic blood pressure of the adherent patients improved by an average of 0.7 mm Hg (*P*=.004) over 5 months, whereas the improvement over 5 months was not significant for nonadherent patients (0.4 mm Hg; *P*=.39; Table 6).

Table 6. Diastolic blood pressure changes between month 1 and month 5 (n=781).

Month	Adherent cohort (n=475; mm Hg)	Nonadherent cohort (n=306[a]; mm Hg)	*P* value[b]
Month 1			
Value, mean (SD)	71.5 (7.9)	74.0 (10.0)	N/A[c]
Month 5			
Value, mean (SD)	70.7 (7.9)	73.6 (9.8)	N/A
Comparison between month 1 and month 5			
Value, mean (SD)	0.7 (5.6)	0.4 (7.9)	.09
P value[d]	.004	.39	N/A

[a]A total of 42 patients were excluded because of missing data.

[b]A two-tailed independent *t* test was performed to compare the diastolic blood pressure changes between the adherent and nonadherent cohorts.

[c]N/A: not applicable.

[d]A two-tailed paired *t* test was performed to analyze the differences in diastolic blood pressure between month 1 and month 5 for each cohort.

Of the 21 patients with an average systolic and diastolic blood pressure reading of more than 140 and 90 mm Hg for the first month, we found that the diastolic blood pressure of adherent patients (7/21, 33%) improved by an average of 11.0 mm Hg (P=.02) over 5 months, whereas that of the nonadherent patients (14/21, 67%) improved over 5 months by an average of 9.7 mm Hg (P=.03).

Discussion

Principal Findings

This study suggests that telemonitoring for hypertension can achieve more than 70% adherence among Medicaid clients. Thus, most patients should be able to check and transmit their blood pressure values and pulse rate after the initial training. Furthermore, much higher levels of adherence (up to 90%) are possible for most patients (475/823, 57.7% of the patients in this study had 80% or more days of transmission) when telemonitoring is accompanied by adherence reminder calls. For these patients, adherence levels seemed to decline slightly over the 5-month period.

Furthermore, many Medicaid patients are likely to have trouble with daily adherence (348/823, 42.3% of the patients in this study). For these patients, adherence reminder calls can be helpful, but many daily transmissions will still be missed (approximately 13 days per patient per month in this study). Such patients can likely be identified within the first month of monitoring (not including the startup period), when their adherence rates without the adherence reminder call fall well below 50% (42.6% in this study). Indeed, 75% (260/348) of the patients in the nonadherent cohort in this study were not adherent in the first month of monitoring. For these patients, adherence rates can be expected to degrade significantly over time (by approximately 16% over 5 months in this study). Of the 823 patients, the 475 (57.7%) adherent patients and the 348 (42.3%) nonadherent patients together generated the need for approximately 350 adherence reminder calls per day, a significant workload. Regardless of the case, patients with adherence problems clearly need more than an adherence reminder call. Indeed, interventions that delve into health behaviors will likely be necessary (but perhaps not sufficient) to bring adherence levels up to 80% and beyond. The data suggest that such interventions should be targeted to weekends and to days after missed transmissions when the likelihood of poor adherence is higher.

Just as additional support for better adherence to daily monitoring is necessary, follow-up on an abnormal clinical condition is also important. A potential benefit of daily monitoring is that health care providers may recognize and address emerging problems before they become urgent. When readings are not transmitted, this opportunity may be lost. If we assume that the percentage of out-of-range transmissions can be applied to the days when data were not transmitted, we can estimate the number of missed transmissions that would have been out of range. Over 150 days of monitoring, this estimate turned out to be 5.1 days ([1 – 0.913] × [0.391] × 150) per patient for the adherent cohort and 29.0 days ([1 – 0.580] × [0.461] × 150) per patient for the nonadherent cohort. This represents a total of 12,528 days (29.04 × 348 + [5.10 × 475]) of unmet needs for 823 patients over 150 days of monitoring (approximately 15 days per patient). In other words, 10.2% (12,528 / [150 × 823]) of the required follow-up was missed because of lack of adherence (for the nonadherent cohort, this was approximately 20%).

On a positive note, it is encouraging that 58.9% (280/475) of the adherent patients and 54.9% (168/306) of the nonadherent patients experienced an improvement in systolic blood pressure values, and 52.2% (248/475) of the adherent patients and 51.6% (158/306) of the nonadherent patients experienced an improvement in diastolic blood pressure values. The mean systolic blood pressure values of both cohorts improved significantly during the study period, and these improvements were significantly higher in the adherent cohort (P=.049). The mean diastolic blood pressure value of the adherent cohort declined significantly during the study period, but the decline was not significant for the nonadherent patients. These results are consistent with those of clinical trials in the literature. In 18 clinical trials of home telemonitoring, the average improvement in systolic and diastolic blood pressure values was 12.1 and 6.3 mm Hg within 6 months [17-34]. Of these 18 clinical trials, eight were restricted to patients with systolic and diastolic blood pressure values of more than 140 and 90 mm Hg at baseline, and the other ten trials were restricted to those with blood pressure readings above or below these values. In our study with patients with systolic and diastolic blood pressure values of more than 140 and 90 mm Hg in the first month, the systolic and diastolic blood pressure values of the adherent patients improved by an average of 14.8 and 11.0 mm Hg, which is higher than the average improvement observed in the 18 clinical trials. However, for nonadherent patients with systolic and diastolic blood pressure values of more than 140 and 90 mm Hg in the first month, only the diastolic blood pressure value significantly improved by an average of 9.7 mm Hg.

Finally, it is important to appreciate that achieving improved adherence requires considerable effort. Patients must be trained in the correct procedures to monitor their blood pressure and pulse; staff members must monitor daily transmissions and contact patients to encourage participation and to resolve technical issues; and, as noted, additional interventions will be needed for many patients. Texas Medicaid payment levels may have been adequate for this level of intervention, but it is not clear whether Medicare or private insurers will reimburse this level of effort in the future. Clearly, the case for reimbursement would be compelling if hypertension telemonitoring could be shown to help avoid even a small number of hospitalizations for stroke and heart disease, which can be extremely expensive.

This study included some limitations. It only examined Texas Medicaid clients. It is not clear whether these findings are generalizable to Medicare, privately insured, or uninsured patients with hypertension. It is also unclear whether these findings are generalizable to people with other chronic conditions who would benefit from ongoing monitoring. This study was limited to patients who were referred to the monitoring program. The analysis would be strengthened if there were a control group to more rigorously examine adherence and the impact of the intervention. Finally, the

monitoring protocol required the data to be transmitted on a daily basis, which was more frequent than the general home blood pressure monitoring guidelines [5]. Excessive and frequent transmission requirements may negatively affect adherence and persistence. In contrast, daily monitoring could help with medication adherence and help avert emergency situations and hospitalizations.

Conclusions

Adherence reminder calls helped most patients with hypertension to achieve higher levels of adherence to blood pressure and pulse monitoring. Telemonitoring improved blood pressure control, similar to the improvement observed in clinical trials. Furthermore, more adherent patients achieved higher levels of blood pressure control. However, the study suggests that additional adherence interventions and support are needed for many patients to achieve high levels of adherence.

Acknowledgments

This study was supported in part by Industrial Seed Fund project from the Industry Board Fees of the NSF PATHS-UP ERC (#1648451), Population Informatics Lab, and the Texas Virtual Data Library at Texas A&M University. Real-world telemonitoring data was provided by Coordination Centric.

Authors' Contributions

SP, HCK, and MAL were involved in the study concept and design and were responsible for data acquisition, governance, and analysis. SP, HCK, MAM, QZ, and MAL were responsible for interpreting the data and drafting the manuscript.

Conflicts of Interest

None declared.

References

1. Centers for Disease Control and Prevention (CDC). Hypertension Cascade: Hypertension Prevalence, Treatment and Control Estimates Among US Adults Aged 18 Years and Older Applying the Criteria From the American College of Cardiology and American Heart Association's 2017 Hypertension Guideline—NHANES 2015–2018. Atlanta, GA: US Department of Health and Human Services; 2021. URL: https://millionhearts.hhs.gov/data-reports/hypertension-prevalence.html [accessed 2021-05-19]
2. Kirkland EB, Heincelman M, Bishu KG, Schumann SO, Schreiner A, Axon RN, et al. Trends in healthcare expenditures among US adults with hypertension: National Estimates, 2003-2014. J Am Heart Assoc 2018 May 30;7(11):e008731 [FREE Full text] [doi: 10.1161/JAHA.118.008731] [Medline: 29848493]
3. Centers for Disease Control and Prevention. Underlying Cause of Death, 1999–2018. CDC WONDER Online Database. Atlanta, GA: Centers for Disease Control and Prevention; 2018. URL: http://wonder.cdc.gov/ucd-icd10.html [accessed 2020-03-09]
4. Siu AL. Screening for high blood pressure in adults: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2015 Oct 13;163(10):778. [doi: 10.7326/m15-2223]
5. Shimbo D, Artinian NT, Basile JN, Krakoff LR, Margolis KL, Rakotz MK, American Heart Associationthe American Medical Association. Self-measured blood pressure monitoring at home: a joint policy statement from the American Heart Association and American Medical Association. Circulation 2020 Jul 28;142(4):42-63. [doi: 10.1161/CIR.0000000000000803] [Medline: 32567342]
6. Duan Y, Xie Z, Dong F, Wu Z, Lin Z, Sun N, et al. Effectiveness of home blood pressure telemonitoring: a systematic review and meta-analysis of randomised controlled studies. J Hum Hypertens 2017 Jul;31(7):427-437. [doi: 10.1038/jhh.2016.99] [Medline: 28332506]
7. Liyanage-Don N, Fung D, Phillips E, Kronish IM. Implementing home blood pressure monitoring into clinical practice. Curr Hypertens Rep 2019 Feb 12;21(2):14. [doi: 10.1007/s11906-019-0916-0] [Medline: 30747350]
8. Omboni S, Gazzola T, Carabelli G, Parati G. Clinical usefulness and cost effectiveness of home blood pressure telemonitoring: meta-analysis of randomized controlled studies. J Hypertens 2013 Mar;31(3):455-467. [doi: 10.1097/HJH.0b013e32835ca8dd] [Medline: 23299557]
9. Tucker KL, Sheppard JP, Stevens R, Bosworth HB, Bove A, Bray EP, et al. Self-monitoring of blood pressure in hypertension: A systematic review and individual patient data meta-analysis. PLoS Med 2017 Sep;14(9):e1002389 [FREE Full text] [doi: 10.1371/journal.pmed.1002389] [Medline: 28926573]
10. Milot J, Birnbaum L, Larochelle P, Wistaff R, Laskine M, Van Nguyen P, et al. Unreliability of home blood pressure measurement and the effect of a patient-oriented intervention. Can J Cardiol 2015 May;31(5):658-663. [doi: 10.1016/j.cjca.2015.03.006] [Medline: 25936491]
11. Holland TM. The new rules for remote patient monitoring reimbursement. Samsung Insights. 2020 Sep 02. URL: https://insights.samsung.com/2020/09/02/the-new-rules-for-remote-patient-monitoring-reimbursement-2/ [accessed 2020-11-18]

12. Texas Medicaid provider procedures manual vol 2. Telecommunication services handbook. Texas Medicaid & Healthcare Partnership (TMHP). 2020 Aug. URL: https://www.tmhp.com/sites/default/files/file-library/resources/provider-manuals/tmppm/pdf-chapters/2020/august/2_Telecommunication_Srvs.pdf [accessed 2020-11-17]
13. Rate changes. Texas Medicaid & Healthcare Partnership (TMHP). 2020. URL: https://www.tmhp.com/resources/rate-and-code-updates/rate-changes [accessed 2020-11-20]
14. Bloem BR, Dorsey ER, Okun MS. The Coronavirus Disease 2019 crisis as catalyst for telemedicine for chronic neurological disorders. JAMA Neurol 2020 Aug 01;77(8):927-928. [doi: 10.1001/jamaneurol.2020.1452] [Medline: 32329796]
15. Duffy EY, Cainzos-Achirica M, Michos ED. Primary and secondary prevention of cardiovascular disease in the era of the coronavirus pandemic. Circulation 2020 Jun 16;141(24):1943-1945. [doi: 10.1161/circulationaha.120.047194]
16. Remote patient monitoring implementation playbook. American Medical Association (AMA). 2018. URL: https://www.ama-assn.org/practice-management/digital/remote-patient-monitoring-implementation-playbook-overview [accessed 2021-05-20]
17. Artinian NT, Flack JM, Nordstrom CK, Hockman EM, Washington OG, Jen KC, et al. Effects of nurse-managed telemonitoring on blood pressure at 12-month follow-up among urban African Americans. Nurs Res 2007;56(5):312-322. [doi: 10.1097/01.NNR.0000289501.45284.6e] [Medline: 17846552]
18. Bosworth HB, Powers BJ, Olsen MK, McCant F, Grubber J, Smith V, et al. Home blood pressure management and improved blood pressure control: results from a randomized controlled trial. Arch Intern Med 2011 Jul 11;171(13):1173-1180. [doi: 10.1001/archinternmed.2011.276] [Medline: 21747013]
19. Bove AA, Homko CJ, Santamore WP, Kashem M, Kerper M, Elliott DJ. Managing hypertension in urban underserved subjects using telemedicine--a clinical trial. Am Heart J 2013 Apr;165(4):615-621. [doi: 10.1016/j.ahj.2013.01.004] [Medline: 23537980]
20. Carrasco MP, Salvador CH, Sagredo PG, Márquez-Montes J, de Mingo MA, Fragua JA, et al. Impact of patient-general practitioner short-messages-based interaction on the control of hypertension in a follow-up service for low-to-medium risk hypertensive patients: a randomized controlled trial. IEEE Trans Inf Technol Biomed 2008 Nov;12(6):780-791. [doi: 10.1109/TITB.2008.926429] [Medline: 19000959]
21. Earle KA, Istepanian RS, Zitouni K, Sungoor A, Tang B. Mobile telemonitoring for achieving tighter targets of blood pressure control in patients with complicated diabetes: a pilot study. Diabetes Technol Ther 2010 Jul;12(7):575-579. [doi: 10.1089/dia.2009.0090] [Medline: 20597833]
22. Friedman RH, Kazis L, Jette A, Smith MB, Stollerman J, Torgerson J, et al. A telecommunications system for monitoring and counseling patients with hypertension. Impact on medication adherence and blood pressure control. Am J Hypertens 1996 Apr;9(4 Pt 1):285-292. [doi: 10.1016/0895-7061(95)00353-3] [Medline: 8722429]
23. Green BB, Anderson ML, Cook AJ, Catz S, Fishman PA, McClure JB, et al. e-Care for heart wellness: a feasibility trial to decrease blood pressure and cardiovascular risk. Am J Prev Med 2014 Apr;46(4):368-377 [FREE Full text] [doi: 10.1016/j.amepre.2013.11.009] [Medline: 24650839]
24. Kim Y, Shin DG, Park S, Lee CH. Randomized clinical trial to assess the effectiveness of remote patient monitoring and physician care in reducing office blood pressure. Hypertens Res 2015 Jul;38(7):491-497. [doi: 10.1038/hr.2015.32] [Medline: 25787041]
25. Madsen LB, Kirkegaard P, Pedersen EB. Blood pressure control during telemonitoring of home blood pressure. A randomized controlled trial during 6 months. Blood Press 2008;17(2):78-86. [doi: 10.1080/08037050801915468] [Medline: 18568696]
26. Magid DJ, Ho P, Olson K, Brand DW, Welch LK, Snow KE, et al. A multimodal blood pressure control intervention in 3 healthcare systems. Am J Manag Care 2011 Apr;17(4):96-103. [Medline: 21774100]
27. Magid DJ, Olson KL, Billups SJ, Wagner NM, Lyons EE, Kroner BA. A pharmacist-led, American Heart Association Heart360 web-enabled home blood pressure monitoring program. Circ: Cardiovas Qual Outcomes 2013 Mar;6(2):157-163. [doi: 10.1161/circoutcomes.112.968172]
28. Margolis KL, Asche SE, Bergdall AR, Dehmer SP, Groen SE, Kadrmas HM, et al. Effect of home blood pressure telemonitoring and pharmacist management on blood pressure control: a cluster randomized clinical trial. J Am Med Assoc 2013 Jul 03;310(1):46-56 [FREE Full text] [doi: 10.1001/jama.2013.6549] [Medline: 23821088]
29. McKinstry B, Hanley J, Wild S, Pagliari C, Paterson M, Lewis S, et al. Telemonitoring based service redesign for the management of uncontrolled hypertension: multicentre randomised controlled trial. Br Med J 2013 May 24;346:f3030 [FREE Full text] [doi: 10.1136/bmj.f3030] [Medline: 23709583]
30. McManus RJ, Mant J, Bray EP, Holder R, Jones MI, Greenfield S, et al. Telemonitoring and self-management in the control of hypertension (TASMINH2): a randomised controlled trial. Lancet 2010 Jul;376(9736):163-172. [doi: 10.1016/s0140-6736(10)60964-6]
31. McManus RJ, Mant J, Haque MS, Bray EP, Bryan S, Greenfield SM, et al. Effect of self-monitoring and medication self-titration on systolic blood pressure in hypertensive patients at high risk of cardiovascular disease: the TASMIN-SR randomized clinical trial. J Am Med Assoc 2014 Aug 27;312(8):799-808. [doi: 10.1001/jama.2014.10057] [Medline: 25157723]

32. Parati G, Omboni S, Albini F, Piantoni L, Giuliano A, Revera M, TeleBPCare Study Group. Home blood pressure telemonitoring improves hypertension control in general practice. The TeleBPCare study. J Hypertens 2009 Jan;27(1):198-203. [doi: 10.1097/hjh.0b013e3283163caf] [Medline: 19145785]
33. Rifkin DE, Abdelmalek J, Miracle C, Low C, Barsotti R, Rios P, et al. Linking clinic and home: A randomized, controlled clinical effectiveness trial of real-time, wireless blood pressure monitoring for older patients with kidney disease and hypertension. Blood Press Monit 2013 Feb;18(1):8-15 [FREE Full text] [doi: 10.1097/MBP.0b013e32835d126c] [Medline: 23275313]
34. Wakefield BJ, Holman JE, Ray A, Scherubel M, Adams MR, Hillis SL, et al. Effectiveness of home telehealth in comorbid diabetes and hypertension: a randomized, controlled trial. Telemed J E Health 2011 May;17(4):254-261. [doi: 10.1089/tmj.2010.0176] [Medline: 21476945]

Edited by R Kukafka; submitted 22.03.21; peer-reviewed by J Pagan, S Omboni; comments to author 30.03.21; revised version received 31.05.21; accepted 05.07.21; published 06.09.21.

Please cite as:
Park S, Kum HC, Morrisey MA, Zheng Q, Lawley MA
Adherence to Telemonitoring Therapy for Medicaid Patients With Hypertension: Case Study
J Med Internet Res 2021;23(9):e29018
URL: https://www.jmir.org/2021/9/e29018
doi:10.2196/29018
PMID:34486977

Original Paper

Using Administrative Data to Explore Potentially Aberrant Provision of Virtual Care During COVID-19: Retrospective Cohort Study of Ontario Provincial Data

Vess Stamenova[1], PhD; Cherry Chu[1], MSc; Andrea Pang[2], MPH; Mina Tadrous[1,2,3], MSc, PhD, PharmD; R Sacha Bhatia[1,2,4], MD, MBA; Peter Cram[2,5], MD, MBA

[1]Institute for Health System Solutions and Virtual Care, Women's College Hospital, Toronto, ON, Canada

[2]Institute for Clinical Evaluative Sciences, Toronto, ON, Canada

[3]Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, ON, Canada

[4]Department of Medicine, University of Toronto, Toronto, ON, Canada

[5]School of Medicine, University of Texas Medical Branch, Galveston, TX, United States

Corresponding Author:
Vess Stamenova, PhD
Institute for Health System Solutions and Virtual Care
Women's College Hospital
76 Grenville St
Toronto, ON, M5S 1B2
Canada
Phone: 1 416 268 0985
Email: vess.stamenova@wchospital.ca

Abstract

Background: The COVID-19 pandemic has led to a rapid increase in virtual care use across the globe. Many health care systems have responded by creating virtual care billing codes that allow physicians to see their patients over telephone or video. This rapid liberalization of billing requirements, both in Canada and other countries, has led to concerns about potential abuse, but empirical data are limited.

Objective: The objectives of this study were to examine whether there were substantial changes in physicians' ambulatory visit volumes coinciding with the liberalization of virtual care billing rules and to describe the characteristics of physicians who significantly increased their ambulatory visit volumes during this period. We also sought to describe the relationship between visit volume changes in 2020 and the volumes of virtual care use among individual physicians and across specialties.

Methods: We conducted a population-based, retrospective cohort study using health administrative data from the Ontario Health Insurance Plan, which was linked to the ICES Physician Database. We identified a unique cohort of providers based on physicians' billings and calculated the ratio of total in-person and virtual ambulatory visits over the period from January to June 2020 (virtual predominating) relative to that over the period from January to June 2019 (in-person predominating) for each physician. Based on these ratios, we then stratified physicians into four groups: low-, same-, high-, and very high–use physicians. We then calculated various demographic and practice characteristics of physicians in each group.

Results: Among 28,383 eligible physicians in 2020, the mean ratio of ambulatory visits in January to June 2020:2019 was 0.99 (SD 2.53; median 0.81, IQR 0.59-1.0). Out of 28,383 physicians, only 2672 (9.4%) fell into the high-use group and only 291 (1.0%) fell into the very high–use group. High-use physicians were younger, more recent graduates, more likely female, and less likely to be international graduates. They also had, on average, lower-volume practices. There was a significant positive correlation between percent virtual care and the 2020:2019 ratio only in the group of physicians who maintained their practice (R=0.35, P<.001). There was also a significant positive correlation between the 2020:2019 ratio and the percent virtual care per specialty (R=0.59, P<.01).

Conclusions: During the early stages of the pandemic, the introduction of virtual care did not lead to significant increases in visit volume. Our results provide reassuring evidence that relaxation of billing requirements early in the COVID-19 pandemic in Ontario were not associated with widespread and aberrant billing behaviors. Furthermore, the strong relationship between the

ability to maintain practice volumes and the use of virtual care suggests that the introduction of virtual care allowed for continued access to care for patients.

(J Med Internet Res 2021;23(9):e29396) doi:10.2196/29396

KEYWORDS

telemedicine; virtual care; COVID-19; pandemic; virtual health; telehealth; ambulatory visits; physicians; patients; digital health

Introduction

The COVID-19 pandemic has led to a rapid increase in virtual care use across the globe [1-5]. In Ontario, Canada's largest province, virtual care increased from 1.6% of all ambulatory visits pre–COVID-19 to 71% during the first wave of the COVID-19 pandemic [3], a much higher rate compared to those reported in other countries such as the United States (30%) and Australia (42%) [4,6,7].

While Ontario had pre-existing virtual care billing codes before the onset of the pandemic, these codes were allowable for a single government-run online platform and only available to specialists and primary care physicians in rostered patient practices or specialized practices. Primary care physicians outside of rostered practices were not included in this model, in order to support continuity of care [8] and respond to growing concerns about fragmentation and poor quality of care received in virtual walk-in clinics as well as funding disruptions in Canada [9] and abroad [10].

In Ontario, the pandemic led to the introduction of temporary billing codes in mid-March 2020 that reimbursed any physician with identical amounts for in-person, video, or telephone visits and eliminated prior restrictions on practice type or allowable technology platforms. This rapid liberalization of billing requirements, both in Canada and other countries [4], has led to concerns about potential abuse, but empirical data are limited [11].

The objectives of this study were to examine whether there were substantial changes in physicians' ambulatory visit volumes coinciding with the liberalization of virtual care billing rules and to describe the characteristics of physicians who significantly increased their ambulatory visit volumes during this period. We also sought to describe the relationship between visit volume changes in 2020 and the volumes of virtual care use in individual physicians and across specialties.

Methods

We conducted a population-based, retrospective cohort study using health administrative data from the Ontario Health Insurance Plan, which was linked to the ICES Physician Database. Data sets were linked using unique encoded identifiers and analyzed at ICES, an independent, nonprofit research institute. Use of these databases for the purposes of this study was authorized under §45 of Ontario's Personal Health Information Protection Act, which does not require review by a research ethics board. An exemption letter was obtained by the Research Ethics Board at Women's College Hospital, Toronto, Ontario.

We identified a unique cohort of providers based on physicians' billings for in-person and virtual ambulatory visits. We excluded visits for non-Ontario residents and those with an invalid or missing health card number. We also excluded all physicians with clinical volumes that were inconsistent with an active practice during the pre–COVID-19 period (<10 ambulatory visits during the period from January to June 2019).

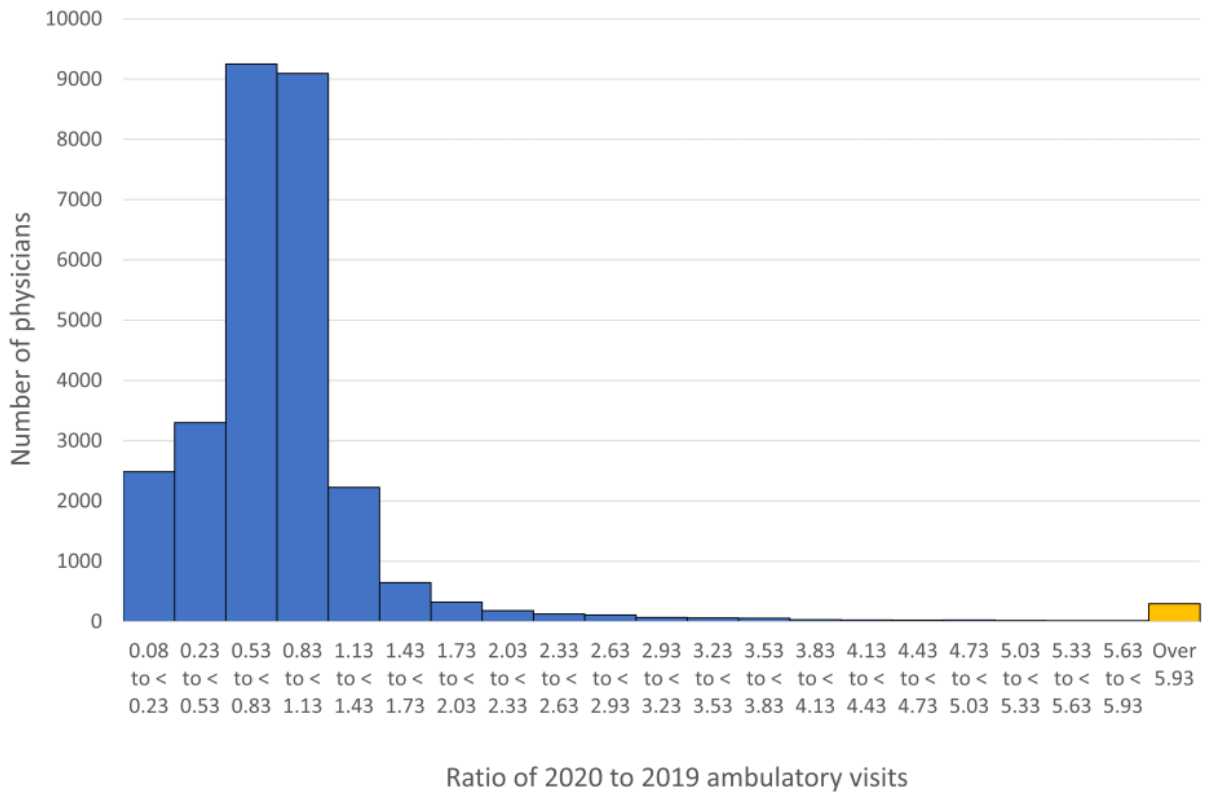
We then calculated the ratio of total in-person and virtual ambulatory visits over the period from January to June 2020 (virtual predominating) relative to that over the period from January to June 2019 (in-person predominating) for each physician. We included first-quarter data in 2020 as they cover the beginning of the pandemic. Data extending past the second quarter of 2020 were unavailable. Based on these ratios, we then stratified physicians into four groups: (1) low-use physicians had ratios from 0 to 0.50 (ie, a 50% or greater reduction in visits in 2020 compared to 2019), (2) same-use physicians were those with ratios over 0.50 but less than 1.25, (3) high-use physicians were those with ratios of at least 1.25 but less than 6.0, and (4) very high–use physicians were those with ratios equal to or greater than 6.0 (ie, an at-least 6-fold increase in visits in 2020 compared to 2019). To explore whether the proportions of physicians falling into each category differed much from previous years, we also calculated the number of physicians falling into each group—defined as the same ratio ranges—for the periods of January to June 2019 relative to January to June 2018.

For all physicians, we also obtained demographic and practice characteristics, including, age, sex, years since graduation, training location, practice type (ie, specialist, family practice, or focused family practice physicians focusing 50% or more of their practice in a specific type of care, such as psychotherapy [12]), and specialty. We also calculated the number of unique patients seen, number of total visits, number of virtual visits, and number of visits per day, virtual or any.

Results

Among 28,383 eligible physicians in 2020, the mean ratio of ambulatory visits in January to June 2020:2019 was 0.99 (SD 2.53; median 0.81, IQR 0.59-1.0). Only 291 physicians (1.0%) were very high users, 2672 physicians (9.4%) were high users, and 5422 (19.1%) were low users (Figure 1). In comparison, the previous year (2019:2018 visit ratio), among 27,709 eligible physicians, 289 (1.0%) were very high users, 3395 (12.3%) were high users, and 2937 (10.6%) were low users.

Figure 1. Histogram showing the total number of physicians by ambulatory visit volume ratio (2020:2019). Proportions of <1 indicate fewer visits in the period of January to June 2020 when compared to January to June 2019, while a proportion of >1 suggested increased visits in the period of January to June 2020. Note, the last bin in orange represents anybody with a ratio over 5.93, representing the top 1%.



High-use physicians were younger, more recent graduates, more likely to be female, and less likely to be international graduates than those who maintained their volumes (ie, same-use group) (P<.001) (Table 1). These effects were even more pronounced in the very high–use group. They were also more likely to be specialists than primary care providers (P<.001), and physicians in focused primary care practices were more likely to be high or very high users relative to specialists (P<.001).

Physicians in the two high-use groups had, on average, lower-volume practices in both 2020 and 2019 (P<.001), and volumes were especially smaller in 2019 with fewer patients seen (P<.001). The effect was larger for the very high users (P<.001). Providers in the two high-use groups had fewer visits per billing day and fewer total patients seen in 2020 (P<.001), but they had a higher percentage of virtual care visits (P<.001) and a similar number of virtual care visits per day compared to the same-use group (Table 1).

Specialties with a large percentage of their total physician population being in the high-use group included emergency medicine (41/230, 17.8%), psychiatry (378/2061, 18.3%), and internal medicine (147/999, 14.7%). The same specialties were common among the very high–use group (Table 2).

Pearson correlations between the 2020:2019 visit ratio and percent of visits completed virtually showed a significant positive correlation only among physicians from the same-use group (R=0.35, P<.001) (Figure 2).

Finally, we also calculated the Pearson correlation coefficient between the average 2020:2019 ratio per specialty and the percent virtual care used per specialty. We excluded emergency medicine, diagnostic radiology, and nuclear medicine, as they were outliers and had both the highest 2020:2019 visit ratios (1.6, 1.3, and 2.9, respectively) and the lowest percentages of virtual care (12.4%, 1.5%, and 8.2%, respectively). There was a significant positive correlation between the 2020:2019 ratio and the percent virtual care across specialties (R=0.59, P<.01) (Figure 3).

Table 1. Physician characteristics stratified by ratio of ambulatory visits in January to June 2020:2019.

Characteristic	Physicians					*P* value
	All (N=28,383)	Low use: 0 to 0.50 visits[a] (n=5422)	Same use: >0.50 to <1.25 visits (n=19,998)	High use: 1.25 to <6.0 visits[b] (n=2672)	Very high use: ≥6.0 visits (n=291)	
Age (years), mean (SD)	53.9 (11.2)	56.1 (12.6)	53.7 (10.8)	51.9 (11.24)	49.2 (11.5)	<.001
Years since graduation, mean (SD)	25.1 (13.3)	26.7 (15.5)	25.4 (12.5)	20.5 (12.60)	15.8 (11.7)	<.001
Sex (female), n (%)	11,864 (41.8)	2287 (42.5)	8116 (40.6)	1268 (48.4)	148 (55.4)	<.001
Canadian or internationally trained (Canadian trained), n (%)	16,218 (76.4)	2928 (77.2)	11,912 (75.7)	1282 (81.6)	96 (83.5)	<.001
Number of unique patients seen in 2019 (January to June)						
Mean (SD)	729.2 (795.0)	496.53 (675.0)	858.28 (828.8)	308.81 (433.4)	55.46 (60.5)	<.001
Median (IQR)	539 (163-999)	257 (75-689)	691 (295-1129)	153 (50-402)	35 (17-72)	<.001
Number of unique patients seen in 2020 (January to June)						
Mean (SD)	556.7 (618.1)	155.2 (297.0)	684.5 (639.5)	428.0 (577.4)	439.3 (408.9)	<.001
Median (IQR)	408 (100-789)	35 (0-183)	559 (241-910)	236 (81-567)	347 (133-590)	<.001
Percent virtual care visits in 2020 (January to June)						
Mean (SD)	31.9 (22.7)	14.95 (24.3)	34.03 (20.0)	40.3 (26.2)	42.3 (28.6)	<.001
Median (IQR)	34 (11-49)	0 (0-22)	37 (19-49)	44 (18-60)	42 (21-61)	<.001
Number of virtual care visits in 2020 (January to June)						
Mean (SD)	400.7 (571.0)	52.1 (139.5)	471.3 (570.6)	411.9 (771.3)	294.8 (315.1)	<.001
Median (IQR)	231 (30-551)	1 (0-37)	324 (98-639)	199 (32-494)	205 (70-432)	<.001
Number of total visits in 2020 (January to June)						
Mean (SD)	1101.2 (1228.9)	292.2 (484.9)	1300.0 (1257.7)	877.9 (1293.1)	738.0 (655.7)	<.001
Median (IQR)	780 (240-1503)	108 (23-370)	1004 (454-1715)	495 (151-1105)	538 (274-976)	<.001
Number of total visits in 2019 (January to June)						
Mean (SD)	1304.8 (1420.4)	800.5 (1103.6)	1561.5 (1483.3)	541.4 (855.9)	66.1 (73.7)	<.001
Median (IQR)	934 (286-1799)	392 (101-1128)	1212 (539-2078)	267 (74-658)	43 (20-84)	<.001
Number of visits per billing day in 2020 (January to June)						
Mean (SD)	12.1 (9.5)	8.1 (8.3)	13.2 (9.5)	9.7 (9.8)	10.2 (7.3)	<.001
Median (IQR)	9 (5-15)	5 (2-10)	11 (6-16)	7 (3-12)	7 (4-12)	<.001
Number of virtual visits per billing day in 2020 (January to June)						
Mean (SD)	8.6 (6.7)	5.0 (5.2)	9.1 (6.5)	8.1 (7.6)	7.4 (5.4)	<.001
Median (IQR)	7 (4-11)	3 (2-6)	7 (4-11)	6 (3-10)	6 (3-9)	<.001
Practice type, n (%)						
Specialist	15,201 (53.6)	3058 (56.4)	10,419 (52.1)	1590 (59.5)	134 (46.0)	<.001
Primary care provider	9393 (33.1)	1366 (25.2)	7560 (37.8)	436 (16.3)	31 (10.7)	N/A[c]
Focused primary care provider[d]	426 (1.5)	120 (2.2)	190 (1.0)	97 (3.6)	19 (6.5)	N/A
Miscellaneous	3363 (50.0)	878 (13.1)	1829 (27.2)	549 (8.2)	107 (1.6)	N/A

[a]Visits of 0 to 0.50 correspond to physicians who had a 50% or greater reduction in ambulatory visits between 2019 and 2020.

[b]Visits of 1.25 to <6 correspond to physicians who had a 25% 6-fold increase in visits between 2019 and 2020.

[c]N/A: not applicable; a single test was conducted across all four groups in this section and the *P* value is reported in the row for the first group.

[d]Focused primary care providers are primary care providers who specialize in a specific care (eg, palliative care).

Table 2. Physicians per specialty across each user group.

Specialty	Low use: 0 to 0.5, n (%)	Same use: >0.5 to <1.25, n (%)	High use: 1.25 to <6.0, n (%)	Very high use: ≥6.0, n (%)
Family medicine (n=13,244)	2366 (17.9)	9715 (73.4)	1042 (7.9)	121 (0.9)
Medicine				
Emergency medicine (n=230)	57 (24.8)	126 (54.8)	41 (17.8)	6 (2.6)
Internal medicine (n=999)	233 (23.3)	604 (60.5)	147 (14.7)	15 (1.5)
Infectious diseases (n=148)	27 (18.2)	101 (68.2)	20 (13.5)	0 (0)
Critical care (n=99)	29 (29.3)	53 (53.5)	13 (13.1)	≤5 (≤5.1)
Endocrinology (n=269)	18 (6.7)	215 (79.9)	34 (12.6)	≤5 (≤1.9)
Nuclear medicine (n=48)	15 (31.3)	24 (50.0)	6 (12.5)	≤5 (≤10.4)
Hematology (n=217)	23 (10.6)	164 (75.6)	27 (12.4)	≤5 (≤2.3)
Cardiology (n=684)	77 (11.3)	527 (77.0)	78 (11.4)	≤5 (≤0.7)
Respirology (n=308)	32 (10.4)	237 (76.9)	35 (11.4)	≤5 (≤1.6)
Geriatric medicine (n=152)	29 (19.1)	105 (69.1)	16 (10.5)	≤5 (≤3.3)
Rheumatology (n=213)	19 (8.9)	172 (80.8)	22 (10.3)	0 (0)
Anesthesiology (n=1179)	480 (40.7)	597 (50.6)	101 (8.6)	≤5 (≤0.4)
Nephrology (n=242)	23 (9.5)	198 (81.8)	20 (8.3)	≤5 (≤2.1)
Clinical immunology (n=84)	22 (26.2)	56 (66.7)	6 (7.1)	0 (0)
Gastroenterology (n=345)	39 (11.3)	281 (81.4)	24 (7.0)	≤5 (≤1.4)
Obstetrics and gynecology (n=808)	98 (12.1)	659 (81.6)	45 (5.6)	6 (0.7)
Other specialties				
Psychiatry (n=2061)	328 (15.9)	1326 (64.3)	378 (18.3)	29 (1.4)
Diagnostic radiology (n=624)	179 (28.7)	355 (56.9)	84 (13.5)	6 (1.0)
Medical oncology (n=269)	34 (12.6)	200 (74.3)	35 (13.0)	0 (0)
Pediatrics (n=1487)	395 (26.6)	938 (63.1)	146 (9.8)	8 (0.5)
Radiation oncology (n=211)	14 (6.6)	178 (84.4)	19 (9.0)	0 (0)
Neurology (n=410)	53 (12.9)	321 (78.3)	33 (8.0)	≤5 (≤1.2)
Physical medicine and rehabilitation (n=211)	48 (22.7)	146 (69.2)	16 (7.6)	≤5 (≤2.4)
Dermatology (n=236)	52 (22.0)	173 (73.3)	11 (4.7)	0 (0)
Surgery (n=2868)	536 (18.7)	2212 (77.1)	104 (3.6)	16 (0.6)
Remaining smaller specialties (n=384)	91 (23.7)	246 (64.1)	43 (11.2)	4 (1.0)
Miscellaneous (n=353)	105 (29.7)	69 (19.5)	126 (35.7)	53 (15.0)

Figure 2. Correlation between the 2020:2019 visit ratio and percent virtual care in the four groups of providers: those who reduced (0 to 0.50), maintained (>0.5 to <1.25), increased (1.25 to <6), and significantly increased (≥6) their practice in 2020 relative to 2019.

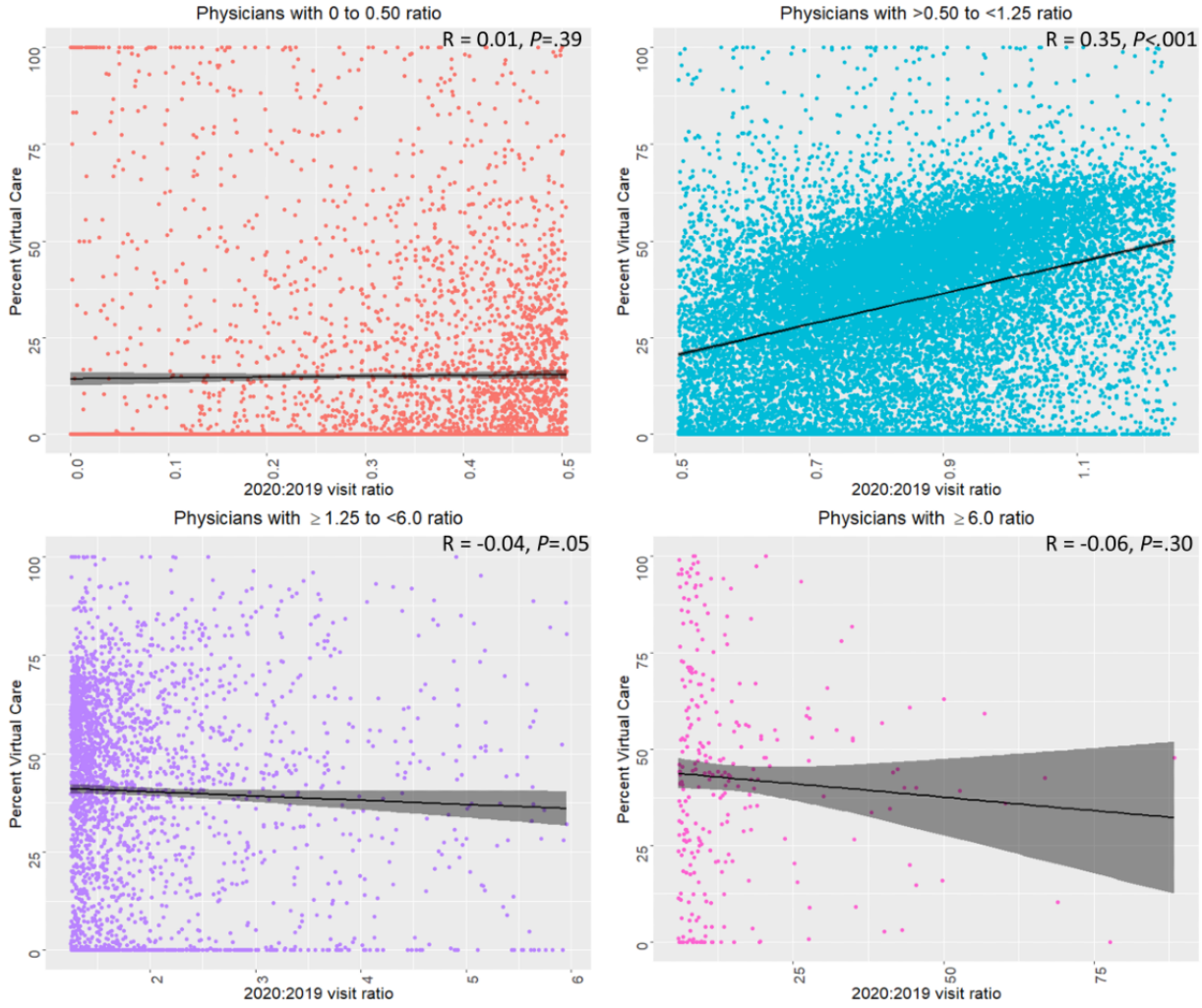
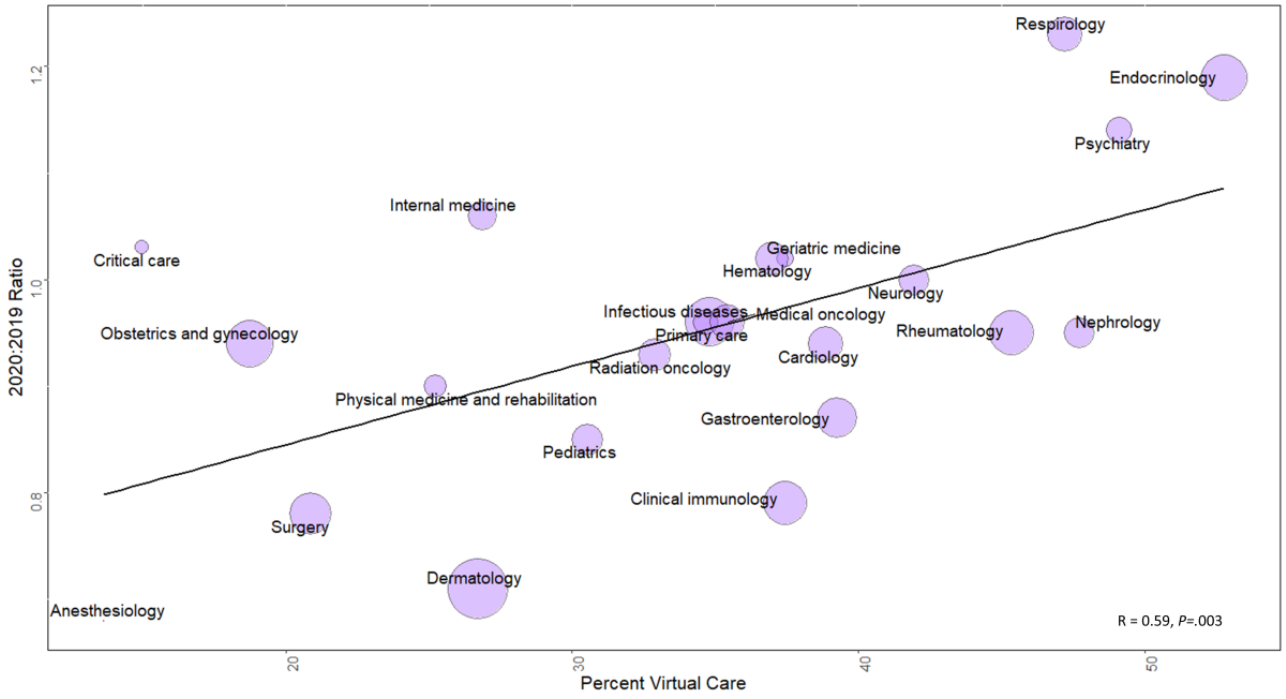


Figure 3. Correlation between the 2020:2019 visit ratio and virtual care adoption across specialties. The size of each sphere indicates the number of visits completed.



Discussion

During the early stages of the pandemic, the introduction of virtual care did not lead to significant increases in visit volume. Only about 10% of physicians increased their visit volumes by 25% or more in 2020 relative to 2019. In total, our results provide reassuring evidence that relaxation of billing requirements early in the COVID-19 pandemic in Ontario were not associated with widespread and aberrant billing behaviors.

Providers who increased their visit volumes tended to be specialists, younger, more recent graduates, and more likely female. Among providers who increased their practice volumes, there was no relationship between the magnitude of increase and virtual care adoption. A significant relationship was observed, however, among providers who maintained their practice. This relationship was also maintained at the specialty level. Endocrinology, respirology, and psychiatry maintained their practices the best and had higher rates of virtual care adoption.

Our results are consistent with data from the United States that showed that despite the introduction of virtual care, overall visit volumes decreased in the early periods of the COVID-19 pandemic [6]. In fact, the introduction of virtual care during the pandemic allowed physicians to maintain their practices. Higher rates of virtual care use among providers who maintained their practice volumes were associated with better maintenance of visit volumes during the pandemic. This trend was also observed in the United States [6]. Here, we confirm these findings with an analysis of the entire physician and patient population in a health care system with a single insurance plan where the introduction of virtual care payment policies occurred at the same time for the entire population.

At least two specialties that showed high virtual care adoption rates and good maintenance of visit volumes during the pandemic were consistent in both Ontario and the United States [6]: psychiatry and endocrinology. Mental health care has the potential to be better suited for virtual care as it often does not require a physical exam and it has been successful in adopting virtual care services both before [13] and after the pandemic [14]. Successful adoption in endocrinology during the pandemic has also been reported [15].

Limitations to our study include a relatively brief time window for evaluating the impact of billing code liberalization, which makes it unclear whether the trends will be maintained in the long term. Our reliance on administrative data also precludes us from robustly evaluating appropriateness of individual visits.

In total, our study suggests that liberalization of virtual care billing requirements coinciding with the COVID-19 pandemic was not associated with an alarming increase in individual physician visit volumes and should serve to assuage concerns over widespread fraud. Furthermore, the strong relationship between the ability to maintain practice volumes and the use of virtual care suggest that the introduction of virtual care allowed continued access to care for patients.

Conflicts of Interest

None declared.

References

1. Baum A, Kaboli PJ, Schwartz MD. Reduced in-person and increased telehealth outpatient visits during the COVID-19 pandemic. Ann Intern Med 2021 Jan;174(1):129-131. [doi: 10.7326/m20-3026]
2. Patel SY, Mehrotra A, Huskamp HA, Uscher-Pines L, Ganguli I, Barnett ML. Trends in outpatient care delivery and telemedicine during the COVID-19 pandemic in the US. JAMA Intern Med 2021 Mar 01;181(3):388-391. [doi: 10.1001/jamainternmed.2020.5928] [Medline: 33196765]
3. Bhatia RS, Chu C, Pang A, Tadrous M, Stamenova V, Cram P. Virtual care use before and during the COVID-19 pandemic: A repeated cross-sectional study. CMAJ Open 2021;9(1):E107-E114 [FREE Full text] [doi: 10.9778/cmajo.20200311] [Medline: 33597307]
4. Mehrotra A, Bhatia RS, Snoswell CL. Paying for telemedicine after the pandemic. JAMA 2021 Feb 02;325(5):431-432. [doi: 10.1001/jama.2020.25706] [Medline: 33528545]
5. Digital transformation: Shaping the future of European healthcare. Deloitte United Kingdom.: Deloitte Centre for Health Solutions; 2020 Sep. URL: https://www2.deloitte.com/uk/en/pages/life-sciences-and-healthcare/articles/european-digital-health.html [accessed 2021-03-01]
6. Patel SY, Mehrotra A, Huskamp HA, Uscher-Pines L, Ganguli I, Barnett ML. Variation in telemedicine use and outpatient care during the COVID-19 pandemic in the United States. Health Aff (Millwood) 2021 Feb;40(2):349-358. [doi: 10.1377/hlthaff.2020.01786] [Medline: 33523745]
7. Alexander GC, Tajanlangit M, Heyward J, Mansour O, Qato DM, Stafford RS. Use and content of primary care office-based vs telemedicine care visits during the COVID-19 pandemic in the US. JAMA Netw Open 2020 Oct 01;3(10):e2021476 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.21476] [Medline: 33006622]
8. OMA's Virtual Care FAQs. Toronto, ON: Ontario Medical Association; 2019 Nov 13. URL: https://ontariorheum.ca/wp-content/uploads/Virtual-Care-OTN-FAQs-Nov-2019.pdf [accessed 2021-08-27]
9. Glauser W. Virtual care has potential to fragment primary care and disturb continuity of care, warn doctors. CMAJ News. 2019 Aug 29. URL: http://cmajnews.com/2019/08/29/virtual-care-has-potential-to-fragment-primary-care-and-disturb-continuity-of-care-warn-doctors/ [accessed 2021-03-04]

10. O'Dowd A. Doctors question Hancock's idea of GP video consultations for all. BMJ 2018 Sep 14;362:k3934. [doi: 10.1136/bmj.k3934] [Medline: 30217932]
11. Verma S. Early impact of CMS expansion of Medicare telehealth during COVID-19. Health Affairs Blog. 2020 Jul 15. URL: https://www.healthaffairs.org/do/10.1377/hblog20200715.454789/full/ [accessed 2021-04-01]
12. Schultz SE, Glazier RH. Identification of physicians providing comprehensive primary care in Ontario: A retrospective analysis using linked administrative data. CMAJ Open 2017 Dec 19;5(4):E856-E863 [FREE Full text] [doi: 10.9778/cmajo.20170083] [Medline: 29259018]
13. Mehrotra A, Huskamp HA, Souza J, Uscher-Pines L, Rose S, Landon BE, et al. Rapid growth in mental health telemedicine use among rural Medicare beneficiaries, wide variation across states. Health Aff (Millwood) 2017 May 01;36(5):909-917. [doi: 10.1377/hlthaff.2016.1461] [Medline: 28461359]
14. Connolly SL, Stolzmann KL, Heyworth L, Weaver KR, Bauer MS, Miller CJ. Rapid increase in telemental health within the Department of Veterans Affairs during the COVID-19 pandemic. Telemed J E Health 2021 Apr;27(4):454-458. [doi: 10.1089/tmj.2020.0233] [Medline: 32926664]
15. Madduri S, Chowdhary R, Sethu Reddy S. Telehealth adoption among endocrinologists during the COVID-19 pandemic. Endocr Pract 2020 Aug;26(8):846-856. [doi: 10.4158/ep-2020-0237]

Edited by C Basch; submitted 05.04.21; peer-reviewed by E van der Velde, P Yellowlees; comments to author 21.04.21; revised version received 30.04.21; accepted 05.05.21; published 07.09.21.

Please cite as:
Stamenova V, Chu C, Pang A, Tadrous M, Bhatia RS, Cram P
Using Administrative Data to Explore Potentially Aberrant Provision of Virtual Care During COVID-19: Retrospective Cohort Study of Ontario Provincial Data
J Med Internet Res 2021;23(9):e29396
URL: https://www.jmir.org/2021/9/e29396
doi:10.2196/29396
PMID:34313590

Viewpoint

Running an Internet Hospital in China: Perspective Based on a Case Study

Lihua Zhi[1*], BM; Pei Yin[2*], MM; Jingjing Ren[2], PhD; Guoqing Wei[3], PhD; Jun Zhou[1], BM; Jun Wu[1], BM; Qun Shen[1], BM

[1]Department of Internet Hospital Office, The First Affiliated Hospital Zhejiang University School of Medicine, Hangzhou, China

[2]Department of General Practice, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

[3]Department of Medical Administration, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

[*]these authors contributed equally

Corresponding Author:
Guoqing Wei, PhD
Department of Medical Administration
The First Affiliated Hospital
Zhejiang University School of Medicine
3rd Fl, No. 17 Bldg, 79 Qingchun Rd
Shangcheng District
Hangzhou
China
Phone: 86 057187231515
Email: weiguoqing2018@zju.edu.cn

Abstract

Internet hospitals, as a new forum for doctors to conduct diagnosis and treatment activities based on the internet, are emerging in China and have become integral to the development of the medical field in conjunction with increasing reforms and policies in China's medical and health system. Here, we take the Internet Hospital of the First Affiliated Hospital, Zhejiang University (FAHZU Internet Hospital) as an example to discuss the operations and functional positioning of developing internet hospital medical services in relation to physical hospitals. This viewpoint considers the platform operation, management, and network security of FAHZU Internet Hospital, and summarizes the advantages and limitations in the operation to provide a reference for other areas with interest in developing internet hospitals.

(J Med Internet Res 2021;23(9):e18307) doi:10.2196/18307

KEYWORDS

internet hospitals; telemedicine; medical service; medical procedures; operation management; network security

Introduction

Internet hospitals are emerging in China as a new way for doctors to perform health care services and education by using telecommunications technology. Internet hospitals are similar to telemedicine, meaning that doctors who are registered in internet hospitals can conduct diagnosis and treatment activities on the internet, and provide subsequent visiting services such as online consultation, prescribing, and dispensing for patients diagnosed with common or chronic diseases in offline hospitals. Family physician services are also provided [1-3]. Internet hospitals are virtual hospitals that are initiated by the government, hospitals, or companies. Internet hospitals initiated by hospitals are equipped with a complete medical team of doctors, nurses, pharmacists, and managers, and patients can also receive examinations at offline hospitals. The scope of medical insurance is the same as that of offline hospitals, and the medical experience is very similar to that of physical hospitals [4].

The Chinese government issued the "Opinion on Actively Pushing Forward the Development of 'Internet+' Action" in July 2015, which clearly encouraged the development of internet hospitals as part of the "Health China" strategy [5]. In 2018, the "Opinion on Promoting the Development of 'Internet+' Healthcare," "Administration Regulations on Internet Diagnoses and Treatments (Trial)," "Administration Regulations on Internet Hospitals (Trial)," and "Administration Specifications for Telemedicine Services (Trial)" were successively issued to propose specific requirements for operating and supervising the diagnosis and treatment activities performed by internet hospitals

[6]. The government, hospitals, and enterprises in China are enthusiastic about building internet hospitals. The first internet hospital was founded at the end of 2014. As of May 2019, 158 internet hospitals were subordinate to various medical institutions in China [7,8].

Telemedicine models of care, in which doctors working in hospitals provide medical care over the internet using approved telehealth platforms, are used by many countries, but they differ with respect to the details of their operation. A selection of representative telehealth platform service providers from different countries is provided in Table 1 for comparison [9-12]. During the COVID-19 pandemic, many countries in addition to China have expanded the scope of diagnosis and treatment, including newly diagnosed patients, and expanded the medical insurance reimbursement limit. However, it is not clear whether this will become a long-term policy [9,13,14].

Table 1. Comparison of operation modes of main telemedicine operators in different countries.

Country	Platform	Operator	Insurance coverage (compared with offline)	Additional royalty	Involved diseases[a]
China	FAHZU[b] Internet Hospital	Hospital	Consistent	Partly[c]	Follow-up after diagnosis of common or chronic diseases
United Kingdom	NHS[d] App	Public body	Consistent	No	Determined by the GP[e] practice
United States	American Well	Company	Inconsistent	Yes	Acute and chronic conditions, consultations regarding prevention and wellness services
Japan	CLINICS	Company	Consistent	Yes	Follow-up after diagnosis of chronic diseases

[a]During the COVID-19 epidemic, all countries except China expanded the scope of diagnosis and treatment, including the initial diagnosis of patients.

[b]First Affiliated Hospital, Zhejiang University.

[c]There is an extra charge for the expert clinic and image-text consultation clinic.

[d]NHS: National Health Service.

[e]GP: general practitioner.

As one of the top hospitals in China, the First Affiliated Hospital of Zhejiang University School of Medicine founded the first large-scale internet hospital in China in February 2016 (FAHZU Internet Hospital), and is representative of this type of hospital in China [15,16]. After 4 years of operation, more than 75,000 patients have received health care services via the internet hospital, leading to a wealth of experience in operating and managing this internet hospital [17,18]. Therefore, this paper aims to describe how an internet hospital in China is run from the perspectives of the service standards, treatment process, charges, platform operation, and management of FAHZU Internet Hospital.

Background

FAHZU Internet Hospital is coordinated by the Internet Hospital Office, which is subordinate to the medical department of the offline hospital with hospital leaders as the primary individuals in charge of the project. The normal operation of the internet hospital is jointly supported by the Information Center, Nursing Department, Quality Management Department, Finance Department, Pharmaceutical Department, Promotion Center, and various inspection departments. In addition, corresponding support in internet technology, logistics, and distribution is offered through cooperation with third-party companies specializing in the internet and medicine. An internet hospital monitoring platform was constructed by the Zhejiang Province Health and Family Planning Commission at the beginning of 2019 to perform online/offline integrated monitoring of internet hospitals, and to ensure the safety and quality of the remote medical service.

FAHZU Internet Hospital provides subsequent visiting services for some patients with common and chronic diseases who receive a diagnosis in offline hospitals. The hospital can also guide patients who have severe illnesses that are not suitable for online consultation to the offline hospital for inspection, diagnosis, and treatment. Both patients and doctors can complete online inquiry diagnoses through their mobile phones or computers. Insurance also offers some types of coverage for internet hospitals.

Health Care Services

Various health services are provided in the internet hospital (Figure 1). Doctors provide patients with online services that include general clinics and expert clinics through face-to-face communication or image-text consultation with patients via remote video systems. With the same clinic hours as the offline hospital, doctors are uniformly assigned to be available at the web-based clinic by the responsible department. Patients register in real time and queue for service. The departments are dominated by common chronic diseases and consultation specialties (Table 2). The expert clinic and image-text consultation clinic are voluntarily administered by senior doctors who have worked independently for more than 3 years and are not limited by their specialties. The training is completed by the Internet Hospital Office. Doctors use their time outside work to provide medical services, and patients can choose the doctor they prefer. It is worth noting that no prescriptions are issued in the image-text consultation clinic since no defined diagnosis is made given the diagnosis limitation, and the charges for the services of the expert clinic and the image-text consultation clinic are priced independently by doctors. Although the clinic

operates outside of routine working hours, which should be paid at the patient's expense, the treatment charges are still covered by insurance.

Figure 1. Various medical services provided in the internet hospital.

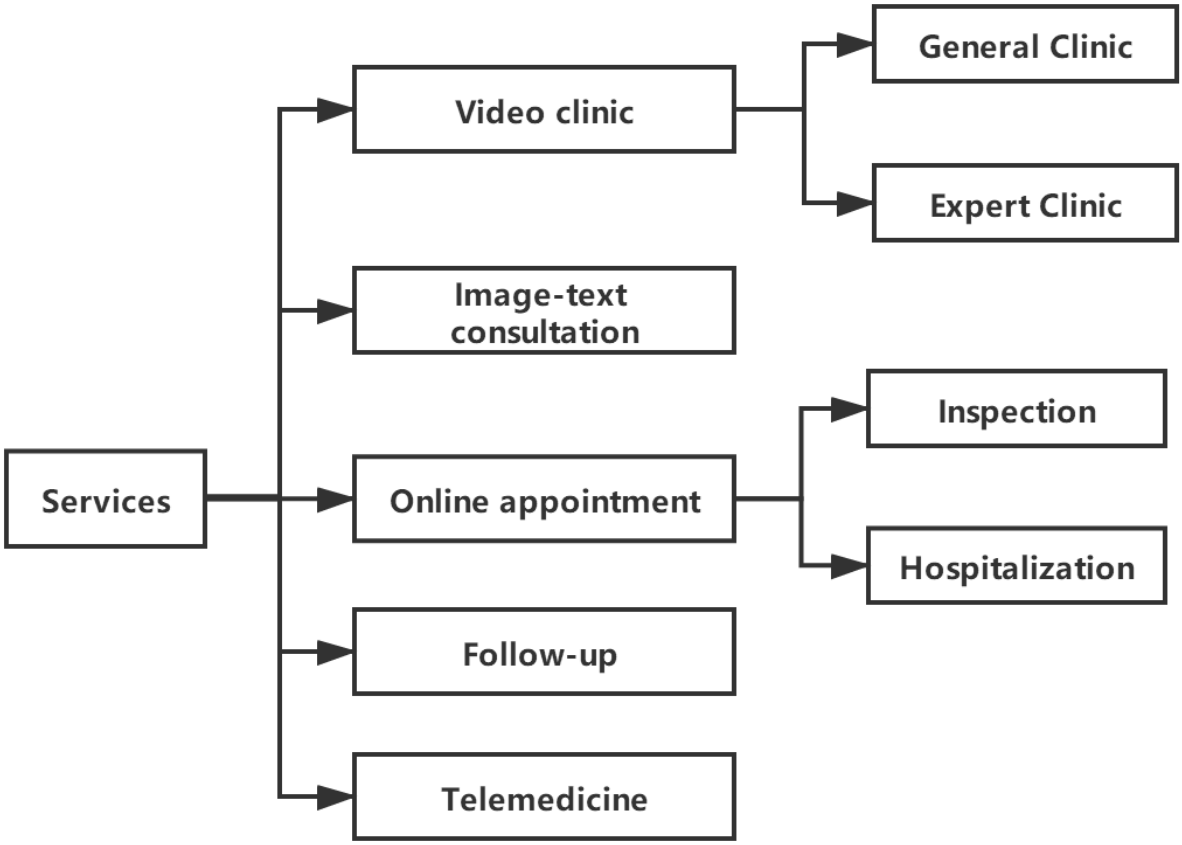


Table 2. General clinic departments associated with the internet hospital.

Departments	Scope of services
Common chronic diseases departments[a]	Provide follow-up management services and health guidance, adjust and prescribe drugs, prescribe examinations, provide services for patients with common and chronic diseases
Health management	Provides health management guidance for healthy people and interpretation of routine physical examination reports
Pharmacy	Provides an online medication guide for patients
Nursing specialties consultation	Provides nursing consultation such as wound and stoma nursing for discharged patients and people with mobility difficulties, as well as nursing guidance for pregnant women, mothers, and babies
Fever clinic[b]	Screens out suspected COVID-19 patients and guides them to the hospital for diagnosis and treatment; patients who are judged to be infected with ordinary viruses are instructed on care at home

[a]Including the Departments of General Practice, Hepatitis Center, Gastroenterology, Cardiology, Endocrinology and Metabolism, Respiratory Medicine, Kidney Disease Center, Neurology, Rheumatology, Pediatrics, Mental Health, and Geriatrics.

[b]Only set up during the COVID-19 pandemic [19].

When a patient uses internet hospital services, they first download the app "Online FAHZU" via a mobile phone or log in to the official website of the First Affiliated Hospital, Zhejiang University via a PC to access the web-based clinic. The patient then selects the consultation department, expert, time, and purpose of consultation; fills in the medical record; and uploads relevant information. After payment via Alipay, patients wait for the doctor's video invitation. After the video diagnosis is complete, patients can also check the diagnosis record and evaluate the service. If there is a drug prescription, the drugs are delivered to the patient's home by express delivery or can be obtained by the patient in the hospital. If a laboratory or imaging examination is needed, patients go to the hospital for the tests within the time determined by the responsible personnel of the internet hospital via a text message. If hospitalization is required, patients can come to the hospital for

the admission procedure at the time determined by the responsible personnel of the internet hospital via a text message (Figure 2).

Doctors can set their available times for the web-based clinic in their spare time upon obtaining permission to conduct web-based clinics. Doctors access the web-based diagnosis clinic in the physician version of the "Online FAHZU" app or log in to the official website of The First Affiliated Hospital, Zhejiang University to access the web-based clinic. After checking the patient's basic information and medical records, the doctor launches a video call with the patient. After the diagnosis, the doctor completes the diagnosis and writes down notes in the medical record. The doctor can also issue online applications for drugs, inspections, and hospitalizations. The diagnosis can be completed after the doctor previews the medical advice. If the prescription is not approved by the pharmacist, it can be revised for approval (Figure 3).

Figure 2. Flowchart of web-based clinic medical services for patients.

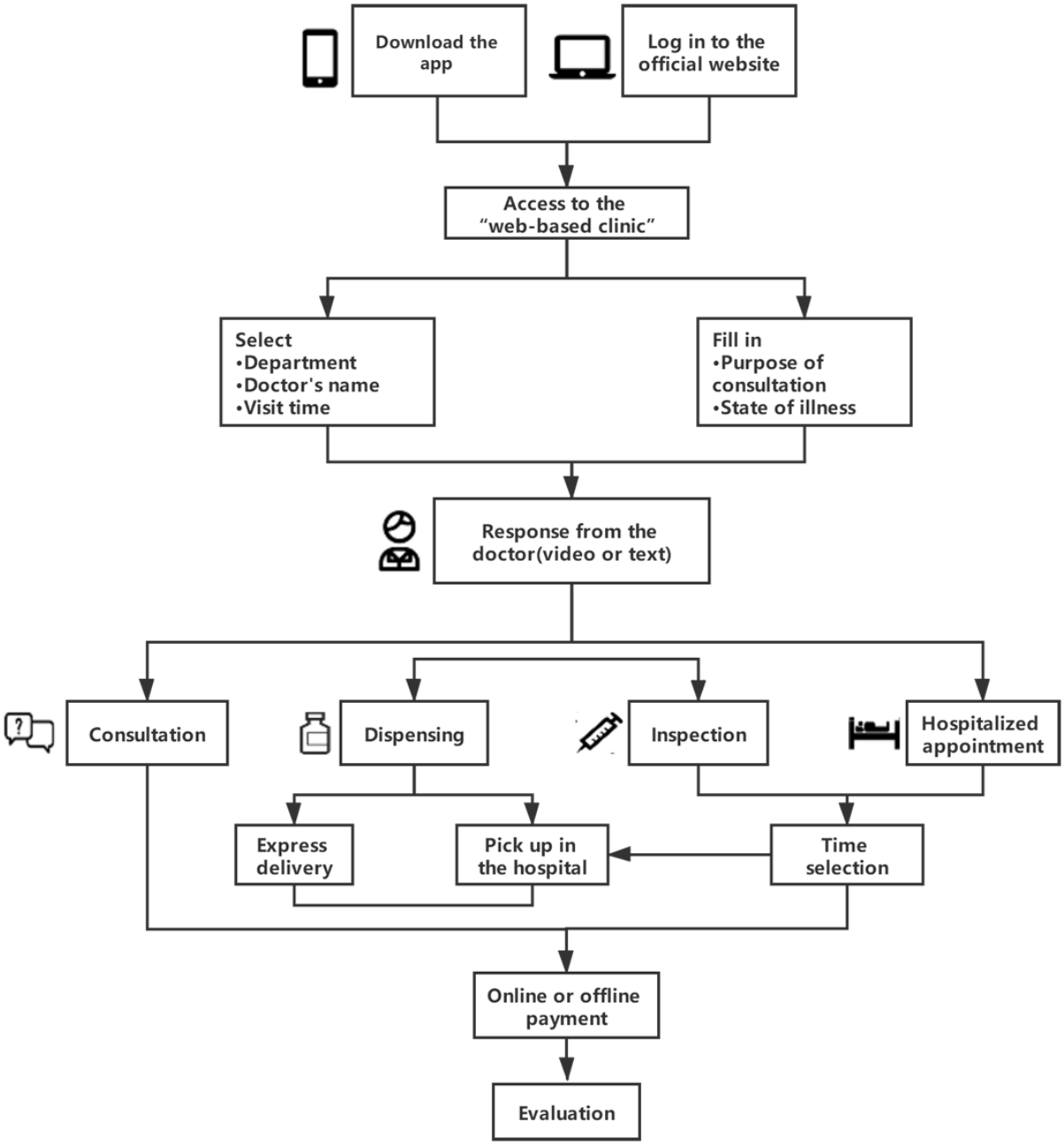
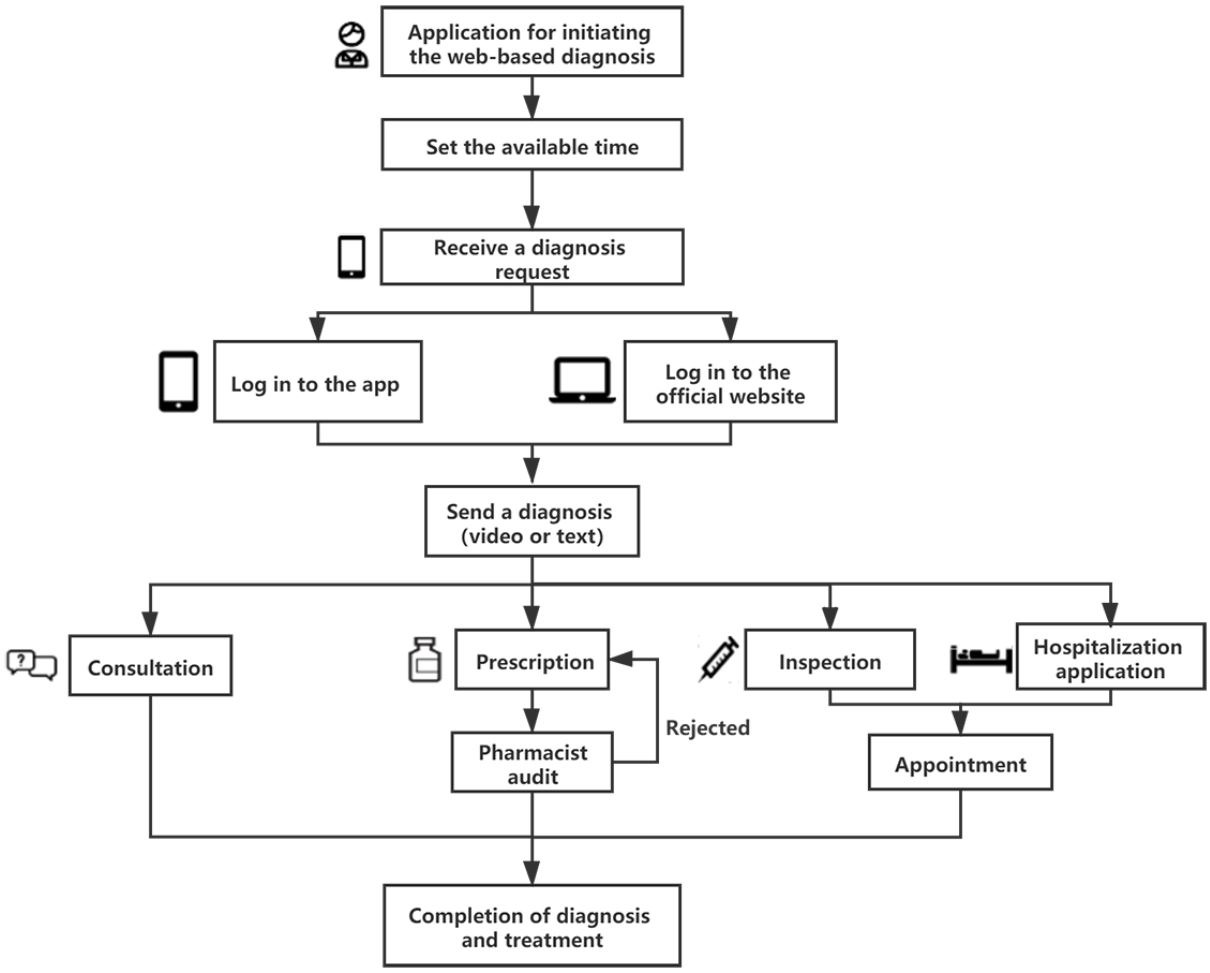


Figure 3. Flowchart of web-based clinic medical services for doctors.



Online Appointments

Unlike other web-based communication platforms that merely provide consultation services, patients of FAHZU Internet Hospital can come to the offline hospital for inspection with an inspection sheet opened by the internet hospital at the inspection time provided via text message by the responsible internet hospital personnel. If the patient is identified for hospitalization while visiting the internet hospital, the physician issues a hospitalization certificate. The patient is admitted to the hospitalization appointment process to await notification from the hospital.

Patient Follow-Up

A discharge follow-up center is subordinate to the internet hospital. The responsible nurse of the ward calls the discharged patient for follow-up inquiry. Any patients who require subsequent visits and inspection appointments are connected to the internet hospital and reexamined by a corresponding physician.

Telemedicine

The internet hospital has remote cooperation with 206 county-level hospitals, 322 community service centers, and 64 pharmacies in China to provide teleconsultation and referral services.

Routine Operation and Management

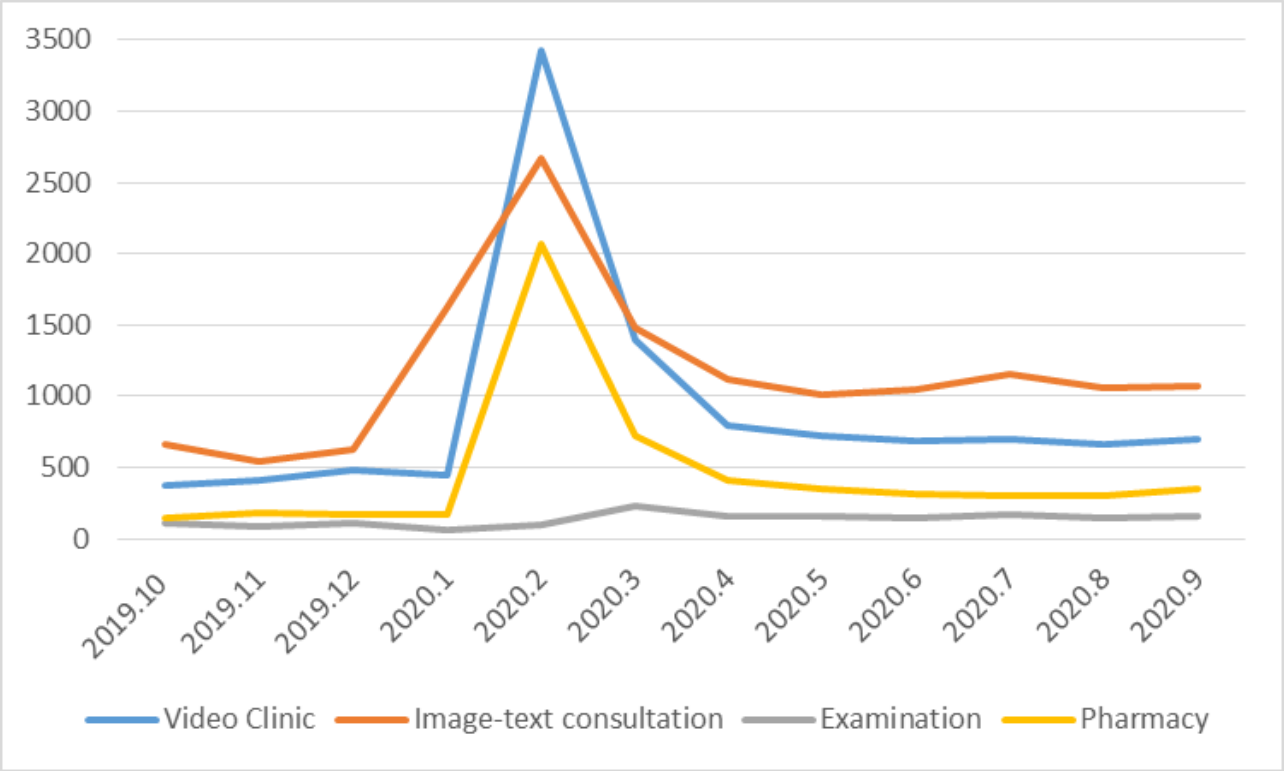
Full-time staff in the Internet Hospital Office are responsible for the management and routine operation of the internet hospital led by the director of the hospital's medical department. After receiving doctors' applications to initiate a web-based clinic, the Internet Hospital Office takes charge of initiating the clinic for candidates after they are trained and qualified in maintaining personal information, authority management, and evaluation performance for doctors engaging in the web-based clinic. In addition to maintaining and ensuring the normal operation of the internet hospital's web-based clinic, the platform status, patients' registered information, diagnosis, drug prescription auditing status, distribution information, timely appointment check, and notification to patients via text message, physician-patient phone consultations are monitored through the backstage operation. The information department, pharmacy department, various inspection departments, medical insurance office, outpatient office, and departments that are closely related to the operation of the internet hospital should be well coordinated with staff cooperation. The hospital staff should also work together with the quality management office of the hospital to formulate related management systems, service processes, emergency plans, and job responsibilities by referring to administration regulations and specifications issued by the government for the internet hospital. All text messages, informed

consent forms, FAQs, and similar files should be prepared in accordance with work requirements. All of these processes should be constantly optimized as per actual situations at work. After inquiring about patients' and doctors' use experience of the internet hospital, related departments should address the items to be improved in a timely manner to enhance the diagnosis and treatment processes by continuously optimizing the system and related functions.

Data Management

Statistics are compiled monthly on online outpatient visits, including basic information of patients, department distribution, visiting purposes, drug prescription and distribution ratio, prescription inspection and the completion ratio of inspections, and patent satisfaction. Based on these data, the internet hospital can be continuously improved. Figure 4 shows the number of patients that received services from the internet hospital from October 2019 to September 2020. From January to April 2020, due to the COVID-19 epidemic, the number of patients using the internet hospital soared. Since then, the epidemic in China has been stable, and most patients have returned to offline treatment; however, the overall number has increased (Figure 4).

Figure 4. Number of patients in internet hospitals in the past year.



Publicity and Promotion

The internet hospital is promoted in a variety of ways. For example, an internet hospital experience center is established in the offline hospital clinic brochures about the internet hospital and doctors' business cards are distributed by staff, plastic bags designed to promote the internet hospital are placed in the pharmacy, and holders for medical records with slogans are put in the outpatient reception area.

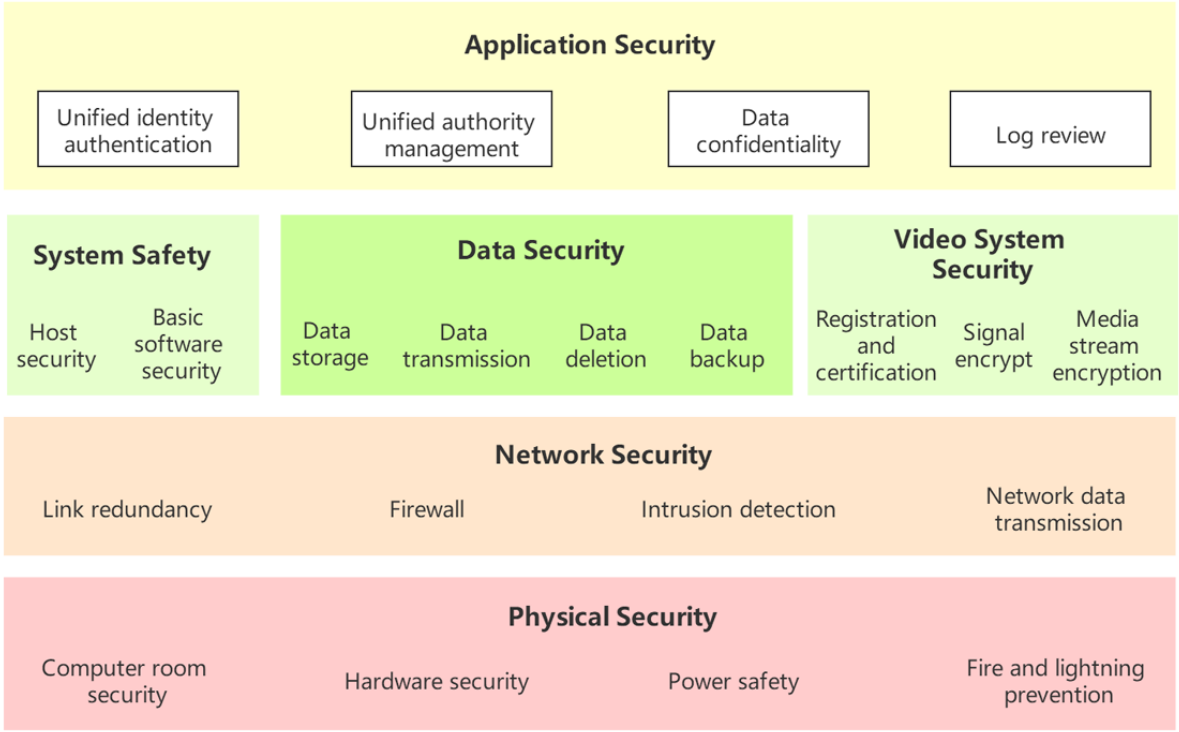
Visit Reception and Experience Sharing

Since its foundation, FAHZU Internet Hospital has shared its working methods and operational experiences with more than 1000 medical institutions worldwide. The FAHZU Internet Hospital team also provided reference opinions for the Administration Regulations on Internet Hospitals issued by the Chinese government in 2018.

Ensuring Network Security

In accordance with the "Guidance on Classified Protection of Information Security in the Health Industry," "Basic Requirements for Hierarchical Protection of Information Security Technology Information Systems," and "Technical Solutions for the Construction of Hospital Information Platforms Based on Electronic Medical Records" of the National Health Commission of China, we have made efforts to develop a technical system that meets the requirements for physical security, network security, host security, app security, and data security. The management system is constructed according to five basic technical requirements: a security management system, security management organization, personnel safety management, system construction management, and system operation and maintenance management (Figure 5).

Figure 5. Safety management and standards.



Overview and Prospects

Since medical services and resources are clustered in large cities, it is difficult for patients in remote areas to access medical resources in large hospitals due to the limitations of economics, distance, and time [20-22]. In addition, it is inconvenient for patients with chronic diseases who live in cities to frequently visit offline hospitals for inspection and drugs (on a regular basis of 15 days or 1 month in China and every 3 months in some cases). Developing internet hospitals might be a good choice to address this dilemma. Patients can obtain quick access to better medical resources through web-based diagnosis and treatment without wasting transportation expenses and time waiting. In a health emergency such as the COVID-19 pandemic, internet hospitals can soothe people's panic, screen out suspected patients, and avoid unnecessary gatherings of patients, which is a good way to supplement offline medical treatment [23]. A summary analysis of the effectiveness of China's internet hospitals found that internet hospitals can ease the difficulty of seeking medical care for patients in remote areas, facilitate patients' communication with doctors in a timely manner, and enable refined tracking of chronic diseases and postoperative patients, thereby improving the effectiveness of patients seeking medical care. At the same time, the convenient medical service experience leads to high patient satisfaction [24,25].

Of course, internet hospitals also have some limitations. First, medical treatment requires practical operations. The network cannot perform physical examinations on patients. Laboratory and imaging examinations also need to be completed offline. Incomplete online medical procedures may affect the accuracy of doctors' judgments. Second, it is difficult for elderly patients who are not experienced with electronic equipment to see a doctor in this way. Third, limited by the hospital's own influence and the influence of the majority of people who prefer the traditional mode, the number of patients in internet hospitals operated by offline hospitals is not growing rapidly.

FAHZU Internet Hospital is independently operated by a large hospital. Doctors, nurses, and management staff, who are the same as those in offline hospitals, can complete tasks that are difficult to complete on third-party platforms, including appointments for examinations and hospitalizations, making the patient treatment process smoother and more complete while controlling patient privacy and data security. However, we also identified some problems in our operations. First, although the patients who visited the clinic provided personal information, including medical insurance information, we were unable to verify whether they actually held this insurance. We should consider adding authentication methods such as face recognition and optical character recognition identification. Second, there is no standard consultation network process, which makes it difficult to guarantee the homogeneity of the work. Third, the relevant laws are not perfect, and there is no legal basis for many issues. Finally, although we train the doctors on how to protect the privacy of patients before they start their work, and the hospital has built an isolated internet hospital outpatient studio for doctors to provide online service during working hours, we cannot guarantee whether a doctor adequately protects the privacy of patients when providing online services outside of work.

Currently, internet hospitals are experiencing rapid growth in China owing to the strong support of the government. Since the First Affiliated Hospital, Zhejiang University has formed a mature operation and management system for FAHZU Internet

Hospital that is subordinate to a large offline hospital with a long operating history, it is expected that our experience can provide a theoretical reference for other areas with interest in developing internet hospitals.

Acknowledgments

We thank the National Natural Science Foundation of China (81870153) for support.

Conflicts of Interest

None declared.

References

1. Opinion on promoting the development of "Internet+ Healthcare". The State Council of the People's Republic of China. 2018 Apr 28. URL: http://www.gov.cn/zhengce/content/2018-04/28/content_5286645.htm [accessed 2020-10-20]
2. Telemedicine and telehealth. HealthIT.gov. URL: https://www.healthit.gov/topic/health-it-health-care-settings/telemedicine-and-telehealth [accessed 2020-10-20]
3. What is telehealth? How is telehealth different from telemedicine? HealthIT.gov. URL: https://www.healthit.gov/faq/what-telehealth-how-telehealth-different-telemedicine [accessed 2020-10-20]
4. Han Y, Lie RK, Guo R. The Internet Hospital as a telehealth model in China: systematic search and content analysis. J Med Internet Res 2020 Jul 29;22(7):e17995 [FREE Full text] [doi: 10.2196/17995] [Medline: 32723721]
5. Opinion on actively pushing forward the development of "Internet+" Action. The State Council of the People's Republic of China. 2015 Jul 04. URL: http://www.gov.cn/zhengce/content/2015-07/04/content_10002.htm [accessed 2020-10-20]
6. Notice on Printing and Distributing Three Documents, including the Administration Regulations on Internet Diagnoses and Treatments (Trial). National Health Commission of the People's Republic of China. 2018 Jul 17. URL: http://www.gov.cn/zhengce/zhengceku/2018-12/31/content_5435436.htm [accessed 2020-10-20]
7. There are 158 internet hospitals in China. National Health Commission of the People's Republic of China. 2019 May 08. URL: http://www.gov.cn/xinwen/2019-05/08/content_5389676.htm [accessed 2020-10-20]
8. Tu J, Wang C, Wu S. The internet hospital: an emerging innovation in China. Lancet Glob Health 2015 Aug;3(8):e445-e446 [FREE Full text] [doi: 10.1016/S2214-109X(15)00042-X] [Medline: 26187488]
9. NHS App privacy policy: online consultation services. National Health Service. URL: https://www.nhs.uk/nhs-app/nhs-app-legal-and-cookies/nhs-app-privacy-policy/online-consultation-services/ [accessed 2020-10-20]
10. How does Virtual Primary Care work? amwell. URL: https://business.amwell.com/service-lines/virtual-primary-care/ [accessed 2020-10-20]
11. Clinics (online medicine). URL: https://clinics.medley.life/ [accessed 2020-10-20]
12. Online healthcare. Ministry of Health Labour and Welfare of Japan. URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/iryou/rinsyo/index_00010.html [accessed 2020-10-20]
13. Telehealth in the time of COVID-19. CCHP. URL: https://www.cchpca.org/covid-19-related-state-actions [accessed 2020-10-20]
14. Online healthcare based on the spread of new coronavirus infection. Ministry of Health Labour and Welfare of Japan. URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/iryou/rinsyo/index_00014.html [accessed 2020-10-20]
15. "FAHZU Internet Hospital", the first online district of a public tertiary hospital in China was officially launched. MedSci. URL: https://www.medsci.cn/article/show_article.do?id=fccb633e5e8 [accessed 2020-10-20]
16. The Internet Hospital of the First Affiliated Hospital, Zhejiang University. URL: http://www.zy91.com/zyyy/hlwyy.html [accessed 2020-10-20]
17. Ma N, Qiu Y, Wei G, Zhou M, Zhou J, Sun C, et al. "Internet-plus" practice focusing on medical service at tertiary general hospitals. Chinese J Hosp Admin 2017;33(2):146-149. [doi: 10.3760/cma.j.issn.1000-6672.2017.02.020]
18. Qiu Y, Liu Y, Ren W, Qiu Y, Ren J. Internet-based and mobile-based general practice: cross-sectional survey. J Med Internet Res 2018 Sep 25;20(9):e266 [FREE Full text] [doi: 10.2196/jmir.8378] [Medline: 30257819]
19. Widely advertising! the First Affiliated Hospital, Zhejiang University opens 24-hour free internet hospital fever consultation clinic!. The First Affiliation Hospital, Zhejiang University. URL: http://www.zy91.com/zxxw/4258.jhtml [accessed 2020-10-20]
20. Arcury TA, Preisser JS, Gesler WM, Powers JM. Access to transportation and health care utilization in a rural region. J Rural Health 2005;21(1):31-38. [doi: 10.1111/j.1748-0361.2005.tb00059.x] [Medline: 15667007]
21. Begashaw B, Tessema F, Gesesew HA. Health care seeking behavior in Southwest Ethiopia. PLoS One 2016;11(9):e0161014 [FREE Full text] [doi: 10.1371/journal.pone.0161014] [Medline: 27626804]
22. Bhan N, Madhira P, Muralidharan A, Kulkarni B, Murthy G, Basu S, et al. Health needs, access to healthcare, and perceptions of ageing in an urbanizing community in India: a qualitative study. BMC Geriatr 2017 Jul 19;17(1):156 [FREE Full text] [doi: 10.1186/s12877-017-0544-y] [Medline: 28724399]

23. Gong K, Xu Z, Cai Z, Chen Y, Wang Z. Internet hospitals help prevent and control the epidemic of COVID-19 in China: multicenter user profiling study. J Med Internet Res 2020 Apr 14;22(4):e18908 [FREE Full text] [doi: 10.2196/18908] [Medline: 32250962]
24. Gu Y, Wei D, Cao X. Analysis of the Reason of the Internet Hospital Occurred in China. Chinese Health Serv Manag 2017;34(6):401-403.
25. Zhu M, Deng C. Systematic review of the advantages of internet hospitals in patients' seeking for medical treatment. Chinese Health Serv Manag 2020;37(6):419-421.

Abbreviations

FAHZU: First Affiliated Hospital of Zhejiang University School of Medicine

Edited by R Kukafka, G Eysenbach; submitted 19.02.20; peer-reviewed by T Aslanidis, S Qanungo, H Fatoum; comments to author 31.08.20; revised version received 25.11.20; accepted 02.08.21; published 16.09.21.

Please cite as:
Zhi L, Yin P, Ren J, Wei G, Zhou J, Wu J, Shen Q
Running an Internet Hospital in China: Perspective Based on a Case Study
J Med Internet Res 2021;23(9):e18307
URL: https://www.jmir.org/2021/9/e18307
doi:10.2196/18307
PMID:34342267

Original Paper

Remote Consultations Versus Standard Face-to-Face Appointments for Liver Transplant Patients in Routine Hospital Care: Feasibility Randomized Controlled Trial of myVideoClinic

Sarah Damery[1], BA, MA, PhD; Janet Jones[1], BA, PhD; Elaine O'Connell Francischetto[1], BSc, MSc; Kate Jolly[1], MBChB, MSc, PhD; Richard Lilford[1], MBChB, PhD; James Ferguson[2], MBChB, MD

[1]Institute of Applied Health Research, Birmingham, United Kingdom

[2]National Institute for Health Research, Birmingham Biomedical Research Centre, University of Birmingham, Birmingham, United Kingdom

Corresponding Author:
Janet Jones, BA, PhD
Institute of Applied Health Research
University of Birmingham
Edgbaston
Birmingham, B15 2TT
United Kingdom
Phone: 44 1214147661
Fax: 44 1214146571
Email: j.e.jones@bham.ac.uk

Abstract

Background: Using technology to reduce the pressure on the National Health Service (NHS) in England and Wales is a key government target, and the NHS Long-Term Plan outlines a strategy for digitally enabled outpatient care to become mainstream by 2024. In 2020, the COVID-19 response saw the widespread introduction of remote consultations for patient follow-up, regardless of individual preferences. Despite this rapid change, there may be enduring barriers to the effective implementation of remote appointments into routine practice once the unique drivers for change during the COVID-19 pandemic no longer apply, to which pre-COVID implementation studies can offer important insights.

Objective: This study aims to evaluate the feasibility of using real-time remote consultations between patients and secondary care physicians for routine patient follow-up at a large hospital in the United Kingdom and to assess whether patient satisfaction differs between intervention and usual care patients.

Methods: Clinically stable liver transplant patients were randomized to real-time remote consultations in which their hospital physician used secure videoconferencing software (intervention) or standard face-to-face appointments (usual care). Participants were asked to complete postappointment questionnaires over 12 months. Data were analyzed on an intention-to-treat basis. The primary outcome was the difference in scores between baseline and study end by patient group for the three domains of patient satisfaction (assessed using the Visit-Specific Satisfaction Instrument). An embedded qualitative process evaluation used interviews to assess patient and staff experiences.

Results: Of the 54 patients who were randomized, 29 (54%) received remote consultations, and 25 (46%) received usual care (recruitment rate: 54/203, 26.6%). The crossover between study arms was high (13/29, 45%). A total of 129 appointments were completed, with 63.6% (82/129) of the questionnaires being returned. Patient satisfaction at 12 months increased in both the intervention (25 points) and usual care (14 points) groups. The within-group analysis showed that the increases were significant for both intervention (P<.001) and usual care (P=.02) patients; however, the between-group difference was not significant after controlling for baseline scores (P=.10). The qualitative process evaluation showed that—according to patients—remote consultations saved time and money, were less burdensome, and caused fewer negative impacts on health. Technical problems with the software were common, and only 17% (5/29) of patients received all appointments over video. Both consultants and patients saw remote consultations as positive and beneficial.

Conclusions: Using technology to conduct routine follow-up appointments remotely may ease some of the resource and infrastructure challenges faced by the UK NHS and free up clinic space for patients who must be seen face-to-face. Our findings regarding the advantages and challenges of using remote consultations for routine follow-ups of liver transplant patients have important implications for service organization and delivery in the postpandemic NHS.

Trial Registration: ISRCTN Registry 14093266; https://www.isrctn.com/ISRCTN14093266
International Registered Report Identifier (IRRID): RR2-10.1186/s13063-018-2953-4

(J Med Internet Res 2021;23(9):e19232) doi:10.2196/19232

KEYWORDS

digital health; remote consultation; patient satisfaction; feasibility; VSQ-9; secondary care; liver transplant patients; mobile phone

Introduction

Background

Increasing the use of technology to reduce pressure on services across the National Health Service (NHS) in England and Wales is a key government target [1]. The NHS Long-Term Plan sets out a strategy for digitally enabled outpatient care to become mainstream across the NHS by 2024 [2]. Central to this strategy is the use of remote consultations, which allows real-time clinician-patient interactions at a distance using video- or telephone-based technology [3,4]. Consequently, there is a growing evidence base assessing the effectiveness of remote consultation in a range of clinical specialties using key metrics such as patient and staff perspectives [4,5], patient acceptability and satisfaction [6,7], health care resource use [8], health outcomes [9], and costs [3].

Despite much of the evidence for the impact of digital technologies on care pathways and service delivery coming from primary care [9-11], studies undertaken in secondary care have found a range of benefits to remote consultations in routine follow-up care. Patients frequently cite advantages relating to personal convenience, such as reduced costs, fewer travel issues, and minimal time off work [3]. For NHS services, there is evidence of a range of incremental benefits, including the efficient use of staff time and physical infrastructure [5]. However, despite the potential benefits of remote consultations in providing follow-up care, uptake in clinical practice remains low [12], and evidence for its impact on health inequalities is equivocal [13]. For many patients, acceptability is mediated by perceived usefulness and security [14], confidence in using the relevant technology [15], and the strength of the patient-clinician relationship [8]. Similarly, although good patient satisfaction can be achieved when remote consultations are used, clinician satisfaction may be reduced [9], particularly when technical issues such as poor audiovisual quality impair patient-clinician communication [3].

Practical implementation issues are often explored superficially in existing studies, and emerging evidence suggests that complexities exist in embedding digital technologies into routine care because of disruption to routines in traditional clinics [16]. Since the emergence of COVID-19, there has been a widespread switch to remote consulting (telephone and on the web) to provide clinical services while mitigating disease risk among vulnerable patients [17-19]. In accelerating the use of remote consultations rapidly, patients and health care professionals have been forced to undergo changes in health care delivery regardless of preference. It is likely that the wholesale shift to remote consultations during COVID-19 has been readily accepted by patients and health care professionals as it was perceived as a temporary measure, with the expectation that follow-up care would revert to prepandemic delivery afterward. A case study of the rapid implementation of virtual consultations for outpatient appointments in orthopedics during COVID-19 showed high levels of satisfaction with appointments but low preferences for remote consultation under *normal* circumstances in which personal safety or easing pressure on services were not the primary drivers of uptake [20]. This suggests that there may be substantial value in assessing the feasibility of implementing remote consultations for follow-up appointments outside of the COVID-19 situation, and an awareness of the enduring perceived advantages and pitfalls of remote appointments may be important when planning how to embed remote consultation within routine practice. This study reports our experiences with the implementation of an outpatient video consultation system before COVID-19.

Aim and Objectives

The aim of this study is to evaluate the feasibility of using real-time remote consultations between clinically stable liver transplant patients and their hospital physicians using secure videoconferencing technology for routine follow-up appointments and assess whether patient satisfaction differed for intervention patients and those receiving usual care (face-to-face consultations). The study objectives are to (1) assess the rates of recruitment, retention, and crossover between arms; (2) assess the appointment numbers and the technical performance of the remote consultation software; (3) assess patients' ability to complete clinical testing locally; (4) explore patient satisfaction across key domains of the RAND Visit-Specific Questionnaire-9 (VSQ-9) instrument [21]; (5) monitor questionnaire return rates and data completeness; and (6) assess the feasibility of collecting patient-reported data on health service use, health-related quality of life (HRQoL), and costs. An embedded qualitative process evaluation explored patient and staff experiences.

Methods

Design and Setting

A two-armed, parallel group, statistician-blinded feasibility randomized controlled trial (RCT) of the provision of real-time remote appointments via videoconferencing software compared with standard face-to-face consultations (usual care) for delivering routine follow-up to clinically stable liver transplant patients was set up. Participants were recruited from 4 outpatient liver clinics at the Queen Elizabeth Hospital Birmingham (QEHB): primary sclerosing cholangitis, primary biliary cholangitis, alcohol-related liver disease, and autoimmune hepatitis. The trial was registered with the International Standard RCT Number Registry (trial number: 14093266).

The published study protocol [22] outlined a definitive trial to evaluate the effectiveness of remote consultations, with a recruitment target of 180 patients (90 in each arm) that would provide sufficient power to detect a statistically (and clinically) significant difference among groups in the primary outcome measure. However, once underway, poor recruitment meant that a definitive trial would not be possible, and ethical approval was obtained to formally modify the study design to a feasibility RCT. Rather than aiming to provide a definitive assessment of intervention effectiveness, the feasibility RCT reported here placed emphasis on evaluating recruitment and retention, crossover, the feasibility of administering the intervention and collecting outcome data, and an embedded process evaluation. This paper follows the CONSORT (Consolidated Standards of Reporting Trials) extension for randomized pilot or feasibility trials [23], the CONSORT eHealth checklist v1.6 (Multimedia Appendix 1) [24], and the COREQ (Consolidated Criteria for Reporting Qualitative Studies; Multimedia Appendix 2) [25].

Participants

Clinically stable adult patients (aged ≥18 years) who received a liver transplant 1 to 5 years before baseline were eligible for the study if they (1) could access the myhealth@QEHB patient portal [26]; (2) could arrange for local clinical testing (blood tests, weight, and blood pressure) via their general practitioner (GP) or a dialysis center; (3) had an internet-enabled computer with a camera, running an operating system compatible with the remote consultation software; (4) had follow-up appointments every 3 or 6 months; and (5) could give informed consent. Patients who were unable to speak or read English or those who were involved in another research study involving ongoing questionnaire completion were excluded. Before each clinic, hospital staff screened the day's appointment list to identify potentially eligible patients according to the time elapsed since transplant. Clinical stability was assessed by the consultant using their judgment of the patient's liver function, adherence to immunosuppressant medication, and blood test results. Clinically eligible patients were introduced to the study during their appointment by their consultant, who assessed their eligibility further against the other inclusion criteria. Patients interested in participating gave written consent to a member of the research team after their appointment and completed a baseline questionnaire before randomization. Patients who chose not to participate had their reason or reasons recorded.

Participants were randomized in equal numbers to the intervention (myVideoClinic) or usual care (standard face-to-face consultations) arm of the study using the GraphPad web-based randomization tool [27]. Intervention patients were registered on the myVideoClinic system and given instructions to access web-based software training. All patients were registered on the myhealth@QEHB system [26] for the administration of follow-up questionnaires. myhealth@QEHB is a patient records portal developed by the Trust Informatics department, which is currently used by approximately 10,000 patients across 40 clinical specialties. It allows patients remote access to clinical information, including letters, laboratory results, and referrals. Patients can also view appointments, upload or share files on the system, and interact with other patients to create peer-support networks.

Intervention

Intervention patients received appointment details through post and standard text appointment reminders. Patients were required to undergo clinical tests locally and make results available through myhealth@QEHB [26] before their appointment. For their appointment, patients logged in to myhealth@QEHB to speak to their consultant using an embedded secure videoconferencing platform provided by Vidyo. The Vidyo platform has been used for remote consultations by many health care providers and is underpinned by a comprehensive information governance policy that protects the confidentiality and security of end users according to the national and international governance and data protection standards [28]. Patients could submit 3 questions before their appointment, and consultation audio recordings were available afterward through the myhealth portal. When technical issues occurred during a consultation, the consultant telephoned the patient to finish the appointment and scheduled a face-to-face consultation if necessary.

Patients allocated to the intervention were made aware that it was the option of having a remote consultation and that they could ask for a standard appointment if they wished while still having subsequent appointments remotely. Clinical staff could assign intervention patients to standard appointments at the patient's request, if there was a clinical need, or if the consultant had not seen preappointment tests for two successive consultations. The statistician, analyzing the primary outcome data, was blinded to participants' group allocation.

Usual Care

Patients received standard face-to-face care at the hospital, standard letters notifying them of their appointments, and routine text reminders. They completed clinical tests at the hospital on the day of their appointment, which the consultant reviewed afterward.

Qualitative Evaluation

An embedded qualitative process evaluation used semistructured interviews to explore participants' experiences of remote consultations (patients randomized to the intervention and staff administering the intervention). Interviews with staff and patients were conducted using topic guides (Multimedia Appendix 3). These were piloted with members of the study steering group, including the study's patient representative. Following their 12-month appointment, patients in the intervention arm were contacted by telephone by the research team to ask whether they were willing to take part in an interview. Patients were purposively sampled to ensure diversity in age and sex. At the end of the study, purposively sampled staff (hospital consultants, staff booking appointments, and information technology [IT] support personnel) were invited to take part in an interview. All participants provided written informed consent. Interviews were conducted by 2 experienced female research fellows (JJ, who was qualified at the PhD level, and EOF, who was qualified at the master's level). Neither interviewer knew the patient participants before the study. Field notes were made by the researcher after each interview. Interviews were audio-recorded, transcribed verbatim, and

checked against the recordings for accuracy. Participants were interviewed only once, and they did not have the opportunity to review their transcripts or provide feedback on the findings.

Outcome Measures

The outcome measures and data collection are summarized in Table 1. The primary outcome was the combined satisfaction score for the three domains of VSQ-9 (convenience of location, getting through to the office by phone, and length of waiting time). Participants rated their satisfaction on a 5-point scale (poor, fair, good, very good, and excellent), which was transformed into a 0 to 100 linear scale, with higher scores denoting greater satisfaction. A 10-point difference between the groups at 12 months was considered clinically significant [29]. Secondary outcomes included recruitment, retention and crossover rates, questionnaire completion rates and return format, system performance, health service use, feasibility of obtaining clinical tests locally, clinical contacts, satisfaction in the other six VSQ-9 domains, and the feasibility of collecting data on patient costs and HRQoL (using EQ-5D-5L) [30].

Table 1. Outcome measures and data collection.

Outcome measure	Data collection instrument	Schedule or format[a]
Primary outcome		
(Change in) satisfaction in the VSQ-9[b] domains of *convenience of location*, *getting through to the office by phone*, and *length of time waiting* [21]	VSQ-9 patient questionnaire	Baseline and 3, 6, 9, and 12 months
Secondary outcomes		
Participant recruitment, retention, and crossover between arms	Routinely collected data	Throughout
Clinical contacts and nonattendance	Number of appointments	Case report form (consultant completed)
myVideoClinic system performance	Failed appointments and telephone consultations	Routinely collected metrics
Patient completion of clinical tests locally	Blood tests, blood pressure, and weight	Case report form (consultant completed)
Patient satisfaction in the other six VSQ-9 domains	VSQ-9 and patient questionnaire	Baseline and 3, 6, 9, and 12 months
Questionnaire completion rates	Number of questionnaires completed and mode of return	Baseline and 3, 6, 9, and 12 months
Feasibility of collecting patient-reported health service use data	Patient questionnaire	Baseline and 3, 6, 9, and 12 months
Feasibility of collecting patient-reported HRQoL[c] data	EQ-5D-5L [20] and patient questionnaire	Baseline and 3, 6, 9, and 12 months
Feasibility of collecting patient-reported cost data (appointment and clinical testing)	Patient questionnaire	Baseline and 3, 6, 9, and 12 months
Patient and staff experiences of virtual clinics	Semistructured interviews	Semistructured interviews

[a]Data at the 3- and 9-month follow-ups were collected from patients on a 3-month follow-up schedule only; data at 6 and 12 months were collected from all patients.

[b]VSQ-9: Visit-Specific Questionnaire-9.

[c]HRQoL: health-related quality of life.

Patients stayed in the study for 12 months. All patients completed baseline questionnaires recording sociodemographics (postcode, sex, age, ethnicity, and employment status), time elapsed since transplant, VSQ-9 scores, HRQoL (using EQ-5D-5L), health care use in the previous 3 months, and costs (travel or personal expenses) associated with their baseline appointment. Patients on a 3-month follow-up schedule received questionnaires via myhealth@QEHB [26] up to 7 days after their 3-, 6-, 9-, and 12-month appointments. Patients seen every 6 months received questionnaires at 6 and 12 months. The 6- and 12-month questionnaires collected data on VSQ-9, HRQoL, costs, IT issues, and health service use. The 3- and 9-month questionnaires covered VSQ-9 and costs only. A short questionnaire was also sent to eligible patients who chose not to participate, to understand their decision.

The qualitative interviews covered patients' experiences of the study (both arms) and recommendations for improvement (intervention arm). Staff interviews focused on experiences of remote consultations and perceived advantages and disadvantages.

Data Analysis

An intention-to-treat analysis was performed using SPSS version 25 (IBM Corporation) [31]. Participant characteristics at baseline were summarized descriptively and compared between study arms using two-tailed *t* tests or chi-square tests, as appropriate. Analysis of the primary outcome was undertaken using analysis of covariance tests to compare intervention and usual care group satisfaction scores at the study end while controlling for baseline scores. For patients with missing 12-month VSQ-9 data, their most recently available data were used. Secondary outcomes

were analyzed descriptively, and feasibility outcomes were presented overall and by group with counts and percentages. Subgroup analyses were not undertaken because of small participant numbers.

For the qualitative process evaluation, interviews were audio-recorded and transcribed verbatim. Transcripts were uploaded to NVivo 12 Plus (QSR International) [32], and with no a priori expectations of findings, the interviews were analyzed thematically [33]. The analysis followed the Braun and Clarke [33] recommended six-stage process of analysis: (1) the researcher spending time familiarizing themselves with the data; (2) the researcher generating initial codes for the data; (3) the researcher starting to develop themes from the codes; (4) the researcher reviewing the themes and codes; (5) the researcher defining, refining, and naming the themes; and (6) the researcher producing an analytic narrative of the findings. Data coding was undertaken by 1 researcher, with 10% of the transcripts independently coded by a second researcher. An initial coding framework was developed using the first 10% of transcripts. If data did not fit the codes in this framework, they were discussed within the team, and where appropriate, new codes were generated, or amendments were made until all data were analyzed. Findings from both patient and staff data sets were compared to identify any similarities or differences.

Sample Size

The protocol [22] outlined a definitive trial in which a required sample size of 90 patients in each arm (180 total) would be sufficient to detect a 10-point difference in VSQ-9 scores between groups at 80% power and α=.05. This was based on the estimated annual figures of 267 clinically stable liver transplant patients attending routine follow-up in the clinics of interest 1 to 5 years posttransplant, an estimated 60% participation rate, and 30% attrition [34]. However, the actual number of eligible patients, as well as the recruitment rate, were substantially lower than expected, and the decision was made to formally change the study design to a feasibility RCT in which as many participants as possible would be recruited over 4 months without a formal sample size target [35].

Ethical Approval

This study received a favorable ethical opinion from the West Midlands Solihull Research Ethics Committee on October 24, 2017 (reference: 17/WM/0338). Research governance approval was obtained from the University Hospitals Birmingham NHS Foundation Trust in February 2018 (reference: RRK6080). The study was sponsored by the University of Birmingham.

Patient and Public Involvement

The patient and public involvement group of the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care West Midlands (NIHR CLAHRC WM) provided advice on the study design, data collection tools, and outcome selection. A patient representative sat on the study steering committee.

Results

Recruitment, Retention, and Crossover

Recruitment took place between March 12, 2018, and July 19, 2018. Overall, 203 patients were potentially eligible according to age and time elapsed since transplant (Multimedia Appendix 4). After further screening, 63.1% (128/203) patients were considered eligible. Of the 128 eligible patients, 72 (56.3%) were not recruited because of the following reasons: 57% (41/72) were not seen by the research team after their appointment, and 43% (31/72) declined participation. Of the 31 decliners, 15 (48%) decliners subsequently returned a questionnaire explaining their decision: 47% (7/15) liked attending hospital or lived nearby, 27% (4/15) lacked appropriate computer equipment or access to myhealth@QEHB [25], 20% (3/15) felt that it would be too difficult to obtain test results or medication locally, and 7% (1/15) disliked the idea of remote consultations.

Of the 203 patients, 56 (27.6%) were recruited. Of the 56 recruited patients, 29 (52%) were allocated to the intervention, and 27 (48%) were allocated to usual care. Approximately 7% (2/27) of usual care patients withdrew after randomization, giving an adjusted recruitment rate of 26.6% (54/203). All remaining patients continued in the study until it closed. Crossover was substantial, with 45% (13/29) of patients changing to usual care: 69% (9/13) because of patient request; 15% (2/13) for health reasons, and 15% (2/13) were changed by the consultant after missed remote appointment or appointments. No harm or adverse events were reported during the study.

Participant Characteristics

Mean participant age was 48.9 (SD 13.8) years, and 59% (32/54) of patients were men (Table 2). Approximately 37% (20/54) of patients had follow-up appointments every 3 months. Patients in the intervention arm were slightly younger than those receiving usual care, and a greater proportion was men, although none of the baseline characteristics showed a statistically significant difference between groups when means or proportions were assessed. The mean distance between the hospital and participants' home postcode was 75.3 (range 7-209) miles for all patients—79.6 (range 7-209) miles for intervention patients and 70.3 (range 8-145) miles for usual care patients.

Table 2. Baseline characteristics by the study arm (N=54).

Characteristics	Intervention (n=29)	Usual care (n=25)	Overall	Comparison		
				t test[a] (*df*)	Chi-square (*df*)	*P* value
Age (years)				1.78 (52)	N/A[b]	.08
Value, mean (SD)	45.8 (14.1)	52.4 (12.9)	48.9 (13.8)			
Value, median (IQR)	35.0 (35-57.5)	53.0 (43-62.5)	50.0 (35-62.0)			
Sex, n (%)				N/A	.03 (1)	.86
Male	18 (62)	14 (56)	32 (59)			
Female	11 (38)	11 (44)	22 (41)			
Follow-up, n (%)				N/A	.02 (1)	.89
3 months	10 (35)	10 (40)	20 (37)			
6 months	19 (65)	15 (60)	34 (63)			
Distance from hospital (miles)				0.72 (52)	N/A	.47
Value, mean (range)	79.6 (7-209)	70.3 (8-145)	75.3 (7-209)			
Employment, n (%)				N/A	1.05 (3)	.79
Full time	13 (45)	11 (44)	24 (44)			
Part time	4 (14)	5 (20)	9 (17)			
Retired	9 (31)	8 (32)	17 (32)			
Unemployed	3 (10)	1 (4)	4 (7)			
Qualifications, n (%)				N/A	2.06 (3)	.56
School level	19 (66)	16 (64)	35 (65)			
Degree level	6 (21)	5 (20)	11 (20)			
Postgraduate	3 (10)	1 (4)	4 (7)			
None	1 (3)	3 (12)	4 (7)			

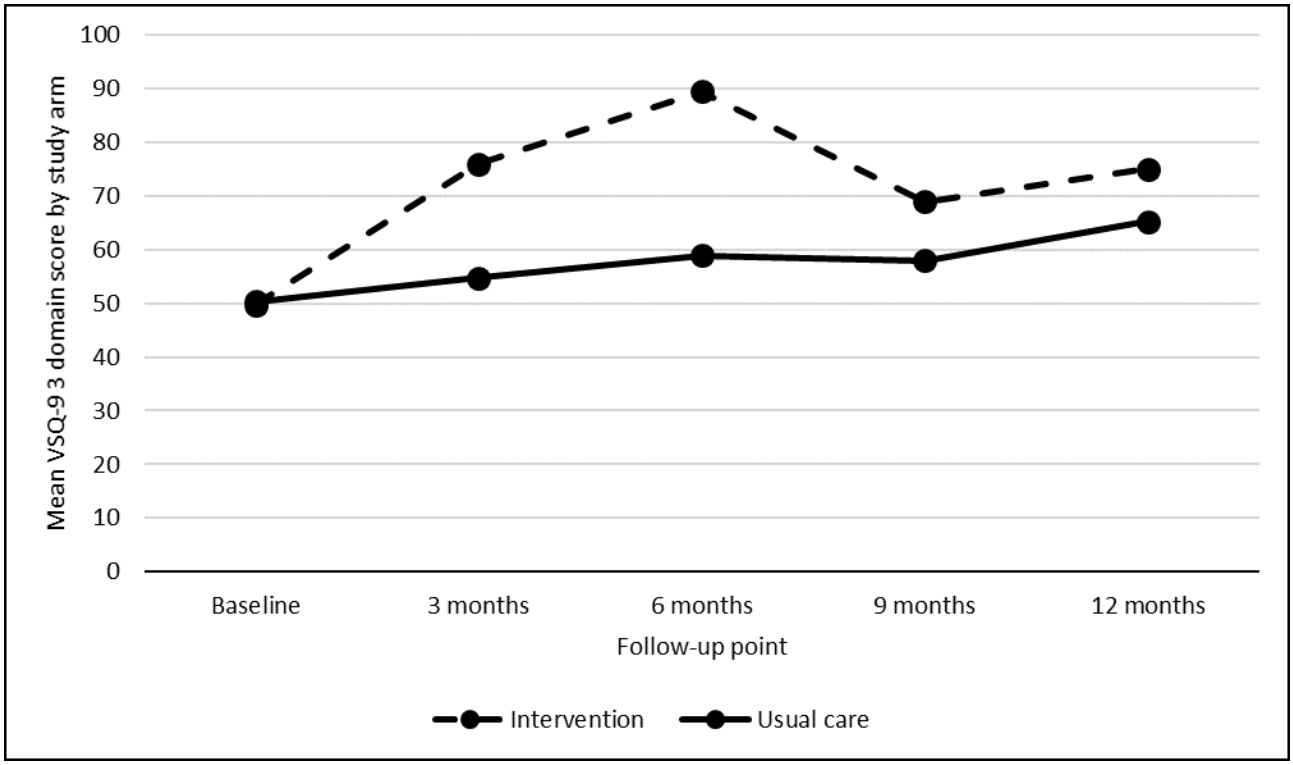
[a]Two-tailed *t* test.

[b]N/A: not applicable.

Primary Outcome

The mean VSQ-9 baseline score for the three domains of *convenience of location*, *getting through by phone*, and *length of time waiting* was 49.7 (SD 17.9) for patients in the intervention arm and 50.3 (SD 19.5) for usual care patients. At each subsequent time point, scores were substantially higher for intervention patients than for usual care patients (Figure 1). At 12 months, scores had increased from baseline in both groups: by 24.7 points for intervention patients (74.4, SD 25.3) and by 14.2 points for usual care (64.5, SD 28.4). Within-group analysis showed that the increase was significant for both intervention ($P<.001$) and usual care ($P=.02$) patients. Analysis of covariance showed no significant difference in satisfaction between groups after controlling for baseline ($F_1=2.84$; $P=.10$). However, the study was underpowered given its feasibility design.

Figure 1. Mean Visit-Specific Questionnaire-9 scores in the three domains comprising the primary outcomes stratified by group. VSQ-9: Visit-Specific Questionnaire-9.



Secondary Outcomes

Clinical Contacts and System Performance

A total of 129 appointments took place, of which 42 (32.5%) were intended as remote consultations across 19 patients (10 patients had switched to usual care before their first remote appointment). Of the 42 planned remote consultations, 18 (43%) took place over video, and 24 (57%) took place at least partially over the telephone because of technical issues. Only 26% (5/19) of patients received all appointments over video. Approximately 47% (9/19) of patients received a mix of video and telephonic appointments, and the remaining 26% (5/19) received only telephonic appointments. Technical issues reported by questionnaire respondents related to problems with software or browser compatibility, audiovisual issues, system freezes or crashes, log-in problems, and system time-outs.

Local Clinical Testing (Intervention Patients)

Blood pressure results were available to the consultant on 81% (34/42) of occasions and weight results on 86% (36/42) of occasions. Blood test results were more variable, with 62% (26/42) of consultations having full blood results, 17% (7/42) having partial results, and 21% (9/42) being without results. The most common reasons for the lack of blood test results were that tests had been done but results not passed on; tests were done, but key analyses were omitted (usually urea, electrolytes, and tacrolimus); the patient did not undergo the tests; or because GPs needed clarification about what was required. These issues decreased over time, with only 1 patient lacking preappointment test results by their 12-month appointment.

VSQ-9 Other Domains

The mean VSQ-9 baseline score for the six domains not assessed within the primary outcome (*wait for appointment*, *time with consultant*, *explanation of what was done*, *technical skill of clinician*, *personal manner*, and *overall impression*) was 73.7 (SD 14.0) for intervention patients and 66.9 (SD 14.8) for usual care patients. The scores remained higher for intervention patients at each follow-up point. At the end of the study, satisfaction scores were 80.9 (SD 15.6) for patients receiving the intervention and 72.7 (SD 19.6) for patients receiving usual care.

Questionnaire Return and Data Completeness

Approximately 63.6% (82/129) of postappointment questionnaires were returned. A greater proportion of questionnaires was returned by intervention patients: 35.6% (46/129) compared with usual care patients 27.9% (36/129). Approximately 20% (11/54) of patients returned no questionnaires, 30% (16/54) returned one or more questionnaires, and 50% (27/54) returned all questionnaires. Of the 82 questionnaires returned, only 22 (27%) were sent back electronically (11 in each arm); the remainder were submitted on paper. Data on health service use were provided by all intervention and usual care patients, and all questionnaires that were returned at baseline, 6 months, and 12 months from patients in both study arms contained full HRQoL data. Data on patient costs were poorly completed, with fewer than half of all returned questionnaires containing data that could inform an economic evaluation.

Patient and Staff Experience

Of the 12 interviews conducted for the qualitative process evaluation, 8 (67%) were with patients receiving the intervention

and 4 (33%) with staff (2 consultants, 1 member of IT support, and 1 hospital administrator; Table 3). All patient participants preferred their interviews to be conducted over the telephone. All staff interviews were conducted face-to-face on hospital premises. No one else was present at the interviews apart from the participants and researchers. The interview length ranged from 15 to 45 minutes. Six overarching themes were interpreted from the data: (1) the recruitment process; (2) using the myVideoClinic software (including barriers and facilitators); (3) perceptions of remote consultations (including satisfaction with care and views on remote consultations compared with face-to-face appointments); (4) local clinical testing (barriers and facilitators to obtaining tests and ensuring results were made available); (5) perceived benefits of remote consultations to the hospital; and (6) implications for implementation. The findings for each theme are described below, with key verbatim quotations cited in the text. Quotations are attributed using participant number, sex, and age group for patients and using participant number and role for staff. Multimedia Appendix 5 summarizes the key themes (including definitions) and provides supplementary supporting quotations for each theme.

Table 3. Characteristics of interview participants (N=12).

Characteristics	Participants, n (%)
Patients (n=8)	
Sex	
Male	4 (50)
Female	4 (50)
Age (years)	
30-39	1 (12)
40-49	2 (25)
50-59	2 (25)
60-69	2 (25)
70-79	1 (12)
Staff (n=4)	
Sex (n=4)	
Male	1 (25)
Female	3 (75)
Information technology (n=1)	
Male	0 (0)
Female	1 (100)
Administration (n=1)	
Male	0 (0)
Female	1 (100)
Consultant (n=2)	
Male	1 (50)
Female	1 (50)

Interview Themes

The Recruitment Process

Both consultants who were interviewed agreed that appointments in which the study was explained to patients were slightly longer than usual. However, recruitment was not onerous for clinical and administrative staff, and patients were generally keen to participate:

> *So broadly, it was a straight forward study in some respects to recruit for, because it's not a particularly evasive intervention for patients. But of course recruiting people in a busy clinical setting is always hard remembering and trying to make sure that you pick the right people.* [Participant 02, consultant]

However, both consultants also reported that time constraints meant that not all clinic lists could be prescreened to identify potentially eligible patients before appointments began, which affected recruitment rates.

Using myVideoClinic Software

Audio and visual problems with at least one remote consultation were reported by 75% (6/8) of patients, and it was rare that sound and video worked together, even when preappointment device testing showed no incompatibility between patients' home system and the Trust. Connections were often slow, and

1 participant was particularly concerned about the confidentiality of their appointment:

> *If you've got something very private or something you're really worried about you'd not be very comfortable if the consultant is going "I can't hear you," "I can't hear you."* [Participant 08, female, aged 60-69 years]

Both consultants also experienced problems, and many planned video appointments became telephone consultations. IT support was not always timely because of pressure on clinic time:

> *It was always after the appointment had failed then you'd contact the help desk and they kind of try and work out what had happened at the time which is never quite so good because they can't see what the problem is.* [Participant 01, consultant]

All the patients (8/8, 100%) welcomed the facility to upload questions before their appointment, although none used it, nor did any patients report accessing the audio recording of their appointment.

Perceptions of Remote Consultations

All patients (8/8, 100%) reported satisfaction with the received care; however, 63% (5/8) explicitly noted that this might not be satisfactory if they were unwell. For example, 1 patient was worried about being unable to demonstrate physical symptoms to her consultant:

> *The disadvantage is that I can't physically show anything on my body, so at one stage I did have these rashes, but doctor couldn't physically examine me.* [Participant 08, female, aged 60-69 years]

Of the 8 patients, 6 (75%) thought that their patient-physician relationship remained the same and 2 (25%) reported some awkwardness with remote consultations:

> *It makes you feel a little bit more distant and I would say that I always do feel more reassured when I see them in person.* [Participant 12, male, aged 50-59 years]

Consultants reported that appointments tended to be more business-like than face-to-face interactions, with less informal discussions. All patient participants (8/8, 100%) reported saving time and money:

> *If I use the car, then it's an 80 mile, 2 hour minimum drive. And some day's it been a 4 hour drive, and that's each way.* [Participant 05, male, aged 70-79 years]

For patients reporting negative health impacts from travel (1/8, 13%), the option of a remote appointment was strongly welcomed:

> *To Birmingham would take me 4 hours, and 4 hours back.... To be perfectly honest, it made me ill for days.* [Participant 06, male, aged 60-69 years]

Local Clinical Testing

Of the 8 patients, 4 (50%) reported that it was convenient to obtain test results locally, but for others, it was challenging, at least at the start of their involvement in the study. GPs were often unsure which tests were needed and whether they were permitted to order them. Sending results to the hospital required pragmatism: 50% (4/8) of the participants uploaded their results via the myhealth portal, but the consultant could not always see them; the remaining 50% (4/8) relied on their local center (eg, GP) to send results; however, this could be problematic:

> *I had to chase them up a couple of times. When they did, the results weren't actually transferred onto myhealth.* [Participant 08, female, aged 60-69 years]

A lack of blood test results was reported by both consultants as making appointments more difficult, as they may not have had the full set of clinical information required for decision-making about patient treatment:

> *Well, it makes it more difficult to – so you wouldn't make a decision about changing medications or adjusting doses without blood test results.... If patients say "oh yes I had my bloods and they told me the results were fine" I think it's important to know what the results actually were rather than just relying on their recollection that the results were fine.* [Participant 01, consultant]

However, both consultants agreed that when results were available, remote consultations became more efficient compared with standard appointments:

> *But actually it's much better than previous [face-to-face appointments], because obviously normally what happens is I take bloods [in clinic] and then I get the results a few days later. Whereas then we have the result discussed in the [video] clinic.* [Participant 02, consultant]

Benefits to the Hospital

Although a health service cost analysis could not be undertaken for this study, remote consultations have the potential to release clinic spaces for those needing face-to-face consultations. This was seen by both consultants as potentially facilitating cost savings for the Trust:

> *They [patients] don't have a nurse that's weighing them and measuring their blood pressure and taking bloods and sending their bloods to the lab for processing. All that I guess is cost saving.* [Participant 01, consultant]

Being seen as innovative and making use of available technology was also regarded as important, although the need for appropriate space in the hospital for remote consultations was highlighted as a challenge:

> *You just need an environment. I mean an office is maybe not the best. You almost need a quiet room with the right computer facilities, you know, and no extra noise and no one coming in or out.* [Participant 02, consultant]

Implications for Implementation

Despite some reservations being expressed about issues such as clinical testing and audio or visual problems, all patients and

staff who were interviewed welcomed remote consultations and saw these as becoming crucial for routine follow-up care in the future, as long as patients could exercise choice about their appointment type. Trust support for the long-term implementation of remote consultations was seen as essential. In this study, only consultants undertook remote appointments. If the technology was to be used in other specialties, other members of the clinical team, such as registrars, would need to conduct appointments. Consultants were apprehensive about this and about the adjustments required:

> *I would find it more difficult to, because then I will be still be retaining responsibility for that patients' care that the registrar has seen in a virtual clinic and although we're all relatively used to the concept of taking responsibility for patient care when they're seen by somebody other than you in clinic, I guess it's just something new that I might get used to.* [Participant 01, consultant]

However, both consultants considered it prudent to train staff to highlight the differences between face-to-face and remote consultations. With regard to future appointments, 63% (5/8) of patients expressed a clear preference to continue with remote appointments, with the option of attending hospital if necessary:

> *Maybe I think alternate it, so one on the virtual and another one.... Sort of physically.* [Participant 08, female, aged 60-69 years]

In contrast, the remaining 38% (3/8) of patients reported that face-to-face appointments would still be their preference. Both consultants felt that remote consultations would be increasingly required, given the rising demand for hospital services while acknowledging that they may not suit all patients:

> *So I think we have to do something to try and mitigate against that and it's an obvious solution to some of those issues and I think it will work really well for a number of patients. It won't work for everybody but I think it would substantially reduce or potentially reduce the number of patients that actually have to come to [the hospital].* [Participant 01, consultant]

Discussion

Principal Findings

Recruitment to our study was lower than expected: the pool of eligible patients was comparatively small, and the recruitment rate was only 26.6% (54/203). Consequently, the study design was altered from a planned definitive trial to one that focused on the feasibility of administering the intervention and collecting evaluative data. Although no participants left the study, the crossover rate from intervention to usual care was high at 45% (12/29), with most changes to the study arm made at the request of the patient before their first remote appointment. This impaired our ability to fully assess processes and outcomes for patients in the intervention group and implies that future evaluations of outpatient remote consultations may benefit from a cohort rather than a randomized design.

Limitations

The principal limitation of this study was the need to change design from the definitive trial outlined in the published protocol [22] to a feasibility RCT following difficulties with recruitment. Many patients were ineligible because of clinical or technological issues, and recruitment rates were low, with only 54 patients recruited in total. There was a substantial crossover from intervention to usual care, leaving only 30% (16/54) of patients receiving the intervention. Consequently, the study was unable to definitively evaluate the impact of remote consultations on patient satisfaction, and the incidence of technical issues that affected the mode or quality of remote consultations was self-reported and not objectively measured. Similarly, although the baseline characteristics of patients in each study arm were comparable, we could not compare the characteristics of patients recruited to the study with those of the wider liver transplant patient cohort at QEHB, which affects the external validity of our findings. However, our in-depth qualitative process evaluation and assessment of the feasibility of implementing remote consultations for routine follow-up care in the hospital setting may offer important insights to others attempting to establish similar services.

Comparison With Prior Work

Many existing studies in this field have been undertaken in primary care rather than secondary care, and there are comparatively fewer randomized studies in which a remote consultation intervention is compared with standard or usual care or another appropriate control group. Most importantly, a few studies have assessed some of the practical challenges and facilitators affecting implementation in clinical practice (particularly the views of health care professionals on the feasibility of remote consultations), and to our knowledge, only one other study has assessed remote consultations for recipients of liver transplants [8]. Thus, we believe that our study adds to what is already known on this topic and can provide useful insights for practitioners seeking to implement remote consultations in routine care following their widespread use during the COVID-19 response.

Our overall findings suggest that remote consultations were effective, and despite technical challenges, interview data suggested enthusiasm for care to be delivered remotely in the future. The consultants involved in administering the intervention were similarly enthusiastic for the option of remote consultations to be available in routine practice. As reported by others [9,36], patients in both study groups were satisfied with their care, although the between-group difference was not significant after controlling for baseline scores. However, the substantial increase of 27 points between baseline and study end for patients receiving remote consultations may indicate that appointments undertaken *at a distance* can replace traditional face-to-face consultations without impairing patient satisfaction. This is particularly pertinent with regard to satisfaction scores on the other six domains of VSQ-9—scores for intervention patients remained higher than those in the usual care group even for features of the consultation that may be adversely affected by remote appointments, such as satisfaction with the personal manner of the consultant and perceptions of

their technical skills. Maintenance of positive patient-practitioner interactions may have been aided by relationships already having been established in previous face-to-face appointments [37,38]. This may indicate a key advantage of using remote consultations to deliver routine follow-up care within the hospital setting.

Consultants delivering remote consultations reported that the duration of remote consultations was typically shorter than that of face-to-face appointments and more business-like, with small talk kept to a minimum [16,39]. We were unable to assess health service costs; however, it may be reasonable to assume that remote consultations have the potential to save resources for the NHS. It is important to ensure that remote consultations do not simply displace costs associated with routine patient follow-up to other parts of the health and social care system [39]. Our study demonstrated the feasibility of collecting data on health service use, and this should be an important element of future research assessing the service and cost implications of offering remote consultations, particularly when clinical test results must be obtained for review during remote appointments. Obtaining blood test results proved to be challenging for many participants. Although workarounds were found, it is clear that careful consideration of viable processes for local clinical testing is required to make this less burdensome for patients. It did not prove feasible to collect meaningful data on patient costs; however, qualitative evidence showed that remote consultations were perceived to offer cost and time savings for many participants. The opportunity for remote consultation was also beneficial for patients who lived far from the hospital and for whom the journey was considered detrimental to their health [40].

Implications for Implementation in Clinical Practice

Given the growing demand for hospital services and the increasing pressure on hospital infrastructure, it is likely that more intensive use of remote consultations will be needed in the future to deliver routine patient follow-up care. Our study highlights a number of areas that require consideration to support effective implementation, and our findings also give important insights into whether remote consultations remain acceptable to patients and health care professionals beyond their unprecedented expansion during the COVID-19 response. First, for conditions where the results from clinical tests must be available during consultations, a robust system for patients to obtain these tests locally should be put in place [3,4]. Second, as has been reported by others, technical issues with the myVideoClinic software were frequently experienced by both consultants and patients [9,16], and less than half of remote consultations took place using video as planned. Technology is constantly evolving, and the facility (currently being developed) to offer access to remote consultations via smartphones is likely to resolve many of these technical issues. Other studies have also reported reluctance to use remote consultations on the part of consultants [38], and training should be offered to support the implementation of virtual models of care and address any associated ethical challenges [37]. There may be a core group of patients who do not wish to have remote consultations under any circumstances. Our survey of nonparticipants showed reluctance to have a remote appointment for a range of reasons, including dislike of the idea of remote follow-up, poor engagement with technology, preference for attending in-person because they lived nearby or enjoyed coming to the hospital, or a perception that obtaining medication and clinical tests locally would be too challenging. This suggests that remote consultations should be offered subject to patient choice and support traditional models of follow-up care rather than replacing them [12]. Unanswered questions remain regarding the equity of access to virtual consultations (particularly for rural patients who may not only live far from hospital sites but also experience poor technological infrastructure) and their suitability for diverse clinical specialties in which face-to-face consultation may continue to be necessary.

Conclusions

The NHS faces substantial pressure on resources and infrastructure because of increasing patient demand. Using technology to support routine patient care may ease some of these pressures. This study demonstrates that using remote consultations in routine follow-up for liver transplant patients is not detrimental to patients' overall care experience and may have a positive impact on patient satisfaction. During the COVID-19 pandemic, telephone and video-based appointments, rather than face-to-face contacts, have been offered almost exclusively by many NHS Trusts to slow down virus transmission and reduce the strain on NHS services. Although patients and practitioners may have accepted these changes as a temporary measure, our study shows that a number of technical and process issues must be resolved if the routine use of remote consultations is to be acceptable to patients and staff in the postpandemic NHS.

Acknowledgments

This research was funded by the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care West Midlands (NIHR CLAHRC WM), now recommissioned as the NIHR Applied Research Collaboration West Midlands (NIHR ARC WM). The views expressed in this publication are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. The authors would like to thank the members of the myVirtualClinic RCT Steering Group, as follows: Gill Combes, Sarah Damery, Eric Deeson, James Ferguson (chair), Foyzal Miah, Pamela Nayyar, Elaine O'Connell Francischetto, Aziz Sheikh, and Katie Squire. The data monitoring committee consisted of Eleanor Walsh (University of Bristol), Theo Arvanitis (University of Warwick), and Richard Aspinall (Portsmouth Hospitals NHS Trust). The study sponsor was the University of Birmingham. The study sponsor had no role in the study design; the collection, analysis, and interpretation of data; the writing of the report; or the decision to submit the paper for publication. The corresponding author had full access to all of the study data and had the final responsibility for the decision to submit for publication.

Authors' Contributions

JF, SD, and EOF conceived and designed the study. JF had overall oversight of the study, and JF, EOF, and JJ undertook recruitment. EOF and JJ performed the data collection, JJ undertook the qualitative data analysis, and SD carried out the quantitative data analysis. JJ and SD produced the first draft of the manuscript. JF, EOF, KJ, and RL critically reviewed the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V1.6.1).
[PDF File (Adobe PDF File), 366 KB - jmir_v23i9e19232_app1.pdf]

Multimedia Appendix 2
COREQ (Consolidated Criteria for Reporting Qualitative Studies) 32-item checklist.
[DOC File , 18 KB - jmir_v23i9e19232_app2.doc]

Multimedia Appendix 3
Topic guide for patient and staff interviews.
[DOC File , 21 KB - jmir_v23i9e19232_app3.doc]

Multimedia Appendix 4
Participant flow diagram.
[DOC File , 29 KB - jmir_v23i9e19232_app4.doc]

Multimedia Appendix 5
Summary of themes and supplementary quotations from the qualitative process evaluation.
[DOC File , 17 KB - jmir_v23i9e19232_app5.doc]

References

1. NHS Next steps on the NHS five year forward view. National Health Service. 2017. URL: https://www.england.nhs.uk/wp-content/uploads/2017/03/NEXT-STEPS-ON-THE-NHS-FIVE-YEAR-FORWARD-VIEW.pdf [accessed 2020-06-01]
2. NHS long term plan. National Health Service. 2019. URL: https://www.longtermplan.nhs.uk [accessed 2020-02-03]
3. Almathami HK, Win KT, Vlahu-Gjorgievska E. Barriers and facilitators that influence telemedicine-based, real-time, online consultation at patients' homes: systematic literature review. J Med Internet Res 2020 Feb 20;22(2):e16407 [FREE Full text] [doi: 10.2196/16407] [Medline: 32130131]
4. Al-Izzi T, Breeze J, Elledge R. Clinicians' and patients' acceptance of the virtual clinic concept in maxillofacial surgery: a departmental survey. Br J Oral Maxillofac Surg 2020 May;58(4):458-461. [doi: 10.1016/j.bjoms.2020.03.007] [Medline: 32199652]
5. Ditchburn J, Marshall A. Renal telemedicine through video-as-a-service delivered to patients on home dialysis: A qualitative study on the renal care team members' experience. J Ren Care 2017 Sep;43(3):175-182. [doi: 10.1111/jorc.12207] [Medline: 28614637]
6. Gilbert AW, Jaggi A, May CR. What is the patient acceptability of real time 1:1 videoconferencing in an orthopaedics setting? A systematic review. Physiotherapy 2018 Jun;104(2):178-186. [doi: 10.1016/j.physio.2017.11.217] [Medline: 29361298]
7. Fatehi F, Martin-Khan M, Smith AC, Russell AW, Gray LC. Patient satisfaction with video teleconsultation in a virtual diabetes outreach clinic. Diabetes Technol Ther 2015 Jan;17(1):43-48. [doi: 10.1089/dia.2014.0159] [Medline: 25296189]
8. Le LB, Rahal HK, Viramontes MR, Meneses KG, Dong TS, Saab S. Patient satisfaction and healthcare utilization using telemedicine in liver transplant recipients. Dig Dis Sci 2019 May;64(5):1150-1157. [doi: 10.1007/s10620-018-5397-5] [Medline: 30519848]
9. Ignatowicz A, Atherton H, Bernstein CJ, Bryce C, Court R, Sturt J, et al. Internet videoconferencing for patient-clinician consultations in long-term conditions: A review of reviews and applications in line with guidelines and recommendations. Digit Health 2019;5:2055207619845831 [FREE Full text] [doi: 10.1177/2055207619845831] [Medline: 31069105]
10. Salisbury C, Murphy M, Duncan P. The impact of digital-first consultations on workload in general practice: modeling study. J Med Internet Res 2020 Jun 16;22(6):e18203 [FREE Full text] [doi: 10.2196/18203] [Medline: 32543441]

11. Daniel H, Sulmasy LS, Health and Public Policy Committee of the American College of Physicians. Policy recommendations to guide the use of telemedicine in primary care settings: an American College of Physicians position paper. Ann Intern Med 2015 Nov 17;163(10):787-789. [doi: 10.7326/M15-0498] [Medline: 26344925]
12. O'Cathail M, Sivanandan MA, Diver C, Patel P, Christian J. The use of patient-facing teleconsultations in the National Health Service: scoping review. JMIR Med Inform 2020 Mar 16;8(3):e15380 [FREE Full text] [doi: 10.2196/15380] [Medline: 32175911]
13. Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, Hanson P, et al. Advantages and limitations of virtual online consultations in a NHS acute trust: the VOCAL mixed-methods study. Health Serv Deliv Res 2018;6:21. [doi: 10.3310/hsdr06210] [Medline: 29939519]
14. Cimperman M, Brenčič MM, Trkman P, Stanonik MD. Older adults' perceptions of home telehealth services. Telemed J E Health 2013 Oct;19(10):786-790 [FREE Full text] [doi: 10.1089/tmj.2012.0272] [Medline: 23931702]
15. Williams OE, Elghenzai S, Subbe C, Wyatt JC, Williams J. The use of telemedicine to enhance secondary care: some lessons from the front line. Future Healthc J 2017 Jun;4(2):109-114 [FREE Full text] [doi: 10.7861/futurehosp.4-2-109] [Medline: 31098445]
16. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method study. J Med Internet Res 2018 Dec 17;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
17. Webster P. Virtual health care in the era of COVID-19. Lancet 2020 Apr 11;395(10231):1180-1181 [FREE Full text] [doi: 10.1016/S0140-6736(20)30818-7] [Medline: 32278374]
18. Howgego G, Sharma I, Kalu P. The rules for online clinical engagement in the COVID era. J Plast Reconstr Aesthet Surg 2020 Dec;73(12):2127-2135 [FREE Full text] [doi: 10.1016/j.bjps.2020.08.045] [Medline: 33051176]
19. Hincapié MA, Gallego JC, Gempeler A, Piñeros JA, Nasner D, Escobar MF. Implementation and usefulness of telemedicine during the COVID-19 pandemic: a scoping review. J Prim Care Community Health 2020;11:2150132720980612 [FREE Full text] [doi: 10.1177/2150132720980612] [Medline: 33300414]
20. Gilbert AW, Billany JC, Adam R, Martin L, Tobin R, Bagdai S, et al. Rapid implementation of virtual clinics due to COVID-19: report and early evaluation of a quality improvement initiative. BMJ Open Qual 2020 May;9(2):1-8 [FREE Full text] [doi: 10.1136/bmjoq-2020-000985] [Medline: 32439740]
21. Visit-Specific Satisfaction Instrument (VSQ-9). RAND Health. URL: http://www.rand.org/health/surveys_tools/vsq9.html [accessed 2018-03-01]
22. Francischetto EO, Damery S, Ferguson J, Combes G, myVideoClinic Randomised Evaluation Steering Group. Video clinics versus standard face-to-face appointments for liver transplant patients in routine hospital outpatient care: study protocol for a pragmatic randomised evaluation of myVideoClinic. Trials 2018 Oct 19;19(1):574 [FREE Full text] [doi: 10.1186/s13063-018-2953-4] [Medline: 30340637]
23. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS Consensus Group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Pilot Feasibility Stud 2016;2:64 [FREE Full text] [doi: 10.1186/s40814-016-0105-8] [Medline: 27965879]
24. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
25. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
26. myhealth patient portal. University Hospitals Birmingham NHS Foundation Trust. URL: https://www.myhealth.uhb.nhs.uk/ [accessed 2018-08-01]
27. Randomly assign subjects to treatment groups. GraphPad Software. URL: http://www.graphpad.com/quickcalcs/randomize1.cfm [accessed 2018-03-01]
28. Software security. Vidyo. URL: https://www.vidyo.com/wp-content/uploads/Vidyo.io/vidyo.io-Security-WhitePaper.pdf?ls=White%2520Paper [accessed 2017-11-01]
29. Wallace P, Haines A, Harrison R, Barber J, Thompson S, Jacklin P, Virtual Outreach Project Group. Joint teleconsultations (virtual outreach) versus standard outpatient appointments for patients referred by their general practitioner for a specialist opinion: a randomised trial. Lancet 2002 Jun 08;359(9322):1961-1968. [doi: 10.1016/s0140-6736(02)08828-1] [Medline: 12076550]
30. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208. [doi: 10.1016/0168-8510(90)90421-9] [Medline: 10109801]
31. IBM SPSS Statistics for Macintosh, Version 25.0. IBM Corp, Armonk, NY. URL: https://www.ibm.com/uk-en/analytics/spss-statistics-software [accessed 2018-01-01]
32. NVivo 12 Plus. QSR International Pty Ltd. 2018. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2018-01-01]

33. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
34. Basudev N, Crosby-Nwaobi R, Thomas S, Chamley M, Murrells T, Forbes A. A prospective randomized controlled study of a virtual clinic integrating primary and specialist care for patients with Type 2 diabetes mellitus. Diabet Med 2016 Jun;33(6):768-776. [doi: 10.1111/dme.12985] [Medline: 27194175]
35. Teare MD, Dimairo M, Shephard N, Hayman A, Whitehead A, Walters SJ. Sample size requirements to estimate key design parameters from external pilot randomised controlled trials: a simulation study. Trials 2014 Jul 03;15:264 [FREE Full text] [doi: 10.1186/1745-6215-15-264] [Medline: 24993581]
36. Palen TE, Price D, Shetterly S, Wallace KB. Comparing virtual consults to traditional consults using an electronic health record: an observational case-control study. BMC Med Inform Decis Mak 2012 Jul 08;12:65 [FREE Full text] [doi: 10.1186/1472-6947-12-65] [Medline: 22769592]
37. Ignatowicz A, Slowther A, Elder P, Bryce C, Hamilton K, Huxley C, et al. Ethical implications of digital communication for the patient-clinician relationship: analysis of interviews with clinicians and young adults with long term conditions (the LYNC study). BMC Med Ethics 2018 Feb 23;19(1):11 [FREE Full text] [doi: 10.1186/s12910-018-0250-0] [Medline: 29475437]
38. Qudah B, Luetsch K. The influence of mobile health applications on patient - healthcare provider relationships: A systematic, narrative review. Patient Educ Couns 2019 Jan 31:1081-1087. [doi: 10.1016/j.pec.2019.01.021] [Medline: 30745178]
39. Jones G, Brennan V, Jacques R, Wood H, Dixon S, Radley S. Evaluating the impact of a 'virtual clinic' on patient experience, personal and provider costs of care in urinary incontinence: A randomised controlled trial. PLoS One 2018;13(1):e0189174 [FREE Full text] [doi: 10.1371/journal.pone.0189174] [Medline: 29346378]
40. Hull SA, Rajabzadeh V, Thomas N, Hoong S, Dreyer G, Rainey H, et al. Do virtual renal clinics improve access to kidney care? A preliminary impact evaluation of a virtual clinic in East London. BMC Nephrol 2020 Jan 10;21(1):1-9 [FREE Full text] [doi: 10.1186/s12882-020-1682-6] [Medline: 31924178]

Abbreviations

CLAHRC: Collaboration for Leadership in Applied Health Research and Care
CONSORT: Consolidated Standards of Reporting Trials
COREQ: Consolidated Criteria for Reporting Qualitative Studies
GP: general practitioner
HRQoL: health-related quality of life
IT: information technology
NHS: National Health Service
NIHR: National Institute for Health Research
QEHB: Queen Elizabeth Hospital Birmingham
RCT: randomized controlled trial
VSQ-9: Visit-Specific Questionnaire-9
WM: West Midlands

Edited by G Eysenbach; submitted 09.04.20; peer-reviewed by F Kaliyadan, J Sturt, S Bidmon; comments to author 26.10.20; revised version received 01.02.21; accepted 02.08.21; published 17.09.21.

Please cite as:
Damery S, Jones J, O'Connell Francischetto E, Jolly K, Lilford R, Ferguson J
Remote Consultations Versus Standard Face-to-Face Appointments for Liver Transplant Patients in Routine Hospital Care: Feasibility Randomized Controlled Trial of myVideoClinic
J Med Internet Res 2021;23(9):e19232
URL: https://www.jmir.org/2021/9/e19232
doi:10.2196/19232
PMID:34533461

Viewpoint

A Blueprint for the Conduct of Large, Multisite Trials in Telemedicine

Patricia Commiskey[1], MA, DPH; April W Armstrong[2], MPH, MD; Tumaini R Coker[3,4], MBA, MD; Earl Ray Dorsey[5], MBA, MD; John C Fortney[6,7], PhD; Kenneth J Gaines[1], MBA, MD; Brittany M Gibbons[8], MPH; Huong Q Nguyen[9], RN, PhD; Daisy R Singla[10,11,12], PhD; Eva Szigethy[13], MD, PhD; Elizabeth A Krupinski[14], PhD

[1]Division of Stroke, Department of Neurology, Vanderbilt University Medical Center, Nashville, TN, United States

[2]Department of Dermatology, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States

[3]Department of Pediatrics, University of Washington School of Medicine, Seattle, WA, United States

[4]Seattle Children's Research Institute, Seattle, WA, United States

[5]Center for Health + Technology, University of Rochester Medical Center, Rochester, NY, United States

[6]Division of Population Health, Psychiatry & Behavioral Sciences, University of Washington, Seattle, WA, United States

[7]Center of Innovation for Veteran-Centered and Value-Driven Care, Health Services Research and Development, Department of Veteran's Affairs, Seattle, WA, United States

[8]Keck School of Medicine, University of Southern California, Los Angeles, CA, United States

[9]Division of Health Services Research & Implementation Science, Kaiser Permanente Southern California, Pasadena, CA, United States

[10]Center of Addiction and Mental Health, Toronto, ON, Canada

[11]Lunenfeld Tanenbaum Research Institute, Toronto, ON, Canada

[12]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

[13]Center for High Value Health Care, UPMC Insurances Division, Department of Psychiatry, Medicine, and Pediatrics, University of Pittsburgh, Pittsburgh, PA, United States

[14]Department of Radiology and Imaging Sciences, Emory University School of Medicine, Atlanta, GA, United States

Corresponding Author:
Patricia Commiskey, MA, DPH
Division of Stroke
Department of Neurology
Vanderbilt University Medical Center
1161 21st Avenue South
A-0118 Medical Center North
Nashville, TN, 37232-2551
United States
Phone: 1 615 875 7988
Email: patricia.commiskey@vumc.org

Abstract

Recent literature supports the efficacy and efficiency of telemedicine in improving various health outcomes despite the wide variability in results. Understanding site-specific issues in the implementation of telemedicine trials for broader replication and generalizability of results is needed. Lessons can be learned from existing trials, and a blueprint can guide researchers to conduct these challenging studies using telemedicine more efficiently and effectively. This viewpoint presents relevant challenges and solutions for conducting multisite telemedicine trials using 7 ongoing and completed studies funded by the Patient-Centered Outcomes Research Institute portfolio of large multisite trials to highlight the challenges in implementing telemedicine trials. Critical issues of ensuring leadership and buy-in, appropriate funding, and diverse and representative trials are identified and described, as well as challenges related to clinical, informatics, regulatory, legal, quality, and billing. The lessons learned from these studies were used to create a blueprint of key aspects to consider for the design and implementation of multisite telemedicine trials.

(J Med Internet Res 2021;23(9):e29511) doi:10.2196/29511

KEYWORDS

telemedicine trials; randomized trials; challenges; multisite; mobile phone

Introduction

Multisite clinical trials are critical to advance the detection, diagnosis, and treatment of disease and to drive valid and reliable knowledge that can be generalized to broad integrated populations and health care settings. Techniques to conduct multisite clinical trials vary as a function of many factors, including the intervention under consideration, patient population, institutional setting, country, and so on, and a vast amount of literature exists on how to conduct clinical trials of all sorts. Telemedicine technologies and trials provide researchers not only the opportunity to assess the impact of telemedicine interventions in different populations but also to use these technologies in the evaluation of other clinical interventions where it is not the telemedicine aspect being evaluated.

Why do we need a blueprint for researchers conducting multisite telemedicine trials? A key reason is that telemedicine trials present new challenges to researchers that are typically not encountered in more traditional clinical trials. Despite the steady growth in acceptance and use, telemedicine is still far from being mainstream, and despite the creation of practice guidelines by the American Telemedicine Association and a variety of professional societies [1], there is little uniformity in practice and thus outcomes. Although existing literature supports the efficacy and efficiency of telemedicine in improving outcomes as measured using a variety of metrics, there is wide variability in these results. An extensive body of literature on modern telemedicine already exists, going back well over 25 years, with at least two mainstream peer review telemedicine journals dedicated to presenting research results with an increasing emphasis on outcomes. Systematic literature reviews and meta-analyses are being conducted at accelerating rates, which could not happen reliably if a large enough body of published studies to draw from were not available. However, this body of literature as a whole includes many mixed, inconclusive, and even contradictory results [2-6].

This really is not surprising. Medical research literature as a whole is often characterized as having mixed or inconclusive evidence, especially when, like telemedicine, the topic of interest is a new technology, procedure, treatment, or intervention. When a new tool or technology is introduced, others jump on board with their (often proprietary) versions, and isolated studies at single institutions are conducted on different patient and provider populations, with slightly (or very) different protocols, using different workflows, different metrics, benchmarks, and statistical analyses. In addition, the clinical application of telemedicine in medical and behavioral care rapidly expanded during the COVID-19 emergency state, allowing providers to reach many patients virtually because of relaxation of the Centers for Medicare & Medicaid Services regulations and state licensures and stay at home orders around the country [7-9]. Although there are many case reports of successes, large-scale systematic studies evaluating the effectiveness of telemedicine across different populations and settings are needed more than ever [10].

What can we do to get “better evidence in telemedicine”? As with any new medical intervention, we need more valid and reliable evidence to conduct large, multisite telemedicine trials. We need to understand and control (or at least account for) site-specific issues in implementation for broader replication and generalizability of the results. We need to acknowledge and understand the heterogeneity of treatment effects in diverse populations, and when examining the compendium of results, account for this heterogeneity appropriately before claiming a given telemedicine intervention does or does not *work*.

Given this opportunity and the difficulties of conducting multisite trials in telemedicine described in this paper, a blueprint can point the way for researchers to conduct these challenging studies more efficiently and effectively. This will build an evidence base for understanding the true impact and use of telemedicine.

Multisite Telemedicine Clinical Trials

This paper examines 7 ongoing and completed studies funded by the Patient-Centered Outcomes Research Institute (PCORI) portfolio of large multisite trials to highlight the challenges in implementing telemedicine trials; each trial is described in Table 1. Key issues (eg, clinical, informatics, regulatory, legal, quality, and billing) are summarized and examples of how they were addressed are described, thereby providing a blueprint of topics that will typically need to be addressed in the planning and implementation of telemedicine clinical trials.

PCORI has funded more than 25 large, multisite, pragmatic trials in telemedicine, most of which are conducted across multiple states, with some in more than 10 different states. More than three-quarters of them are ongoing and at various stages of implementation. Representative trials were selected based on the stage of implementation and discussion with PCORI on which current trial participants might be willing to share their experiences.

Table 1. Selected large multisite trials funded by PCORI[a].

Principal investigator	Study title	Summary	PCORI contract number	ClinicalTrials.gov number
AWA	Improving Specialty-Care Delivery in Chronic Skin Diseases	This study evaluated an innovative CCH[b] model where patients and primary care providers could access dermatologists on the web directly and asynchronously via a pragmatic RCT[c] to test whether a CCH model results in equivalent improvements in disease severity, quality of life, and mental health, and whether the model provides better access to specialty care, compared with usual in-person care for psoriasis management.	IHS-071502-IC	NCT02358135
TRC	Using Telehealth to Deliver Developmental, Behavioral, and Mental Health Services in Primary Care Settings for Children in Underserved Areas	Using community-engaged research principles, FQHC[d] (n=6 sites) and CHMCs[e] (n=2 sites) designed a telehealth-based intervention to improve the referral system for children being referred to specialty mental health care from primary care. We conducted an RCT involving 342 children, aged 5-12 years, with mental health concerns to receive either a routine referral or the new telehealth-enabled referral system.	IH-12-11-4168-IC	NCT02396576
ERD	Connect.Parkinson	Connect.Parkinson assessed the feasibility, value, and benefits of telemedicine visits with a specialist for individuals with Parkinson disease (n=200) through randomization to usual care or their routine care enhanced by video visits with a Parkinson specialist in approximately 20 US states.	AD-12-11-4701	NCT02038959
KJG	C3FIT[f] Stroke Care Trial	C3FIT uses team-based, enhanced collaboration to follow patients from presentation at the Emergency Department through 12 months post discharge to compare joint commission-certified stroke care with an Integrated Stroke Practice Unit care model that uses nurse and lay health educator care teams to visit patients and caregivers at home or in rehabilitation or skilled nursing facilities to assess function and quality of life using telehealth technology for patients at 18 US clinical sites.	PCS-2017C3-9081	NCT04000971
HQN, Mularski RA	Noninferiority Comparative Effectiveness Trial of Home-Based Palliative Care (HomePal) Trial	HomePal is a CER[g] comparing two models of home-based palliative care, standard approach with nurses and physicians making home visits versus nurses in patients' home facilitating remote physician consults.	PLC-1609-36108	NCT03694431
DRS	SUMMIT[h] Trial	SUMMIT examines whether psychotherapy delivered via telemedicine is as effective as in-person sessions for perinatal populations in Toronto, Canada; Chapel Hill, North Carolina; and Chicago, Illinois.	PCS-2018C1-10621	NCT04153864
ES	Specialty Medical Homes to Improve Outcomes for Patients With IBD[i] and Behavioral Health Conditions Trial	This trial compares traditional face-to-face delivered to telemedicine delivered team-based care within a subspecialty medical home integrating behavioral and medical care for patients with IBD in Pittsburgh, Pennsylvania, New York City, New York, and Cleveland, Ohio.	IHS-2017C3-8930	NCT03985800

[a]PCORI: Patient-Centered Outcomes Research Institute.

[b]CCH: Collaborative Connected Health.

[c]RCT: randomized controlled trial.

[d]FQHC: federally qualified health center.

[e]CHMC: community mental health clinic.

[f]C3FIT: Co-ordinated, Collaborative, Comprehensive, Family-based, Integrated, Technology-Enabled Care.

[g]CER: community-engaged research.

[h]SUMMIT: Scaling Up Maternal Mental Healthcare by Increasing Access to Treatments.

[i]IBD: inflammatory bowel disease.

Issues and Challenges in Implementing Multisite Telemedicine Clinical Trials

As researchers consider moving to multisite clinical trials using telemedicine, several issues need to be considered. Telemedicine has shown success in overcoming geographic and time limitations in medical care delivery, and evidence shows improved access to care, higher patient satisfaction, and enhanced quality of care and value of care can result [11]. Using telemedicine as a means of conducting a multisite clinical trial or using multiple sites to assess telemedicine intervention also raises some important decisions and considerations that should be addressed during study design and implementation. Textbox 1 summarizes the specific study elements and key considerations for assessing the impact of telemedicine interventions and the use of telemedicine to conduct multisite trials that were derived from reviewing completed and ongoing PCORI trials.

Textbox 1. Study elements and key considerations for assessing the impact of telemedicine interventions.

Technology

- Selecting a telemedicine platform
- Information technology reviews and approvals
- Technology support for electronic failure and vendor changes
- Potential financial cost to patients and ways to mitigate

Regulatory and reimbursement

- Credentialing and privileging
- Insurance empaneling
- Health Insurance Portability and Accountability Act or security standards
- Billing and remuneration
- Contracting and budgeting
- Institutional review boards

Design and initiation

- Comparators and outcomes
- Cross-contamination
- Clinical workflow
- Patient and participant characteristics
- Training
- Health disparities

Implementation and sustainability

- Buy-in
- Patients and family caregiver engagement
- Sustainability

Technology

Overview

Selecting a telemedicine platform is a key first step in ensuring the success of the research project and should be carefully considered. In some cases, the research project will simply use the telehealth platform in which the institution or institutions are already using clinically. If a platform is not available, selecting one is a key aspect of study planning and execution and typically must be done in conjunction with the health care system. Considering the current availability, study context, cost, security, and compliance are essential. Furthermore, considering user-centeredness ensures that the study platform is acceptable and feasible for both the provider and target population. For example, the SUMMIT (Scaling Up Maternal Mental Healthcare by Increasing Access to Treatments) trial [12] used specific criteria to facilitate the selection of telemedicine platforms, including simplicity of use for patients and providers, the availability of easy-to-follow instructions, whether Health Insurance Portability and Accountability Act (HIPAA) or Personal Health Information Protection Act guidelines were met, and technical support. Finally, integration of the telemedicine platform with an institution's electronic medical record (EMR) was an important consideration. A robust telemedicine platform may not always integrate with the institution's EMR, which can present challenges in study visits, clinical care documentation, and continuity of care. In this case, using two EMR systems (one for institutional EMR and the

other for telemedicine EMR) can be used, but the study protocol must specify the integration of both systems and procedures for reconciling differences to ensure that documentation is complete in both systems. Understanding the impact of using multiple systems is essential, and differences in the integration of the telemedicine platform across different study sites need to be evaluated in confounder analyses. The SUMMIT trial used an existing platform that was free and readily used by all registered health professionals in Toronto, Canada, but an existing platform was not available in Chicago, Illinois; therefore, the study team opted for a commercial product. In all cases, the platform offered a version that was compliant with either the United States HIPAA security rule or the Canadian Personal Health Information Protection Act. User-centeredness requires ensuring a platform that is acceptable and feasible for the provider and target population. The SUMMIT trial used specific criteria to facilitate telemedicine platform implementation, including simplicity of use with easy instructions, ability to try out the platform, and technical support.

Information Technology Reviews and Approval

Researchers must work closely with their institution's information technology (IT) department (university, hospital, or both) during evaluation and selection to ensure that telemedicine platforms meet federal, state, and institutional privacy and security standards and provide robustness for efficient patient-provider communication. As research requirements and approvals vary, involving institutional IT representatives early is critical. The study team may need to take the lead in evaluating telemedicine platforms, involving IT to understand the *must-have* standards, including encryption, data security, confidentiality, digital certificates, log-in IDs and passwords, auto-log-offs, audit trails, and disaster recovery plans. The Co-ordinated, Collaborative, Comprehensive, Family-based, Integrated, Technology-enabled Care (C3FIT) trial added a high-level IT leader at their main institution to the trial, providing a concrete effort for an IT liaison for all study years. Furthermore, this person was involved in design well before applying for funding, which allowed the IT leader to be fully vested in the study, providing both consultation and guidance through IT review and while setting up agreements with sites, as well as throughout the study to assist with technology changes and user issues. The importance of this as the COVID-19 pandemic expanded cannot be overstated. All of the trials worked closely with their IT teams for both technology selection and deployment, such as holding dedicated workshops to help train providers and patients, providing training materials either on the internet or through written materials, sending IT support to patient homes if necessary, or helping them remotely and often training providers to troubleshoot issues to some degree.

Technology Support for Electronic Failure and Vendor Changes

Ensuring institutional support is essential both for the technology and to ensure seamless integration [13] into existing appointment, scheduling, documentation, panel management, and billing systems. Avoiding workarounds specific to research requires advance planning and ample time for reviews, approvals, and coordination across multiple business groups. Electronic failure and poor signal quality are unavoidable, especially in remote areas [14]. Contingency plans must be in place, for instance, to quickly switch from video to phone when connectivity is not reliable to avert unnecessary frustrations. Having service line agreements in place will ensure that IT is highly responsive to troubleshooting technical issues. Moreover, as change is the only constant with technology vendors, multiyear studies must be agnostic and agile in building for, using, and adapting to multiple platforms over time.

To this end, the inflammatory bowel disease Specialty Medical Home trial included contingency plans at all sites to use a video platform outside the EMR as a backup, a flexibility facilitated by COVID-19-related reduction in HIPAA regulatory requirements. Furthermore, HomePal began their study using two different video platforms across providers and sites with the knowledge that they would be replaced with another enterprise-wide platform about a year later. Prepping providers and patients ahead of time about that possibility is important and allowed for setting expectations regarding additional training and limited attachment to one platform. This is true for both clinic-to-clinic and clinic-to-home settings, especially the latter, where patients do not have the same access to IT support as they would in a clinic.

Regulatory and Reimbursement Navigation

Overview

Telemedicine studies may face fewer regulatory and reimbursement issues in integrated systems of care (ie, the Veteran's Administration and Kaiser Permanente) as distant and host sites are in the same health care system. All patients are enrolled in the same EMR, and no additional credentialing, privileging, or insurance paneling of telemedicine providers is required. Credentialing is the process of assessing and confirming the license or certification, education, training, and other qualifications of a licensed or certified health care practitioner), and privileging is the process of authorizing a health care practitioner's specific scope and content of patient care services [15].

In nonintegrated systems of care, scheduling and delivering health care via interactive video can be approached through enrollment in the health care system of the distant site, which requires no additional credentialing, privileging, or insurance paneling for providers. Billing for telemedicine encounters is efficient when done at scale, but providers do not have direct access to EMRs at the host site, increasing the potential for duplicative services (eg, medical tests), over-or contraindicated-prescribing, and lack of care and research coordination. Patients might also face unfamiliar health care and scheduling systems, as well as cost-sharing requirements. A more integrated, patient-centered, and potentially safer approach is to have distant site providers credentialed and privileged to practice, chart, and bill from the host site [16].

Billing and Remuneration

Medicare coverage analysis is typically performed to determine which expenses are considered *study-related* and which expenses are considered *standard of care*. For billing of study-related expenses, setting up a protocol for invoicing to the prime institution that includes payment terms, timelines, and an invoice template that reflects the terms of the executed agreement will ensure there are no gaps in research activities as a result of billing or payment issues. If a Medicare coverage analysis determines that telemedicine service is the standard of care, then the Centers for Medicare & Medicaid Services regulations and payer procedures will dictate how telemedicine services are reimbursed as a standard of care as a function of state [17].

Contracting and Budgeting (Indemnity)

The prime institution is responsible for working with participating sites to establish subcontracts. Executing a subaward involves negotiation of terms and fees, which often vary according to each site's local laws and institutional policies. The scope of work and proposed budget should be prepared ahead of time to allow participating sites to make revisions based on current institutional rates, anticipated personnel effort, and applicable fees for licensing and credentialing. Valuable time may be saved during contract execution if the budget has been completed and reviewed by the subrecipient before the negotiations. Investigators with the Improving Specialty-Care Delivery for Chronic Skin Diseases study engaged with performance sites at the beginning of the proposal development timeline to establish working relationships with the relevant administration and understand institutional contracting procedures and policies. The primary study team was able to receive approval for the financial aspects of all subrecipient efforts before the notice of the award. With these subawards preapproved and key entities from each institution actively engaged, project researchers saved valuable time and were able to begin study conduct at the satellite sites as much as 8 weeks sooner than if budgetary considerations had not already been finalized.

Malpractice is a concern for telemedicine encounters because of the lack of legal precedents. If the distant site specifies the host site as an approved *site of practice*, it will ensure that their distant site's indemnification coverage extends to the provider. However, the distant site's malpractice insurance will not cover the host site if named in a lawsuit; the host site will need to purchase supplementary gap indemnity coverage for telemedicine encounters [16].

Institutional Review Board

The National Institutes of Health policy number NOT-OD-16-094 advises that multisite trials use a single institutional review board (SIRB) to conduct and oversee the protection of human subjects (US HSS, 2016) in an effort to streamline the ethical review process and reduce administrative burden and redundancies inherent to multiple review boards. Although SIRB use may increase the speed of protocol approval, the policy introduces challenges, including the selection of an appropriate institutional review board of record and agreement from institutions to cede review to an outside review board [18].

Use of an SIRB system requires strict adherence to the approved trial protocol from participating sites to ensure the safety of subjects and scientific rigor across the study. Communication and reporting mechanisms for adverse events or unanticipated problems, protocol deviations, and study progress should be established before the initiation of the study. In addition, a standard operating procedure for the storage and maintenance of regulatory documents should be developed and distributed to participating sites to ensure that the sites agree to follow. The coordinating site for reporting and maintenance of regulatory items should establish clear communication and reporting expectations from the outset for the safety and integrity of the trial. The Improving Specialty-Care Delivery for Chronic Skin Diseases trial study team used an SIRB (the University of Southern California Institutional Review Board) to oversee the clinical trial conduct and monitor participant safety across participating sites (University of Colorado and University of California, Davis). Therefore, amendments to procedures and study documents, protocol deviations, and reportable events were submitted to and reviewed exclusively by the University of Southern California Institutional Review Board. For this study, using an SIRB for the multisite trial enabled a streamlined, consistent, and efficient process for oversight of study conduct, patient safety, and enrollment details. In this model, designated individuals from the University of Southern California coordinating team established frequent communication procedures before conducting research. These procedures included weekly team meetings, regular email updates, and a mechanism for requesting ad hoc meetings in the event of a time-sensitive item requiring urgent principal investigator input. The primary site developed and coordinated training activities for all performance site personnel on the institutional review board–approved research protocol and distributed detailed standard operating procedure documents in advance of study activation at each site. Site staff at the primary site were also available for questions and supplemental training, as needed. Frequent and regular communication across sites and standardized and contemporaneous training materials contributed to the well-being of enrolled participants, as study-wide compliance to the research protocol was indicated by the lack of major protocol deviations over the course of the study.

Design and Initiation

Overview

Constantly evolving technology, regulations, and payments create significant challenges in the rigor and selection of the study design (ie, parallel control or stepped wedge, quasi-experimental, cluster, or individual-level randomization) and conduct of any comparative effectiveness telehealth study, as these decisions can have substantial implications for implementation success within most complex, adaptive systems [19,20]. Several study design issues should be considered when implementing large, multisite trials involving telemedicine.

Comparators and Outcome Measurements

Usual care and other comparators change over time as telemedicine becomes more mainstream, making it harder to

maintain the comparison. It should be noted that significant changes in the delivery of telemedicine occurred during the recent pandemic that likely impacted these types of clinical trials. However, it is too early to objectively assess the extent of the impact and how it will affect these and future trials. Furthermore, understanding a study's outcome measures and how the data can be collected pragmatically while making use of EMR data is important. The SUMMIT trial will conduct a thorough process evaluation to determine relevant barriers and facilitators of telemedicine-delivered sessions from a multistakeholder perspective, including study participants, identified significant others, treatment providers, and other clinicians. This evaluation will answer key digital technology questions in mental health care [21], including the risks and benefits of using telemedicine versus in-person treatment for mental health care needs. The inflammatory bowel disease Specialty Medical Home trial carefully tracks fidelity to face-to-face versus telemedicine care delivery and records reasons for any cross-contamination observed. Regular meetings are conducted with clinical and research teams, as well as the study statistician, across the three sites and these are also monitored by the stakeholder advisory board to minimize cross-contamination.

Cross-contamination

In designing multisite trials, it is important to consider the prevention of cross-contamination between arms when examining the efficacy of telemedicine-delivered care compared with another arms. This aspect is closely tied to the study design (eg, randomization at the site, provider, or individual patient level). Developing implementation rules, involving key stakeholders in development, and physically separating implementation platforms can be useful in avoiding cross-contamination. Sharing clear rules of telemedicine-based arms with key stakeholders is essential, including study participants during informed consent, treatment providers during training, and others who may prefer one arm versus another (eg, referring clinicians). Arms can also be separated by training one cadre of providers to use the telemedicine-based platform, whereas the other uses the alternative. Finally, including key stakeholders in study development provides the opportunity to identify relevant barriers [22] and ensures a more holistic and patient-centered approach. The C3FIT study was randomized by clinical site and went to great lengths to keep their two study arms separated, appointing different personnel to work exclusively with each arm, holding separate arm-specific meetings, and sending separate communication, including tailored versions of monthly newsletters and other correspondence. It is also important to acknowledge that researchers cannot control every aspect of a study, including technology. For example, during the COVID pandemic, non–HIPAA-compliant technologies were allowed, and even though a study might require a specific platform, some providers might have chosen to use a different platform because it was more expedient, which is a potential source of cross-contamination.

Telemedicine Platform Training Protocol

Developing a training protocol is critical for ensuring quality patient care and meaningful outcome data. Providers and site staff should be trained accordingly, as additional participating sites will likely introduce increased variability in the use of the platforms. To ensure standardization, a training protocol that includes a detailed, step-by-step description of how and when to use the platform should be developed. In-person training sessions, video tutorials, and informational sheets are helpful resources to provide participating site personnel to ensure consistent adherence to the telemedicine platform procedures. C3FIT included a combination of recorded video training using a web-based training platform combined with short, focused, and ongoing web-based meetings with site personnel to reinforce training and have continued to reinforce these messages as the trial commenced. Role playing and superuser testimonials and tips were helpful components of HomePal's web training sessions. In the Improving Specialty-Care Delivery in Chronic Skin Diseases study, the main training and procedural components mandated for each site before activation were completion of a tutorial use of the selected telemedicine platform. Research coordinators, investigators, and providers were provided with a standardized training session customized to their role in the study and their access within the platform. For example, research coordinators were trained specifically on registration of patients to the web-based system, instructions for patients on how to use the platform to submit cases, and use of scheduling and messaging functionality of the system. Training for dermatologists and primary care physicians focused primarily on how to review patient cases, submit recommendations and prescriptions, and communicate securely with the patients, other providers, and the study team via the web-based platform. Training materials developed by the coordinating center were available for review by participating site personnel at any time to ensure consistent adherence to the telemedicine platform procedures.

Health Disparities

The fundamental aim of telemedicine is to increase access to care [23]. However, as with any new technology, adoption is often not equitable, and considering the impact on disparate populations is essential. Contributors to disparities [24] include the global divide between high- and low-income societies, a social divide between the information-rich and information-poor, and a democratic divide between those who use digital tools to engage in public life and those who do not. The potential for telemedicine initiatives to improve health care access [23] and reduce health disparities has been widely noted [25-28], particularly given ongoing physician and specialty shortages in rural areas. Telemedicine offers an almost unprecedented opportunity to break down geospatial barriers to care and expand access for underserved populations. Likewise, well-designed multisite trials have the potential to identify, refine, and provide evidence on how telemedicine can be used to reduce socioeconomic, racial, ethnic, and geographic disparities in care.

Although great promise exists, disparities in access to care continue and are driven by complex issues that often require

broad infrastructure investments to correct [27,29]. Rural and underserved areas considering telemedicine implementation face high equipment and start-up costs, limited personnel and reimbursement, and interoperability issues that limit collaboration [27,29], as well as policy and regulation restrictions [27]. Furthermore, inadequate access to broadband, particularly among rural residents, can limit access to telemedicine services [30], as can sociodemographic differences [27].

Since its beginning, telemedicine has sought to bridge these divides, but researchers must be conscious of these inequities and make efforts to not perpetuate them in the design and conduct of their studies. Multisite telemedicine intervention trials may maintain or increase existing disparities rather than decrease disparities for a number of reasons, as described below:

1. 1. The telemedicine program or intervention must be designed for clinics, hospitals, or institutions that serve the most disadvantaged populations. If clinical sites can more easily use and implement the program or intervention, there is an increased likelihood of a stronger health impact for disadvantaged populations. For example, if limited English proficiency patients and families cannot be accommodated during telemedicine visits, the potential impact may be limited to an important and often disadvantaged population [31]. In the Telehealth-Co-ordinated Referral project, researchers brought a federally qualified health center, local community mental health clinics, and parents together to design and implement an intervention to fit the resources, capacity, and needs of its stakeholders (primary care clinicians, specialty mental health care clinicians, and parents) [32].
2. 2. The program or intervention must be acceptable to its users, including providers, staff, patients, and families. For example, providers who care for more underserved patients often work in clinical facilities that are less resourced [33]. If these providers find the intervention too cumbersome for participation because of limited resources (eg, lack of sufficient nursing staff), they may be less likely to fully engage, thus altering the potential impact on the most disadvantaged patients. Another example is direct-to-consumer (DTC) telemedicine, which should be accessible preferentially by patients in more rural areas and in areas designated as Health Resources and Services Administration and primary health care professional shortage areas; however, a recent study of DTC demonstrated that users of one DTC service were not preferentially located in rural or primary health care professional shortage areas [34].
3. 3. The study protocol must be designed with the most disadvantaged group. For example, if the study protocol requires patients to complete a lengthy written survey, patients with lower literacy levels may be less likely to complete it; thus, the impact on these patients will not be well understood [35]. If the study requires patients to provide a social security number to receive a study incentive, undocumented immigrant patients may decline enrollment because of fear of immigration enforcement [36-38].

Connect.Parkinson, one of the first national randomized controlled trials to assess the feasibility, value, and benefits of specialist-led telemedicine visits for Parkinson disease, demonstrates the ability of telemedicine to address geographic barriers to care. Participants from roughly 20 US states, many in remote locations, participated in >90% of the visits completed. Each visit saved patients and their caregivers about 100 miles of travel and 3 hours of time. Thousands expressed interest in this telemedicine study, and the care model, assessed in 2014, foresaw the large latent demand and high receptivity for telemedicine among many older adults that emerged during the COVID-19 pandemic [39,40]. A limitation of this trial was that although it did address geographic disparities, it did not address social disparities, as the majority of participants were White and college educated. However, given the positive outcomes for these patients with Parkinson disease, this trial could readily be translated to more disadvantaged populations.

Researchers engaging in multisite telemedicine trials can do several things to enhance their study's ability to reduce health disparities:

1. Understand the underlying pre-existing health and health care disparities that relate to telemedicine intervention under study in the trial.
2. Incorporation of health systems, providers, staff, and patients that represent the most disadvantaged groups as partners and collaborators in the study design, implementation, planning, and conduct.
3. Design the trial to collect reliable demographic data on race and ethnicity, language, income, and rurality (at least) for all participants.
4. Power the study so that intervention impact can be studied by participant demographic factors (ie, race and ethnicity, and language). This may require certain populations to be overrepresented in their recruitment strategy.

HomePal addressed disparities in digital access and literacy by using *telepresenters* in the form of nurses who were at home to facilitate a remote video consultation with physicians. These nurses could work out any technology and clinical issues, which lessened the impact on patients and caregivers, which is particularly important for those without a supportive home environment. Furthermore, C3FIT sought to decrease access and financial barriers when COVID-19 prohibited in-person visits by providing internet-enabled tablets to participants who did not have access to a smartphone or computer to conduct study visits. The Connect.Parkinson trial mailed a webcam to participants who did not have one, and the SUMMIT study provided tablets to participants lacking access to technology.

Implementation and Sustainability

Overview

Sustainability may differ depending on the specific intervention being evaluated (eg, telemonitoring, telediagnostics, video consultations, and telecounseling). Implementation planning will benefit from using several frameworks to assess potential facilitators and barriers [41-43], readiness for large-scale trials

[44], and adjustment of study design or approach to maximize pragmatism in real-world settings [45].

Ensuring Health System Leaders, Providers, and Support Staff Buy-in

Pearl [46] identified five key areas to improve physician engagement in telehealth that readily apply to getting buy-in for research projects:

1. 1. Communicate the *Why:* Understanding the research and patient value, including the convenience of receiving medical care without travel or missing school and work, tends to resonate with physicians and other personnel. HomePal physicians were sold on using video visits and saw improved efficiency in seeing patients and reductions in windshield time; however, this benefit was not seen by home-based nurses, who managed both technology and synching up with physician schedule and availability [47].
2. 2. Compensate appropriately: compensating physicians based on relative value units or production incentives can be incentives for the adoption of telehealth and engagement in research, but including them as coinvestigators on grants and authors on publications can incentivize participation as well. Setting performance targets and incentives may assist by motivating the competition.
3. 3. Keep it simple: Cumbersome and flawed IT that requires multiple applications and clicks to move through will impede physician adoption and potentially impact data quality.
4. 4. Mind the workload: In general, telemedicine visits require the same time commitment as in-person visits; the provider should be given adequate support and staff to encourage its use and to carry out any additional research tasks.
5. 5. Invest in the culture: a shared EMR and technology and ensuring that rather than being penalized, providers will be encouraged to use telemedicine innovations will improve adoption and research engagement.

Patients and Family Caregiver Engagement

To optimize the engagement and representativeness of the study sample, designs must allow the care team to frame and set expectations that telemedicine tools will be used as appropriate for routine care to best meet the needs of patients and families. This approach allows for tailoring the technology to the needs and preferences of patients and families and the clinical situation, which ultimately reflects real-world practices. Video visits may not be appropriate for many older patients with hearing or vision impairment; however, video facilitates lip reading and is considered superior to phone if transportation to a clinic is not possible. Furthermore, patients and families must be assured that they can still easily access clinic visits or have a clinician to make a home visit if needed. Using home health staff to facilitate video consults with providers may overcome access disparities for the most vulnerable homebound populations [48,49]. This was especially important for HomePal, as their physicians could still make home visits if the video consults suggested a need to do so, which led to increased assurance by the patient and family regarding quality of care. Using video allowed multiple family caregivers to join the session if needed to ensure a care plan agreement.

Sustainability

Positive developments with telehealth regulations and payments in the United States have set the stage for telehealth to be an integral tool for care delivery. Thus, research questions will focus less on whether telehealth tools are effective or as good as the in-person gold standard, and more on how specific tools can be best leveraged, under what conditions to enhance quality of care and outcomes as systems and regulations evolve [50]; the telehealth intervention tested at the start of the study may not be the same at the end of the study. The pragmatic study design helps, in general, to foster sustainability after the study is completed.

Leadership and Funding

The importance of strong leadership and adequate funding in the development and implementation of successful telemedicine trials is especially important and cannot be understated. Telemedicine and telehealth service research has stressed the importance of identifying *champions* who can promote uptake, legitimize services, build relationships, and work through implementation challenges [20,51]. Having champions when building and executing telemedicine trials is essential, not only at the lead site or sites but also at clinical sites to allow working through institution-specific processes and barriers.

Although understanding the specific resources required to support telemedicine trials is still evolving, these types of trials are often expensive and require more resources than traditional trials. To obtaining sufficient funding, leaders who can advocate for and incorporate creative approaches to support may frequently be necessary. One question is whether to pay for clinical services out of a grant. Although possible, this should be used with caution to avoid creating potential bias in the trial (ie, including only insured patients). As with traditional clinical trials, every effort should be made to include as many patient-related expenses in the grant budget as much as possible. It is important to include telemedicine-specific items such as devices (eg, smartphones, tablets, electronic scales, blue-tooth-enabled blood pressure monitors), as in the C3FIT study. In many trials, these can be cycled through different patients as each patient completes the trial.

In many cases, grant funding alone is not adequate to cover all trial expenses. Researchers need to be creative at times to cover these expenses. For example, there are a number of companies with telemedicine products and platforms that are willing to collaborate on trials sometimes providing technology for free or at reduced prices with the hope that they will be included in future clinical budgets if they perform well. There are also state and federal programs designed to help providers and patients with telemedicine technology support. For example, during the COVID-19 pandemic the Federal Communications Commission had at least two rounds of funding, with one having dedicated funds to help patients with device and internet access.

Discussion

This manuscript provides a blueprint of lessons learned from seven PCORI trials in various stages of development and design through training and implementation, as well as the impact on

disparities. A feasible approach for telemedicine implementation is a necessary part of any telemedicine research study, and must be integrated into the study protocol and timeline, and this blueprint represents perspectives and provides potential solutions from researchers across the continuum, from just beginning their research to further in the process of completing multisite telemedicine trials.

Consumers, major drivers of telemedicine use, point to improved convenience and access to care as important benefits. Research has found high numbers of consumers who rate the convenience and access improvements important and are willing to try telemedicine visits [52], as well as extremely high levels of satisfaction [53]. This suggests that use of telemedicine and its use in multisite trials will only increase if designed and implemented successfully.

Minority and low-income populations have difficulty accessing care because of age, health, social support, location of medical centers, and numerous other social issues. There must be a greater positive impact on the most disparate or most disadvantaged group [54-56]. If the intervention works equally well in all groups or populations, disparity will be maintained. If the intervention works better in the more advantaged group, the disparity will widen. Without this type of blueprint for multisite telemedicine trials, we may miss the opportunity to create and implement telemedicine programs that actually improve existing disparities.

The populations that telemedicine serves will also expand from episodic to chronic conditions. Motivation for use will also expand from convenience to access to better patient-centered care. Finally, telemedicine applications will expand from hospitals and clinics to home and mobile devices [57]. As this migration occurs, many patients will require the support of local clinicians, such as those in community-based outpatient clinics that support millions of veterans [58]. These changes, combined with increasing access to the internet and familiarity with its health applications, will enable larger and more representative populations to participate in research. The challenge will remain to meet these individuals on their terms and to support them with services that are tailored to their needs.

Acknowledgments

The research reported in this publication was partially funded by the PCORI Awards (HIS-071502-IC, IH-12-11-4168-IC, AD-12-11-4701, PCS-2017C3-9081, HSRP-2018-1686, PCS-2018C1-10621, HIS-2017C3-8930). The views, statements, and opinions in this publication are solely the responsibility of the authors and do not necessarily represent the views of the PCORI, its Board of Governors, or Methodology Committee. PC is supported by a grant from PCORI (grant PCS-2017C3-9081) and partially supported by a grant from the United States Department of Agriculture Distance Learning and Telemedicine program (grant TN 754-A17). AWA is supported by a grant from PCORI (grant IHS-071502-IC). TRC is supported by a grant from PCORI (grant IH-12-11-4168-IC). ERD is supported by a grant from PCORI (grant AD-12-11-4701). JF was supported by a grant from the PCORI (grant PCS-1406-19295) and by a Research Career Scientist award from the Department of Veterans Affairs. KJG is supported by a grant from PCORI (grant PCS-2017C3-9081). BG is supported by a grant from PCORI (grant IHS-071502-IC). HQN is supported by a grant from PCORI (grant PLC-1609-36108). DRS is supported by a grant from PCORI (grant PCS-2018C1-10621) and by an Academic Scholars Award from the Department of Psychiatry at the University of Toronto. ES is supported by a grant from PCORI (grant IHS-2017C3-8930).

Authors' Contributions

PC and EAK led the development, compiled sections, and refined and edited the manuscript. All authors participated in the development of the concept, contributed content, and reviewed the final manuscript.

Conflicts of Interest

ERD has received honoraria for speaking at American Academy of Neurology courses, American Neurological Association, MCM Education, Physician's Education Resource, LLC, Stanford University, University of California Irvine, and University of Michigan. ERD also received compensation for consulting services from 23andMe, Abbott, Abbvie, Acadia, Acorda, Biogen, BrainNeuroBio, Clintrex, Curasen Therapeutics, DeciBio, Denali Therapeutics, Eli Lilly, Grand Rounds, Karger, MC10, Medopad, Michael J. Fox Foundation, Olson Research Group, Origent Data Sciences, Inc, Otsuka, Pear Therapeutics, Praxis, Roche, Sanofi, Spark, Sunovion Pharma, Theravance, and Voyager Therapeutics. In addition, ERD received research support from Abbvie, Biogen, Biosensics, Burroughs Wellcome Fund, Food and Drug Administration, Greater Rochester Health Foundation, Huntington Study Group, Massachusetts Institute of Technology, Michael J Fox Foundation, National Institutes of Health and National Institute of Neurological Disorders and Stroke and National Center for Advancing Translational Sciences, PICORI, Pfizer, Photopharmics, Roche, and Safra Foundation; editorial services for Karger Publications; and ownership interests with Grand Rounds (second opinion service).

References

1. Krupinski EA, Antoniotti N, Bernard J. Utilization of the American Telemedicine Association's clinical practice guidelines. Telemed J E Health 2013 Nov;19(11):846-851 [FREE Full text] [doi: 10.1089/tmj.2013.0027] [Medline: 24050615]

2. Timpel P, Oswald S, Schwarz PE, Harst L. Mapping the evidence on the effectiveness of telemedicine interventions in diabetes, dyslipidemia, and hypertension: an umbrella review of systematic reviews and meta-analyses. J Med Internet Res 2020 Mar 18;22(3):e16791 [FREE Full text] [doi: 10.2196/16791] [Medline: 32186516]
3. Harst L, Timpel P, Otto L, Richter P, Wollschlaeger B, Winkler K, et al. Identifying barriers in telemedicine-supported integrated care research: scoping reviews and qualitative content analysis. J Public Health 2019 Apr 09;28(5):583-594. [doi: 10.1007/s10389-019-01065-5]
4. Yang YT, Weintraub E, Haffajee RL. Telemedicine's role in addressing the opioid epidemic. Mayo Clin Proc 2018 Sep;93(9):1177-1180 [FREE Full text] [doi: 10.1016/j.mayocp.2018.07.001] [Medline: 30097301]
5. Bashshur RL, Howell JD, Krupinski EA, Harms KM, Bashshur N, Doarn CR. The empirical foundations of telemedicine interventions in primary care. Telemed J E Health 2016 May;22(5):342-375 [FREE Full text] [doi: 10.1089/tmj.2016.0045] [Medline: 27128779]
6. Bashshur RL, Shannon GW, Bashshur N, Yellowlees PM. The empirical evidence for telemedicine interventions in mental disorders. Telemed J E Health 2016 Feb;22(2):87-113 [FREE Full text] [doi: 10.1089/tmj.2015.0206] [Medline: 26624248]
7. Ohannessian R, Duong TA, Odone A. Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: a call to action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]
8. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
9. Hollander JE, Carr BG. Virtually perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
10. Ekeland AG, Bowes A, Flottorp S. Methodologies for assessing telemedicine: a systematic review of reviews. Int J Med Inform 2012 Jan;81(1):1-11. [doi: 10.1016/j.ijmedinf.2011.10.009] [Medline: 22104370]
11. Foley survey finds healthcare executives pursuing telemedicine despite hurdles. Foley & Lardner Blog. 2014. URL: https://www.foley.com/en/insights/publications/2014/11/foley-survey-finds-health-care-executives-pursuing [accessed 2021-08-12]
12. Singla DR, Meltzer-Brody SE, Silver RK, Vigod SN, Kim JJ, La Porte LM, et al. Scaling Up Maternal Mental healthcare by Increasing access to Treatment (SUMMIT) through non-specialist providers and telemedicine: a study protocol for a non-inferiority randomized controlled trial. Trials 2021 Mar 05;22(1):186 [FREE Full text] [doi: 10.1186/s13063-021-05075-1] [Medline: 33673867]
13. Dang S, Olsan T, Karuza J, Cai X, Gao S, Intrator O, et al. Telehealth in home-based primary care: factors and challenges associated with integration into veteran care. J Am Geriatr Soc 2019 Sep 25;67(9):1928-1933. [doi: 10.1111/jgs.16045] [Medline: 31343732]
14. Ignatowicz A, Atherton H, Bernstein CJ, Bryce C, Court R, Sturt J, et al. Internet videoconferencing for patient-clinician consultations in long-term conditions: a review of reviews and applications in line with guidelines and recommendations. Digit Health 2019 Apr 23;5:2055207619845831 [FREE Full text] [doi: 10.1177/2055207619845831] [Medline: 31069105]
15. Health center program compliance manual. Health Resources and Services Administration. URL: https://bphc.hrsa.gov/programrequirements/compliancemanual/index.html [accessed 2021-04-03]
16. Fortney JC, Veith RC, Bauer AM, Pfeiffer PN, Valenstein M, Pyne JM, et al. Developing telemental health partnerships between state medical schools and federally qualified health centers: navigating the regulatory landscape and policy recommendations. J Rural Health 2019 Jun;35(3):287-297 [FREE Full text] [doi: 10.1111/jrh.12323] [Medline: 30288797]
17. Telehealth services. Centers for Medicare & Medicaid Services. 2021. URL: https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/Downloads/TelehealthSrvcsfctsht.pdf [accessed 2021-08-11]
18. Fortney JC, Heagerty PJ, Bauer AM, Cerimele JM, Kaysen D, Pfeiffer PN, et al. Study to promote innovation in rural integrated telepsychiatry (SPIRIT): rationale and design of a randomized comparative effectiveness trial of managing complex psychiatric disorders in rural primary care clinics. Contemp Clin Trials 2020 Mar;90:105873. [doi: 10.1016/j.cct.2019.105873] [Medline: 31678410]
19. Venkataramani AS, Underhill K, Volpp KG. Moving toward evidence-based policy: the value of randomization for program and policy implementation. JAMA 2020 Jan 07;323(1):21-22. [doi: 10.1001/jama.2019.18061] [Medline: 31730191]
20. Ellimoottil C, An L, Moyer M, Sossong S, Hollander JE. Challenges and opportunities faced by large health systems implementing telehealth. Health Aff (Millwood) 2018 Dec;37(12):1955-1959. [doi: 10.1377/hlthaff.2018.05099] [Medline: 30633667]
21. Hollis C, Sampson S, Simons L, Davies EB, Churchill R, Betton V, et al. Identifying research priorities for digital technology in mental health care: results of the James Lind Alliance Priority Setting Partnership. Lancet Psychiatry 2018 Oct;5(10):845-854. [doi: 10.1016/S2215-0366(18)30296-7] [Medline: 30170964]
22. Kaplan B, Litewka S. Ethical challenges of telemedicine and telehealth. Camb Q Healthc Ethics 2008;17(4):401-416. [doi: 10.1017/S0963180108080535] [Medline: 18724880]
23. Bashshur R, Shannon G. History of Telemedicine: Evolution, Context, and Transformation. New Rochelle, NY: Mary Ann Liebert, Inc; 2009.
24. Norris P. Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide. Cambridge: Cambridge University Press; 2001.

25. Gonzalez BD. Promise of mobile health technology to reduce disparities in patients with cancer and survivors. JCO Clinical Cancer Informatics 2018 Dec;2:1-9. [doi: 10.1200/cci.17.00141] [Medline: 30652578]
26. deShazo RD, Parker SB. Lessons learned from Mississippi's telehealth approach to health disparities. Am J Med 2017 Apr;130(4):403-408. [doi: 10.1016/j.amjmed.2016.11.005] [Medline: 27899245]
27. Dorsey ER, Topol EJ. State of telehealth. N Engl J Med 2016 Jun 14;375(2):154-161. [doi: 10.1056/NEJMra1601705] [Medline: 27410924]
28. Marcin JP, Shaikh U, Steinhorn RH. Addressing health disparities in rural communities using telehealth. Pediatr Res 2016 Jan;79(1-2):169-176 [FREE Full text] [doi: 10.1038/pr.2015.192] [Medline: 26466080]
29. Nelson R. Telemedicine and telehealth: the potential to improve rural access to care. Am J Nurs 2017 Jun;117(6):17-18. [doi: 10.1097/01.NAJ.0000520244.60138.1c] [Medline: 28541980]
30. Drake C, Zhang Y, Chaiyachati KH, Polsky D. The limitations of poor broadband internet access for telemedicine use in rural America: an observational study. Ann Intern Med 2019 Sep 3;171(5):382-384. [doi: 10.7326/m19-0283] [Medline: 31108509]
31. Phelps RA, Coker TR. First, do no harm: improving access in autism diagnostic assessments for children without exacerbating inequities in care. J Dev Behav Pediatr 2019 Apr;40(3):217-218. [doi: 10.1097/DBP.0000000000000663] [Medline: 30907772]
32. Coker TR, Porras-Javier L, Zhang L, Soares N, Park C, Patel A, et al. A telehealth-enhanced referral process in pediatric primary care: a cluster randomized trial. Pediatrics 2019 Mar;143(3):e20182738 [FREE Full text] [doi: 10.1542/peds.2018-2738] [Medline: 30770523]
33. Bach PB, Pham HH, Schrag D, Tate RC, Hargraves JL. Primary care physicians who treat blacks and whites. N Engl J Med 2004 Aug 05;351(6):575-584. [doi: 10.1056/NEJMsa040609] [Medline: 15295050]
34. Uscher-Pines L, Mulcahy A, Cowling D, Hunter G, Burns R, Mehrotra A. Access and quality of care in direct-to-consumer telemedicine. Telemed J E Health 2016 Apr;22(4):282-287 [FREE Full text] [doi: 10.1089/tmj.2015.0079] [Medline: 26488151]
35. Raich PC, Plomer KD, Coyne CA. Literacy, comprehension, and informed consent in clinical research. Cancer Invest 2001;19(4):437-445. [doi: 10.1081/cnv-100103137] [Medline: 11405182]
36. Berk M, Schur C. The effect of fear on access to care among undocumented Latino immigrants. J Immigr Health 2001 Jul;3(3):151-156. [doi: 10.1023/A:1011389105821] [Medline: 16228780]
37. Toomey RB, Umaña-Taylor AJ, Williams DR, Harvey-Mendoza E, Jahromi LB, Updegraff KA. Impact of Arizona's SB 1070 immigration law on utilization of health care and public assistance among Mexican-origin adolescent mothers and their mother figures. Am J Public Health 2014 Feb;104 Suppl 1(Suppl 1):S28-S34 [FREE Full text] [doi: 10.2105/AJPH.2013.301655] [Medline: 24354823]
38. Martinez O, Wu E, Sandfort T, Dodge B, Carballo-Dieguez A, Pinto R, et al. Evaluating the impact of immigration policies on health status among undocumented immigrants: a systematic review. J Immigrant Minority Health 2015 Jun;17(3):947-970. [doi: 10.1007/s10903-013-9968-4] [Medline: 24375382]
39. Beck CA, Beran DB, Biglan KM, Boyd CM, Dorsey ER, Schmidt PN, Connect.Parkinson Investigators. National randomized controlled trial of virtual house calls for Parkinson disease. Neurology 2017 Sep 12;89(11):1152-1161 [FREE Full text] [doi: 10.1212/WNL.0000000000004357] [Medline: 28814455]
40. Early impact of CMS expansion of Medicare telehealth during COVID-19. Health Affairs Blog. 2020. URL: https://www.healthaffairs.org/do/10.1377/hblog20200715.454789/full/ [accessed 2021-08-12]
41. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
42. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
43. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method study. J Med Internet Res 2018 Apr 17;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
44. Baier RR, Jutkowitz E, Mitchell SL, McCreedy E, Mor V. Readiness assessment for pragmatic trials (RAPT): a model to assess the readiness of an intervention for testing in a pragmatic trial. BMC Med Res Methodol 2019 Jul 18;19(1):156 [FREE Full text] [doi: 10.1186/s12874-019-0794-9] [Medline: 31319789]
45. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ 2015 May 08;350:h2147. [doi: 10.1136/bmj.h2147] [Medline: 25956159]
46. Engaging physicians in telehealth. NEJM Catalyst. 2017. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.17.0458 [accessed 2021-08-12]
47. Osuji TA, Macias M, McMullen C, Haupt E, Mittman B, Mularski RA, et al. Clinician perspectives on implementing video visits in home-based palliative care. Palliat Med Rep 2020 Oct 6;1(1):221-226 [FREE Full text] [doi: 10.1089/pmr.2020.0074] [Medline: 34223480]

48. Funderskov KF, Raunkiær M, Danbjørg DB, Zwisler A, Munk L, Jess M, et al. Experiences with video consultations in specialized palliative home-care: qualitative study of patient and relative perspectives. J Med Internet Res 2019 Mar 21;21(3):e10208 [FREE Full text] [doi: 10.2196/10208] [Medline: 30896436]
49. Nguyen HQ, Mularski RA, Edwards PE, Lynn J, Machado MT, McBurnie MA, HomePal Research Group. Protocol for a noninferiority comparative effectiveness trial of home-based palliative care (HomePal). J Palliat Med 2019 Sep;22(S1):20-33. [doi: 10.1089/jpm.2019.0116] [Medline: 31486724]
50. Schwamm LH. Telehealth: seven strategies to successfully implement disruptive technology and transform health care. Health Aff (Millwood) 2014 Feb;33(2):200-206. [doi: 10.1377/hlthaff.2013.1021] [Medline: 24493761]
51. Wade V, Eliott J. The role of the champion in telehealth service development: a qualitative analysis. J Telemed Telecare 2012 Dec;18(8):490-492. [doi: 10.1258/jtt.2012.gth115] [Medline: 23209264]
52. Moses RH, Hurley B, Devan P, Bhat R. What can health systems do to encourage physicians to embrace virtual care? Deloitte Insights. 2018. URL: https://www2.deloitte.com/content/dam/insights/us/articles/4407_Virtual-care-survey/DI_Virtual-care-survey.pdf [accessed 2021-08-12]
53. Polinski JM, Barker T, Gagliano N, Sussman A, Brennan TA, Shrank WH. Patients' satisfaction with and preference for telehealth visits. J Gen Intern Med 2016 Mar 13;31(3):269-275 [FREE Full text] [doi: 10.1007/s11606-015-3489-x] [Medline: 26269131]
54. McPheeters ML, Kripalani S, Peterson NB, Idowu RT, Jerome RN, Potter SA, et al. Closing the quality gap: revisiting the state of the science (vol. 3: quality improvement interventions to address health disparities). Evid Rep Technol Assess (Full Rep) 2012 Aug(208.3):1-475. [Medline: 24422952]
55. Weinick RM, Hasnain-Wynia R. Quality improvement efforts under health reform: how to ensure that they help reduce disparities--not increase them. Health Aff (Millwood) 2011 Oct;30(10):1837-1843. [doi: 10.1377/hlthaff.2011.0617] [Medline: 21976324]
56. Lion KC, Raphael JL. Partnering health disparities research with quality improvement science in pediatrics. Pediatrics 2015 Feb;135(2):354-361 [FREE Full text] [doi: 10.1542/peds.2014-2982] [Medline: 25560436]
57. Dorsey ER, Glidden AM, Holloway MR, Birbeck GL, Schwamm LH. Teleneurology and mobile technologies: the future of neurological care. Nat Rev Neurol 2018 May;14(5):285-297. [doi: 10.1038/nrneurol.2018.31] [Medline: 29623949]
58. Fortney JC, Pyne JM, Kimbrell TA, Hudson TJ, Robinson DE, Schneider R, et al. Telemedicine-based collaborative care for posttraumatic stress disorder: a randomized clinical trial. JAMA Psychiatry 2015 Jan;72(1):58-67. [doi: 10.1001/jamapsychiatry.2014.1575] [Medline: 25409287]

Abbreviations

C3FIT: Co-ordinated, Collaborative, Comprehensive, Family-based, Integrated, Technology-enabled Care
DTC: direct-to-consumer
EMR: electronic medical record
HIPAA: Health Insurance Portability and Accountability Act
IT: information technology
PCORI: Patient-Centered Outcomes Research Institute
SIRB: single institutional review board
SUMMIT: Scaling Up Maternal Mental Healthcare by Increasing Access to Treatments

Edited by G Eysenbach; submitted 09.04.21; peer-reviewed by J Switzer, P Yellowlees; comments to author 30.04.21; revised version received 25.06.21; accepted 27.07.21; published 20.09.21.

Please cite as:
Commiskey P, Armstrong AW, Coker TR, Dorsey ER, Fortney JC, Gaines KJ, Gibbons BM, Nguyen HQ, Singla DR, Szigethy E, Krupinski EA
A Blueprint for the Conduct of Large, Multisite Trials in Telemedicine
J Med Internet Res 2021;23(9):e29511
URL: https://www.jmir.org/2021/9/e29511
doi:10.2196/29511
PMID:34542417

Original Paper

An Ostomy Self-management Telehealth Intervention for Cancer Survivors: Technology-Related Findings From a Randomized Controlled Trial

Ronald S Weinstein[1], MD, FCAP, FATA; Michael J Holcomb[1], BS; Julia Mo[2], BS; Peter Yonsetto[1], BS; Octavio Bojorquez[2], BS; Marcia Grant[3], PhD; Christopher S Wendel[4], MS; Nancy J Tallman, BSN; Elizabeth Ercolano[5], DNSc; Zuleyha Cidav[6], PhD; Mark C Hornbrook[7], PhD; Virginia Sun[3], PhD; Ruth McCorkle[8†], PhD; Robert S Krouse[9], MD

[1]Arizona Telemedicine Program, The University of Arizona Health Sciences, Tucson, AZ, United States
[2]Hospital of the University of Pennsylvania, Philadelphia, PA, United States
[3]Nursing Research and Education, City of Hope National Medical Center, Duarte, CA, United States
[4]The University of Arizona, Tucson, AZ, United States
[5]School of Public Health, Yale University, New Haven, CT, United States
[6]University of Pennsylvania, Philadelphia, PA, United States
[7]Center for Health Research, Kaiser Permanente, Northwest Region, Portland, OR, United States
[8]Yale University, New Haven, CT, United States
[9]Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States
[†]deceased

Corresponding Author:
Ronald S Weinstein, MD, FCAP, FATA
Arizona Telemedicine Program
The University of Arizona Health Sciences
1501 N Campbell AHSL 1156
Tucson, AZ, 85724
United States
Phone: 1 5203733552
Fax: 1 5206264774
Email: rweinstein@telemedicine.arizona.edu

Abstract

Background: An Ostomy Self-management Telehealth (OSMT) intervention by nurse educators and peer ostomates can equip new ostomates with critical knowledge regarding ostomy care. A telehealth technology assessment aim was to measure telehealth engineer support requirements for telehealth technology–related (TTR) incidents encountered during OSMT intervention sessions held via a secure cloud-based videoconferencing service, *Zoom for Healthcare*.

Objective: This paper examines technology-related challenges, issues, and opportunities encountered in the use of telehealth in a randomized controlled trial intervention for cancer survivors living with a permanent ostomy.

Methods: The Arizona Telemedicine Program provided telehealth engineering support for 105 OSMT sessions, scheduled for 90 to 120 minutes each, over a 2-year period. The OSMT groups included up to 15 participants, comprising 4-6 ostomates, 4-6 peer ostomates, 2 nurse educators, and 1 telehealth engineer. OSMT-session TTR incidents were recorded contemporaneously in detailed notes by the research staff. TTR incidents were categorized and tallied.

Results: A total of 97.1% (102/105) OSMT sessions were completed as scheduled. In total, 3 OSMT sessions were not held owing to non–technology-related reasons. Of the 93 ostomates who participated in OSMT sessions, 80 (86%) completed their OSMT curriculum. TTR incidents occurred in 36.3% (37/102) of the completed sessions with varying disruptive impacts. No sessions were canceled or rescheduled because of TTR incidents. Disruptions from TTR incidents were minimized by following the TTR incident prevention and incident response plans.

Conclusions: Telehealth videoconferencing technology can enable ostomates to participate in ostomy self-management education by incorporating dedicated telehealth engineering support. Potentially, OSMT greatly expands the availability of ostomy self-management education for new ostomates.

Trial Registration: ClinicalTrials.gov NCT02974634; https://clinicaltrials.gov/ct2/show/NCT02974634

(J Med Internet Res 2021;23(9):e26545) doi:10.2196/26545

KEYWORDS

telehealth; telemedicine; cloud computing; ostomy; cancer survivors; family caregivers; self-management; patient education; videoconferencing; mobile phone

Introduction

Background

An intestinal stoma, or ostomy, is a surgically created opening in the abdomen that provides an alternate pathway for stool or urine to exit the body. An ostomy may be needed for a patient due to cancer, trauma, inflammatory bowel disease, bowel obstruction, infection, incontinence, or diverticular disease [1]. According to the United Ostomy Associations of America, approximately 100,000 ostomy surgeries are performed annually in the United States [2]. Ostomies have been shown to be associated with multiple health-related quality of life difficulties, irrespective of the type of or reason for ostomy [3]. Ostomy complication rates of 21%-70% have been reported in previous studies [4].

Successfully caring for and living with an ostomy requires the development of specific skills and regimens by the patient. Without proper care, ostomy sites may develop irritation or infection. Ostomates may develop psychosocial complications if they lack knowledge of ostomy self-management or coping skills. Many ostomates face multiple barriers to support and resources, such as travel distances, lack of transportation, financial restrictions, and lack of access to certified wound, ostomy, and continence nurses (WOCNs) [5].

A systematic review conducted by Faury et al [6] found that patient education programs have a positive impact on certain psychosocial and self-management skills for colorectal cancer survivors with ostomies. However, no consensus has been reached regarding the optimal curriculum content and delivery method. Our previous studies have examined the unique patient-reported challenges faced by colorectal cancer survivors with ostomies, highlighting the importance of education, skill building, and emotional support (both formal and informal) [7]. It is critical to develop accessible survivor education programs that address the needs, concerns, and struggles experienced by cancer survivors with ostomies.

Objectives

We developed the Ostomy Self-management Telehealth (OSMT) intervention to help ostomates learn to care for their ostomy sites and adapt to living with an ostomy. The program content was first delivered in an in-person setting in which WOCN educators taught new ostomates about ostomies, ostomy care, and life with an ostomy. The curriculum design provided an opportunity for new ostomates to learn from people with long-term ostomies (peer ostomates) who have successfully adapted to life with an ostomy [5,7,8]. While this program was beneficial for cancer survivors with ostomies, survivors had barriers to participation [8]. Programs that deliver education to patients with advanced disease may face participant attrition due to disease progression and other limitations that would prompt withdrawal from the program [9]. Technology-aided alternatives such as videotaped education or computer-aided instruction provide accessibility but sacrifice the interactivity and active discussion offered by in-person education sessions [10].

Telehealth is potentially a preferred alternative for delivering self-management education to patients with chronic diseases [11,12]. However, patients face potential technology barriers as well, including challenges related to their preparedness to participate in videoconferencing, suitability of various types of videoconferencing devices, and local network connectivity management issues [13-17]. This paper aims to examine technology challenges, issues, and opportunities encountered in the use of telehealth in a cancer survivor education program [12-15].

Methods

Recruitment and Design of the OSMT Intervention

The overall design of the OSMT intervention, including the block randomization method, has been described in detail elsewhere [5]. The OSMT program is a multisite group intervention designed to support self-efficacy and patient activation for ostomy self-management among cancer survivors. This study was approved by the Human Subjects Protection institutional review boards of the participating research sites. The OSMT program was delivered to 21 consecutive groups of ostomates via a series of 4 telehealth sessions conducted by a team of trained WOCN educators and peer ostomates (hereafter referred to as *peers*). A fifth group education session that ostomates did not attend was delivered via telehealth to family caregivers (FCs) or support persons (SPs). OSMT sessions typically lasted 90-120 minutes and were generally held weekly. Study participants were recruited from health care delivery sites in 3 cities: City of Hope National Medical Center in Los Angeles, California; the Hospital of the University of Pennsylvania in Philadelphia, Pennsylvania; and Yale-New Haven Medical Center in New Haven, Connecticut. This study enrolled cancer survivors who were English-speaking, aged ≥21 years, and underwent ostomy surgery at least 6 weeks before the first session. FCs or SPs aged ≥21 years who provided assistive care for the patients were also enrolled, although their

participation was not a requirement for patient enrollment. Participants were randomized to either the intervention arm or the usual care (UC) arm. Participants in both arms were given information postoperatively on ostomy-related resources in their respective areas; however, the study team did not provide UC arm participants with the OSMT program.

Telehealth Design, Preintervention Preparations, and Technological Support for Participants

The OSMT program used telehealth technology managed and supported by the Arizona Telemedicine Program (ATP) at the University of Arizona's College of Medicine, in Tucson, Arizona [5,14,15]. OSMT sessions with groups of ostomates were held via the *Zoom for Healthcare* encrypted videoconferencing service provided by Zoom Video Communications Inc. The *Zoom for Healthcare* service enables secure group video calls (SGVCs). A Health Insurance Portability and Accountability Act business associate agreement between the University of Arizona's ATP and Zoom Video Communications was executed as required.

In this OSMT intervention, the OSMT trainee groups were linked into SGVCs as *virtual training* groups. Before the first session of the OSMT curriculum for each OSMT group, study coordinators from each accrual site paired up with the ATP telehealth engineers to train participants on the technologies they would use to participate in the SGVCs. ATP implemented telehealth technology incident prevention plans (IPPs) and incident response plans (IRPs). IPPs included instructing, testing, and requesting participants to connect up to 30 minutes early for the scheduled OSMT sessions to allow time to resolve any last-minute technical difficulties before the start of the session and after telehealth technology–related (TTR) incident follow-up to ensure that any problems were fully resolved. Instruction was provided to ostomates and respective FCs or SPs on loading the Zoom (hereafter referred to as videoconferencing software) app on their engagement device of choice (a PC, laptop, smartphone, or tablet), on using basic videoconferencing software controls, and on following privacy and SGVC etiquette during the OSMT sessions. These instructions and pre-OSMT-session test video calls were designed to familiarize all participants with commonly used videoconferencing software features, such as muting and unmuting the microphone, enabling and disabling sharing of their video camera's image, and the positioning of their video camera. If participants did not have an engagement device or an internet connection, a study tablet was loaned to them. The IRPs were uncomplicated. ATP telehealth engineers staffed each OSMT session in-person to provide an immediate response to any TTR incidents that occurred. In cases where participants encountered technology difficulties (problems with network connectivity or videoconferencing devices) or needed additional assistance with the videoconferencing software platform, ATP telehealth engineers could immediately and often proactively provide technical support as needed. Ostomates and their respective FC or SP generally participated in OSMT sessions from home but could also optionally connect from an alternative location with adequate privacy and broadband internet connectivity.

PCs and laptops were the most common types of devices used by intervention patients to connect to OSMT sessions, followed by tablets and smartphones (Table 1). Most patients used a single device for all of their OSMT sessions.

Table 1. Devices used to connect to Ostomy Self-management Telehealth (OSMT) sessions by 93 intervention patients that participated in OSMT sessions.

Device	Values, n (%)
PC or laptop	45 (48)
Tablet or iPad (Apple Inc)	26 (28)
Smartphone	11 (12)
Unknown	10 (11)
More than 1 device	1 (1)

Ostomy Self-management Telehealth SGVCs

SGVCs included up to 15 participants per session, including up to 6 ostomates, 6 peers, 2 WOCN educators, and an ATP telehealth engineer. Generally, 2 WOCN educators staffed each OSMT session. This paradigm was not intended for tandem-teaching; rather, 1 WOCN educator was responsible for teaching all 4 patient-training units and the other taught the FC and SP training unit. In addition, 4-6 peers were integral to the delivery of the curriculum and discussion. Each WOCN educator provided backup and support to the other for the OSMT sessions as needed. Session disruptions for any reason during multiparticipant OSMT sessions were of concern because they are distracting and waste multiple participants' time simultaneously. The built-in redundancy of WOCN educators and peers helped alleviate such occurrences.

New cohorts of ostomates for the OSMT program were created on a rolling basis as new ostomates accrued to the intervention arm of the study. The pair of WOCN educators connected to all group sessions from a Health Insurance Portability and Accountability Act–compliant videoconferencing room in the ATP office suite on the first floor of the Arizona Health Sciences Library, located at the University of Arizona campus in Tucson [14]. In general, peers participated from their homes.

Telehealth Support Operations

At the ATP headquarters in Tucson, Arizona, the ATP engineer on duty joined the 2 nurse educators in the ATP videoconferencing room for each OSMT session. This setup allowed for expedited technical assistance and troubleshooting by the ATP telehealth engineer during the sessions. During each OSMT session, the engineer responded immediately to any

technical problems (eg, loss of audio or telecommunications network instability). For complex support incidents requiring more than a few minutes to rectify, the engineer would physically relocate offline to the ATP's Control Room on the same floor as the ATP videoconference room, approximately 50 feet away, and work with the participants experiencing TTR problems to resolve the problem and assist them in reconnecting to their session. The engineer then returned to the videoconferencing room with the WOCNs to continue monitoring and supporting the ongoing session.

The ATP telehealth engineers each had decades of experience working in the telehealth industry. They had extensive prior experience working with rural telehealth site coordinators, nurses, and the public, but had little prior experience working directly with patients.

Based on the study protocol and participant agreement, OSMT sessions were not recorded in the interest of preserving the privacy and confidentiality of the ostomates and FCs or SPs participating in each session. In the United States, it is standard practice in most clinical telehealth applications not to record videoconferencing sessions with patients.

Outcomes and Data Collection

Throughout this study, data on telehealth support, participant acceptance, technology-associated problems, and related outcomes were collected using several approaches. TTR observations from WOCN, peer, and telehealth engineer field notes were stored in the form of contemporaneous notes in the master data set of the OSMT-session data. The comments also included non-TTR observations.

The observations were classified into 2 categories. A *major* incident involved telehealth engineer intervention. A *minor* incident did not require telehealth engineer intervention and could be resolved by an ostomate, a peer, or a WOCN nurse educator. The field notes collected were reviewed to identify and count the TTR incidents experienced in each session. Study coordinators collected the session data and field notes recorded by the OSMT program team and stored them in a study database that contained all OSMT-session written observations. In addition, reasons for declining study participation owing to technical problems or attrition due to technical problems were also recorded as part of the study implementation data.

Data Analysis

All comments containing TTR observations were analyzed by an ATP telehealth engineer to identify TTR support incidents and to determine whether an engineer intervened for each incident. Each TTR incident was counted as major if an engineer intervened and minor if an engineer did not intervene. TTR incidents were also categorized by audio, video, internet connection, software, or equipment problems. TTR incidents with no details about the nature of the problem were categorized as *unspecified.* Specific search terms were used to identify technology-related concepts and comments for analysis purposes (Textbox 1). All categorized data were reviewed by the research team. Data that were discordantly categorized were discussed by the research team, and a consensus decision was made. Counts for all groups and categories of TTR incidents were tallied using descriptive summary statistics (frequencies, percentages, etc).

Textbox 1. Technology search terms.

Search Terms

- *Android*
- *App*
- *Audio*
- *Broadband*
- *Call* (nonspecific, but useful, "technology term" identifiers)
- *Camera*
- *Cellular*
- *Computer*
- *Connect(ed)*
- *Connection*
- *Data*
- *Device*
- *Disconnect(ed)*
- *Display* (nonspecific, but useful, "technology term" identifiers)
- *Drop(ped)*
- *Hear* (nonspecific, but useful, "technology term" identifiers)
- *Install*
- *Internet*
- *iOS*
- *iPad*
- *iPhone*
- *Laptop*
- *Light(ing)* (nonspecific, but useful, "technology term" identifiers)
- *Mic* (nonspecific, but useful, "technology term" identifiers)
- *Microphone*
- *Mute* (nonspecific, but useful, "technology term" identifiers)
- *Network*
- *Phone*
- *Screen*
- *See* (nonspecific, but useful, "technology term" identifiers)
- *Signal*
- *Smartphone*
- *Tablet*
- *Technical*
- *Technology*
- *Telehealth*
- *Telemedicine*
- *Troubleshooting*
- *Uninstall*
- *Unmute*
- *Video*
- *WiFi*

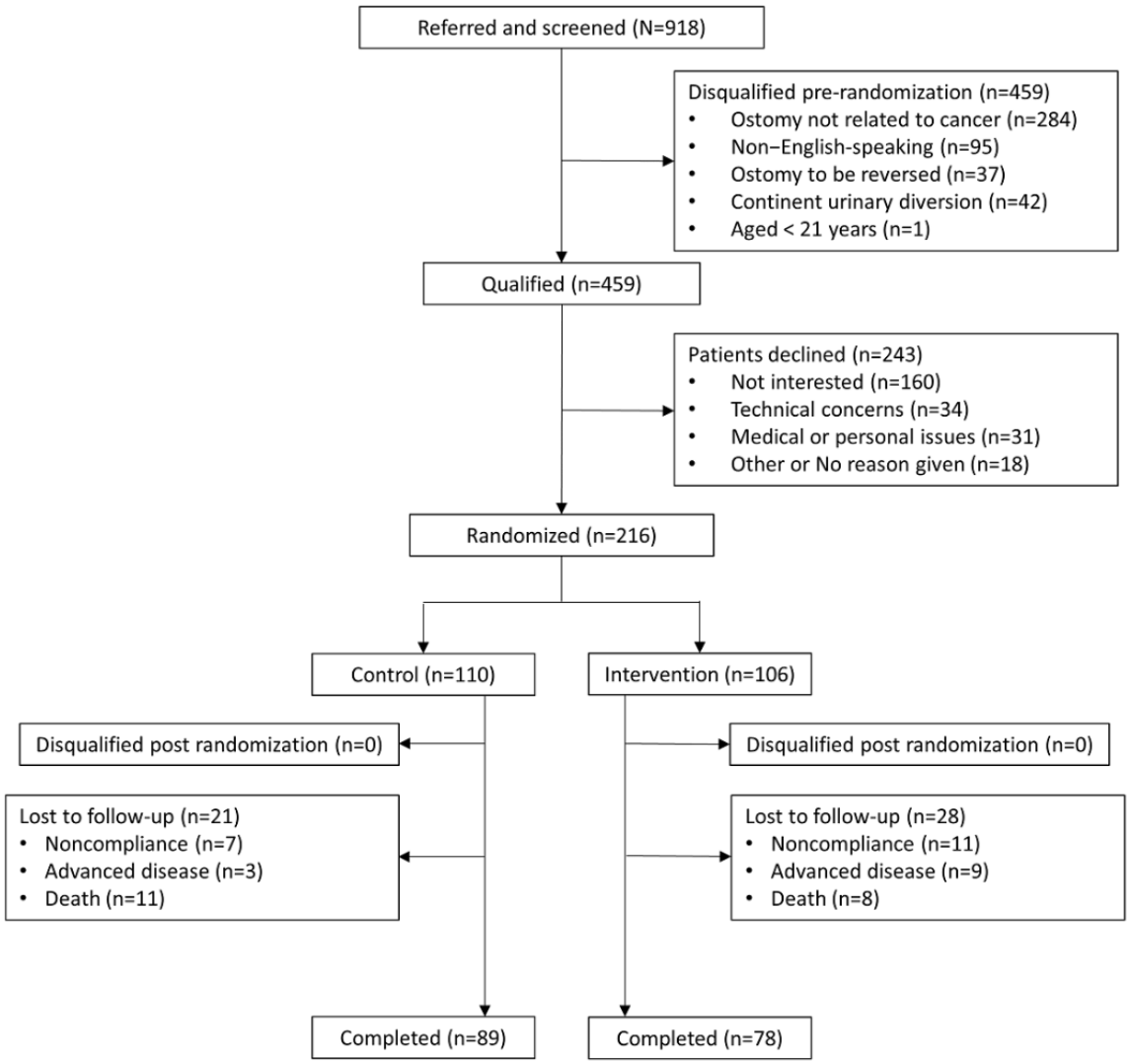
- *Wired* (nonspecific, but useful, "technology term" identifiers)
- *Wireless*
- *Zoom*

Results

Study Participation, Completion, and Telehealth Use

The study CONSORT (Consolidated Standards of Reporting Trials) flow diagram [18] showing recruitment and retention is presented in Figure 1. Of the 459 cancer survivors qualified for participation in the study, 7.4% (34) declined participation due to technology-related concerns. Of the 34 technology-related concerns among qualified patients who declined participation, 88% (30) were related to *fear of technology*, and 12% (4) claimed to have no prior experience with technology at all. Of the 216 survivors who consented to participate in this study, 50.9% (110) were randomized to the UC arm, and 49.1% (106) randomized to the OSMT intervention arm. Of the 106 survivors randomized to intervention, 12.2% (13) subsequently opted not to participate in any OSMT sessions. The reasons for not joining any sessions included the following: technology concerns 31% (4/13), illness 15% (2/13), and opting out of the study 54% (7/13). Of the 106 intervention participants, 87.7% (93) attended at least 1 OSMT session. Of the 93 participants that attended OSMT sessions, 86% (80) completed the OSMT curriculum and 14% (13) did not complete the OSMT curriculum. The reasons for not completing the OSMT curriculum included technology concerns 15% (2/13), advanced illness or death 46% (6/13), and no longer wanting to participate 39% (5/13). Moreover, 21 UC participants and 28 intervention participants did not complete the study for the reason lost to follow-up. Of all participants lost to follow-up, 57% (12/21) of UC participants and 39% (11/28) of intervention participants were lost to follow-up at 6 months.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.



Telehealth Technology–Related Incidents

OSMT-session field notes were logged for all 102 completed group OSMT sessions. Frequencies of noted TTR observations are presented in Table 2.

Occurrences of TTR incidents described in the OSMT-session field notes were divided into two categories: major and minor incidents (Table 3). Major incidents included a telehealth engineer intervening to provide technical support such as working one-on-one with a participant, offline from the session in-progress in some instances, to resolve a range of problems such as lack of audio or poor audio quality, camera placement or image quality, or network performance to enable the participant to rejoin the group. Minor incidents did not require a telehealth engineer intervening. Some examples of minor incidents include peers or nurse educators assisting participants with microphone muting and unmuting controls, a participant disconnecting and then reconnecting to the session on their own, or a participant experiencing a minor degradation of audio or video. Of the 102 completed OSMT sessions, 36.3% (37) had at least 1 TTR incident noted (Table 3). No session had more than 2 TTR incidents noted.

Table 2. Frequencies of Ostomy Self-management Telehealth sessions with 0, 1, or 2 telehealth technology–related incidents noted per session (n=102).

TTR[a] incidents noted per session	Values, n (%)
0 incidents	65 (63.7)
1 incident	28 (27.5)
2 incidents	9 (8.8)

[a]TTR: telehealth technology–related.

Table 3. Frequency distribution of major and minor telehealth technology–related (TTR) incidents noted in 37 sessions that had at least one TTR incident.

Type of telehealth technology–related incident noted	Sessions, n (%)
2 major and 0 minor	4 (11)
1 major and 1 minor	3 (8)
1 major and 0 minor	15 (40)
2 minor and 0 major	2 (5)
1 minor and 0 major	13 (35)

No OSMT sessions were canceled or rescheduled because of TTR failure. A total of 3 sessions were canceled for reasons unrelated to telehealth technology.

A total of 46 TTR incidents, occurring during a total of 37 OSMT sessions, were noted. In total, 21.5% (22/102) sessions had a major incident. From the 46 TTR incidents, audio and internet connection problems accounted for most TTR incidents with a combined total of 36 (78%) incidents, whereas problems with video, software, and equipment totaled 6 (13%) incidents. Furthermore, 9% (4/46) of problems noted did not specify the type of technology problem encountered.

Most TTR incidents were experienced by the ostomates, as compared with the peers or the nurse educators. While 28 noted incidents described technology problems primarily experienced by ostomates (n=26) and caregivers (n=2), 16 technology problems were experienced by peers (n=11), nurse educators (n=3), or both peers and nurse educators (n=2). Two noted incidents did not specify participants who experienced the technology problem.

Telehealth OSMT Intervention Receptivity

The OSMT program was generally well-received, and many intervention participants emphasized the ease and accessibility of the OSMT program provided by the telehealth component. One peer noted, after a particularly upbeat session, how “superior” the telehealth format was compared to in-person support groups because of the “ease in format for meeting and communicating with people who live apart.” Comments from peers and ostomates highlighted important aspects of the secure telehealth real-time videoconference format for the OSMT sessions, citing the “safe environment for expressions of vulnerability and important emotions” and the reliability of the telehealth platform to “function effectively despite severe weather,” which would have discouraged travel to an in-person meeting.

Discussion

Principal Findings

The results of this study are timely owing to the dramatic surge in telehealth usage propelled by the COVID-19 pandemic. The ATP’s telehealth engineers are experienced in supporting and testing complex telemedicine networks in demanding environments such as rural and prison environments with unstable access to broadband telecommunications and challenging cancer clinical patient-training settings such as the multisite SGVC cancer survivor support training sessions for ostomates, as described in this paper.

Multiple patient participants in SGVCs can potentially increase the probability of session interruptions due to TTR incidents. On the other hand, multiple participants in each session are also desirable for enriching the learning experience for the patient

participants and increasing the number of patients who complete the training program. The probability of interruption or failure to complete an OSMT training session for TTR reasons becomes a factor in calculating the level of engineering support needed to approach zero disruptions due to TTR reasons and the cost-effectiveness of this intervention.

The ATP's Clinical Research Unit specializes in implementing and operating high-quality cloud computing, enabling patient encounter environments. In our experience, the videoconferencing platform used in this study was robust and reliable throughout the OSMT sessions. However, there are numerous well-recognized potential technical failure points that can be barriers to successful participation in SGVCs, ranging from problems with hardware, software, and network connectivity to end-user training. For example, generally, all SGVC participants are expected to install videoconferencing software on a compatible device, work with telehealth engineers to troubleshoot any TTR problems encountered when connecting to the videoconferencing platform, learn to appropriately use the microphone mute and unmute controls to prevent unintended audio disruption of the OSMT session, and learn to *produce and manage* their video image on camera.

This study designed and tested a novel telehealth engineering support strategy for a cloud computing–based videoconferencing clinical health care delivery model with a trimodal support model consisting of a combination of IPPs and IRPs: (1) proactive end-user onboarding and training to use their respective technology suite to participate in OSMT SGVC sessions; (2) telehealth engineer real-time in-session monitoring, management, and support of OSMT-session participants, including patients, peers, and nurse educators; and (3) proactive telehealth engineer follow-up technical support and problem-solving to address and resolve any TTR problems not resolved during an OSMT session. In this randomized controlled patient clinical trial, despite the occurrence of TTR incidents in some OSMT sessions, no sessions were canceled due to TTR failure. This high OSMT-session operational success level was achieved by (1) incorporating 15-90 minutes of presession, one-on-one patient-telehealth engineer instruction sessions; (2) assigning an in-the-classroom dedicated telehealth engineer to *all* sessions; and (3) including nurse educators, site coordinators, and peers on the technology assistance team tasked with minor technology problem resolution during the SGVCs, whereas the in-the-classroom telehealth engineer handled both major and minor technical problems encountered by OSMT participants.

OSMT for cancer survivors held via SGVCs, as described herein, involves multiple new ostomates, multiple peers, 1 or 2 nurse educators, and a dedicated telehealth engineer in each training session. The justification for inclusion of the dedicated telehealth engineer was based on our previous experience that disruptions in sessions may have serious negative ramifications for participants. Disruptions of sessions due to technological difficulties can elevate fear and doubt of technology in patients and result in the accumulation of nonproductive time for training group members. Technology issues are of potential concern for multiuser cloud computer-based interventions for cancer support groups [16,17,19-21]. This study was designed to determine the frequency with which telehealth engineer support was needed to address technical failures in OSMT sessions. Over the duration of the OSMT study, no videoconferencing service platform outages disrupted OSMT sessions. There are rare instances when the videoconferencing service platform is known to have experienced an outage or reduction in performance [22]. In this study, participants experienced a range of TTR incidents. Problem areas included establishing sufficient network connectivity for some participants, instructing participants on technology configuration and operations, participants remembering to *mute* or *unmute* their microphones or to enable or disable their device camera, and participants achieving sufficient illumination to allow others on the SGVC to see them.

In anticipation of potential technical incidents affecting patients during SGVC sessions (eg, loss of audio or video camera malfunction), redundancy of personnel was intentionally built into the protocol for both the nurse educator components and the telehealth technology support components of the OSMT sessions. To accomplish this, 2 fully prepared nurse educators were present in the videoconference room used by a pair of nurse educators to lead OSMT groups in 94.1% (96/102) of the sessions. The assigned telehealth engineer was stationed in the videoconference room, rather than being at his usual site of operations, a nearby videoconference control room. Nurse educators became a part of the onsite technology team. They were trained to step in and resolve minor technology issues when the telehealth engineer was occupied with other support tasks. Third, planned redundancy was incorporated into the OSMT curriculum so that each ostomate could have full exposure to the curriculum by attending three out of four training sessions. These measures proved to be effective in minimizing interruptions of OSMT sessions for technical reasons and maximizing OSMT curriculum completion by ostomates.

Currently, rapid changes are occurring in the telemedicine industry [13]. Use of telemedicine (health care services by a physician at a distance) and telehealth (health care services delivered by nonphysicians, eg, nurses, pharmacists, or psychologists) is skyrocketing. Much of this is attributed to the COVID-19 pandemic and the *Stay at Home* mandates of US federal and state governments. It is estimated that there will be close to a billion telehealth cases in the United States in 2020 alone, up from 36 million cases in 2019 [13]. This surge in activity invites more studies on the many facets of telemedicine and telehealth service delivery, bringing quality of services, cost-effectiveness of telehealth, clinical outcomes, user satisfaction, and the applicability of videoconferencing services, into sharper focus [15-17,19].

This study examined questions related to the technological aspects of telehealth-enabled patient training. For example, how can ostomy patients be suitably prepared for support group training sessions? What level of technical support can achieve a zero level of session failure in a cancer survivor group training activity?

The results of this study also provide baseline data on the range of video communication devices used by today's patients enrolled in a community-based OSMT cancer patient–training group setting.

Limitations

One limitation of this study was that recorded TTR incidents comprised free-text comments from a group of individuals with varying levels of technology expertise. Another limitation was that the notes describing TTR incidents generally did not include measures of time spent on either incident resolution or disruptive impact to OSMT sessions. Currently, in a separate study, we are following a more structured protocol with regard to observing and recording TTR incidents.

Conclusions

The delivery of OSMT via SGVCs can enable interactive OSMT for groups of new ostomates from anywhere they have access to broadband internet and sufficient privacy. The technical staffing model and combination of IPPs and IRPs for OSMT described in this paper worked to minimize session disruptions for TTR reasons. Additional research is needed to determine the scalability of OSMT for larger groups across multiple time zones.

Acknowledgments

The authors acknowledge Kristine A Erps and Angelette L Holtrust for assistance in editing the manuscript.

The research reported in this paper was funded by Patient-Centered Outcomes Research Institute award (1507-31690). The statements presented in this paper are solely the responsibility of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute and its Board of Governors or Methodology Committee. Additional funding for this research was provided for Julia Mo by the Benjamin and Mary Siddons Measey Foundation.

Authors' Contributions

Conceptualization was done by RSW, VS, and RSK. Formal analysis was performed by RSW, MJH, JM, CSW, MCH, and RSK. Funding acquisition, methodology, and supervision was carried out by RSW, and RSK. Resources were obtained by RSW, MJH, PY, NJT, and RSK. Visualization was carried out by RSW and MJH. The original draft was written by RSW, MJH, and JM. Reviewing and editing of the manuscript was carried out by RSW, MJH, JM, PY, OB, CSW, NJT, EE, ZC, MCH, VS, and RSK. Data curation was done by MJH, OB, and MG. Investigation was carried out by MJH, PY, OB, and NJT. Validation was performed by MJH. Project Administration was carried out by MJH, NJT, EE, VS, and RM.

Conflicts of Interest

RSW is an uncompensated advisor to Scottsdale, Arizona-based GlobalMed, a telehealth equipment company.

References

1. Ostomy expanded information. American Society of Colon and Rectal Surgeons. URL: https://fascrs.org/patients/diseases-and-conditions/a-z/ostomy-expanded-version [accessed 2020-03-27]
2. What is an ostomy? United Ostomy Associations of America. URL: https://www.ostomy.org/what-is-an-ostomy/ [accessed 2020-03-26]
3. Vonk-Klaassen SM, de Vocht HM, den Ouden ME, Eddes EH, Schuurmans MJ. Ostomy-related problems and their impact on quality of life of colorectal cancer ostomates: a systematic review. Qual Life Res 2016 Jan;25(1):125-133 [FREE Full text] [doi: 10.1007/s11136-015-1050-3] [Medline: 26123983]
4. Shabbir J, Britton DC. Stoma complications: a literature overview. Colorectal Dis 2010 Oct;12(10):958-964. [doi: 10.1111/j.1463-1318.2009.02006.x] [Medline: 19604288]
5. Sun V, Ercolano E, McCorkle R, Grant M, Wendel CS, Tallman NJ, et al. Ostomy telehealth for cancer survivors: design of the Ostomy Self-management Training (OSMT) randomized trial. Contemp Clin Trials 2018 Jan;64:167-172 [FREE Full text] [doi: 10.1016/j.cct.2017.10.008] [Medline: 29051047]
6. Faury S, Koleck M, Foucaud J, M'Bailara K, Quintard B. Patient education interventions for colorectal cancer patients with stoma: a systematic review. Patient Educ Couns 2017 Oct;100(10):1807-1819. [doi: 10.1016/j.pec.2017.05.034] [Medline: 28602564]
7. Sun V, Bojorquez O, Grant M, Wendel CS, Weinstein R, Krouse RS. Cancer survivors' challenges with ostomy appliances and self-management: a qualitative analysis. Support Care Cancer 2020 Apr;28(4):1551-1554. [doi: 10.1007/s00520-019-05156-7] [Medline: 31720825]
8. Krouse RS, Grant M, McCorkle R, Wendel CS, Cobb MD, Tallman NJ, et al. A chronic care ostomy self-management program for cancer survivors. Psychooncology 2016 May;25(5):574-581. [doi: 10.1002/pon.4078] [Medline: 26804708]
9. Ransom S, Azzarello LM, McMillan SC. Methodological issues in the recruitment of cancer pain patients and their caregivers. Res Nurs Health 2006 Jun;29(3):190-198. [doi: 10.1002/nur.20129] [Medline: 16676340]
10. Freda MC. Issues in patient education. J Midwifery Womens Health 2004;49(3):203-209. [doi: 10.1016/j.jmwh.2004.01.003] [Medline: 15134673]

11. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth interventions to support self-management of long-term conditions: a systematic metareview of diabetes, heart failure, asthma, chronic obstructive pulmonary disease, and cancer. J Med Internet Res 2017 May 17;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
12. Bashshur RL, Shannon GW, Smith BR, Alverson DC, Antoniotti N, Barsan WG, et al. The empirical foundations of telemedicine interventions for chronic disease management. Telemed J E Health 2014 Sep;20(9):769-800 [FREE Full text] [doi: 10.1089/tmj.2014.9981] [Medline: 24968105]
13. Barrett J. Case study reveals rapid uptake of telemedicine in health system amid COVID-19. Drug Topics. 2020. URL: https://www.drugtopics.com/view/case-study-reveals-rapid-uptake-telemedicine-health-system-amid-covid-19 [accessed 2020-05-05]
14. Krupinski EA, Weinstein RS. Telemedicine in an academic center - the Arizona Telemedicine Program. Telemed J E Health 2013 May;19(5):349-356. [doi: 10.1089/tmj.2012.0285] [Medline: 23343258]
15. Weinstein RS, Lopez AM, Joseph BA, Erps KA, Holcomb M, Barker GP, et al. Telemedicine, telehealth, and mobile health applications that work: opportunities and barriers. Am J Med 2014 Mar;127(3):183-187. [doi: 10.1016/j.amjmed.2013.09.032] [Medline: 24384059]
16. Foster MV, Sethares KA. Facilitators and barriers to the adoption of telehealth in older adults: an integrative review. Comput Inform Nurs 2014 Nov;32(11):523-535. [doi: 10.1097/CIN.0000000000000105] [Medline: 25251862]
17. Archibald MM, Ambagtsheer RC, Casey MG, Lawless M. Using zoom videoconferencing for qualitative data collection: perceptions and experiences of researchers and participants. Int J Qual Methods 2019 Sep 11;18:1-8. [doi: 10.1177/1609406919874596]
18. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010 Jun 01;152(11):726-732 [FREE Full text] [doi: 10.7326/0003-4819-152-11-201006010-00232] [Medline: 20335313]
19. Baker J, Stanley A. Telemedicine technology: a review of services, equipment, and other aspects. Curr Allergy Asthma Rep 2018 Sep 26;18(11):60. [doi: 10.1007/s11882-018-0814-6] [Medline: 30259201]
20. Gentry MT, Lapid MI, Clark MM, Rummans TA. Evidence for telehealth group-based treatment: a systematic review. J Telemed Telecare 2019 Jul;25(6):327-342. [doi: 10.1177/1357633X18775855] [Medline: 29788807]
21. Doolittle GC, Nelson E, Spaulding AO, Lomenick AF, Krebill HM, Adams NJ, et al. Telehospice: a community-engaged model for utilizing mobile tablets to enhance rural hospice care. Am J Hosp Palliat Care 2019 Sep;36(9):795-800. [doi: 10.1177/1049909119829458] [Medline: 30803247]
22. Narcisi G. Zoom outage tests limits of cloud videoconferencing. CRN News. 2020 Apr 03. URL: https://www.crn.com/news/networking/zoom-outage-tests-limits-of-cloud-videoconferencing [accessed 2020-07-09]

Abbreviations

ATP: Arizona Telemedicine Program
CONSORT: Consolidated Standards of Reporting Trials
FC: family caregiver
IPP: incident prevention plan
IRP: incident response plan
OSMT: Ostomy Self-management Telehealth
SGVC: secure group video call
SP: support person
TTR: telehealth technology–related
UC: usual care
WOCN: wound, ostomy, and continence nurse

Edited by R Kukafka; submitted 21.12.20; peer-reviewed by E Sutsunbuloglu, K Rao; comments to author 09.02.21; revised version received 05.04.21; accepted 06.05.21; published 27.09.21.

Please cite as:
Weinstein RS, Holcomb MJ, Mo J, Yonsetto P, Bojorquez O, Grant M, Wendel CS, Tallman NJ, Ercolano E, Cidav Z, Hornbrook MC, Sun V, McCorkle R, Krouse RS
An Ostomy Self-management Telehealth Intervention for Cancer Survivors: Technology-Related Findings From a Randomized Controlled Trial
J Med Internet Res 2021;23(9):e26545
URL: https://www.jmir.org/2021/9/e26545
doi:10.2196/26545
PMID:34086595

Review

The Effect of Noninvasive Telemonitoring for Chronic Heart Failure on Health Care Utilization: Systematic Review

Stefan L Auener[1], MSc; Toine E P Remers[1], MSc; Simone A van Dulmen[1], PhD; Gert P Westert[1], Prof Dr; Rudolf B Kool[1], MD, PhD; Patrick P T Jeurissen[1], Prof Dr

IQ healthcare, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands

Corresponding Author:
Stefan L Auener, MSc
IQ healthcare
Radboud Institute for Health Sciences
Radboud University Medical Center
Geert Grooteplein Zuid 10
Nijmegen, 6525GA
Netherlands
Phone: 31 243616359
Email: stefan.auener@radboudumc.nl

Abstract

Background: Chronic heart failure accounts for approximately 1%-2% of health care expenditures in most developed countries. These costs are primarily driven by hospitalizations and comorbidities. Telemonitoring has been proposed to reduce the number of hospitalizations and decrease the cost of treatment for patients with heart failure. However, the effects of telemonitoring on health care utilization remain unclear.

Objective: This systematic review aims to study the effect of telemonitoring programs on health care utilization and costs in patients with chronic heart failure. We assess the effect of telemonitoring on hospitalizations, emergency department visits, length of stay, hospital days, nonemergency department visits, and health care costs.

Methods: We searched PubMed, Embase, and Web of Science for randomized controlled trials and nonrandomized studies on noninvasive telemonitoring and health care utilization. We included studies published between January 2010 and August 2020. For each study, we extracted the reported data on the effect of telemonitoring on health care utilization. We used $P<.05$ and CIs not including 1.00 to determine whether the effect was statistically significant.

Results: We included 16 randomized controlled trials and 13 nonrandomized studies. Inclusion criteria, population characteristics, and outcome measures differed among the included studies. Most studies showed no effect of telemonitoring on health care utilization. The number of hospitalizations was significantly reduced in 38% (9/24) of studies, whereas emergency department visits were reduced in 13% (1/8) of studies. An increase in nonemergency department visits (6/9, 67% of studies) was reported. Health care costs showed ambiguous results, with 3 studies reporting an increase in health care costs, 3 studies reporting a reduction, and 4 studies reporting no significant differences. Health care cost reductions were realized through a reduction in hospitalizations, whereas increases were caused by the high costs of the telemonitoring program or increased health care utilization.

Conclusions: Most telemonitoring programs do not show clear effects on health care utilization measures, except for an increase in nonemergency outpatient department visits. This may be an unwarranted side effect rather than a prerequisite for effective telemonitoring. The consequences of telemonitoring on nonemergency outpatient visits should receive more attention from regulators, payers, and providers. This review further demonstrates the high clinical and methodological heterogeneity of telemonitoring programs. This should be taken into account in future meta-analyses aimed at identifying the effective components of telemonitoring programs.

(J Med Internet Res 2021;23(9):e26744) doi:10.2196/26744

KEYWORDS

heart failure; telemonitoring; remote monitoring; health care utilization; eHealth

Introduction

Background

Chronic heart failure (CHF) is one of the most prevalent high-cost chronic diseases affecting at least 1%-2% of the worldwide population [1]. Europe and the United States spent approximately 1%-2% of their national health care budget on this chronic disease [1,2]. Worldwide, the economic burden of CHF is estimated to be approximately US $108 billion per annum, of which US $65 billion can be attributed to direct health care costs [2]. CHF is characterized by an erratic and difficult-to-predict course. A high percentage of incurred costs are due to high readmission rates, as well as the high number of comorbidities [3].

Changes in physiological parameters such as weight, heart rate, blood pressure, and pulse oximetry may precede cardiac events. Signaling such changes through telemonitoring may enable physicians to intervene before the patient needs hospitalization or an emergency department (ED) visit [4]. Telemonitoring has been proposed as a possible strategy to tackle the challenges that CHF brings to health systems, most notably, spare use of the utilization of expensive resources. Telemonitoring has the potential to prevent hospital readmissions, thereby saving costs and improving the quality of life of these patients [5].

The technology and quality of care for patients with CHF have been evolving, which has resulted in a variety of telemonitoring programs consisting of different elements for different populations. In addition, studies on the effect of telemonitoring on health care utilization differ widely according to cointerventions, time horizons, and outcome measures. Such parameters may affect the effectiveness of telemonitoring and the comparability of studies [6]. This heterogeneity at many levels results in debates concerning the effectiveness of telemonitoring.

Studies on telemonitoring have shown mixed results with respect to health care utilization. Although some studies showed a decline, others showed no significant differences or even an increase in health care utilization. Many studies have included cointerventions within the telemonitoring intervention. Structured telephone support (STS) has often been incorporated [7-9]. As STS has been reported to reduce the number of heart failure–related hospitalizations [10], the effects found in studies that use these two interventions simultaneously may not be solely attributed to the telemonitoring program. A Cochrane study from 2015 [10] concluded that, although telemonitoring can reduce heart failure–related admissions, telemonitoring programs were not able to reduce the risk of all-cause hospitalization. A variety of systematic reviews have been performed, mostly exclusively including randomized controlled trials (RCTs). Bashi et al [11] performed an overview of systematic reviews and found that 11 of 19 systematic reviews only included RCTs. Although RCTs are considered as the gold standard for research purposes, observational studies also have merits that should not be ignored. For example, they provide a sense of the real world as opposed to experimental RCT settings [12]. Especially for diseases such as heart failure, which is known for a high degree of multimorbidity [13], RCTs may be limited by their strict inclusion and exclusion criteria. In addition, the vast majority of previous systematic reviews have only analyzed the effect of telemonitoring on hospitalizations and mortality [10,11]. Thus, limited information is available on the effect of telemonitoring programs on other outcomes, such as ED visits, length of stay, and other forms of health care utilization.

Objective

We therefore performed a systematic review on the effect of telemonitoring for CHF on health care utilization, including both RCTs and observational studies. By including a wide variety of telemonitoring programs and outcome measures, we aim to identify the various aspects and broad impact of telemonitoring programs on the health care utilization of patients with CHF.

Methods

Search Strategy

PubMed, Web of Science, and Embase (Ovid) databases were searched. All authors were consulted for additional eligible studies. A detailed description of our search strategy for each database can be found in Multimedia Appendix 1. We excluded articles published before January 1, 2010 because technology and insights have rapidly evolved over the past 10 years. Our search was updated until August 4, 2020 according to the method described by Bramer [14]. Removal of duplications was performed according to the method described by Bramer et al [15].

Eligibility Criteria

Telemonitoring includes a wide variety of definitions and descriptions. Generally, telemonitoring refers to the use of telecommunication to assist in the transmission of medical information and services between health care providers and patients [16]. This definition also includes STS. Many studies do not explicitly distinguish between invasive and noninvasive telemonitoring techniques. However, this review explicitly focuses on noninvasive techniques exclusively because there are significant differences between implanted monitoring devices and noninvasive telemonitoring in terms of costs and eligibility of patients [17]. Thus, we defined telemonitoring as the noninvasive application– or web-based collection and transfer of physiological data, aimed at improving quality of life or decreasing health care utilization in patients with heart failure, or both.

We included peer-reviewed studies that reported direct or indirect measurements of health care utilization such as hospitalizations, ED visits, length of stay, days of hospitalization, visits, and health care costs. We did not apply exclusion parameters concerning study design to include all available evidence. Therefore, we included RCTs, nonrandomized trials, and studies on observational data. However, we excluded studies that were not based on original data such as Markov models because these studies are based on assumptions rather than empirical data. Studies with regular telephone support initiated by nurses or health care providers (without medical indication) were excluded to focus on the

effect of telemonitoring as opposed to the combined effect of telemonitoring and STS. We excluded studies from countries not belonging to the Organization for Economic Co-operation and Development to improve homogeneity in terms of the socioeconomic characteristics of the populations. Finally, studies that were not published in either English or Dutch were included. Textbox 1 presents the inclusion and exclusion criteria.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Telemonitoring equipment within program assessed physiological parameters.
- Telemonitoring measures were shared with at least one health care provider.
- Patients are living independently and are allowed to have access to home care services but should not be admitted to a nursing home during the intervention.
- Subjects have been diagnosed with chronic heart failure.
- Outcomes included direct or indirect measures of health care utilization.
- Paper was a peer-reviewed publication.

Exclusion criteria

- Study was performed in countries not belonging to the Organization for Economic Co-operation and Development.
- Publication date was before January 1, 2010.
- Telemonitoring program included structured telephone support.
- Intervention program used invasive telemonitoring (eg, CardioMEMS).
- Study was not available in the English or Dutch language.
- No quantitative data with accompanied statistical analysis or measure of statistical significance was reported.
- Health care utilization measure was not reported separately (eg, combined end point with death).

Study Selection and Data Extraction

Titles and abstracts were independently screened by 2 researchers (SLA and TEPR) for eligibility. In cases where no conclusive decision could be made, studies were included for full-text screening. Full-text screening was performed in duplicate and independently by SLA and TEPR. Disagreements were resolved by discussion until consensus was achieved or consultation with a coauthor (RBK or SAVD). Thereafter, data were extracted by one researcher (SLA) for both CHF-specific health care utilization and all-cause utilization because a reduction in CHF-specific health care utilization may not translate into an overall reduction [10]. For each study, the effect of telemonitoring on health care utilization was determined by extracting data on hospitalizations, ED visits, length of stay, days of hospitalization, visits, and health care costs. We determined the statistical significance of these results using a cut-off value of $P<.05$. If no P value was reported, we deemed the effect statistically significant if the reported CI did not include 1.00.

Risk of Bias Assessment

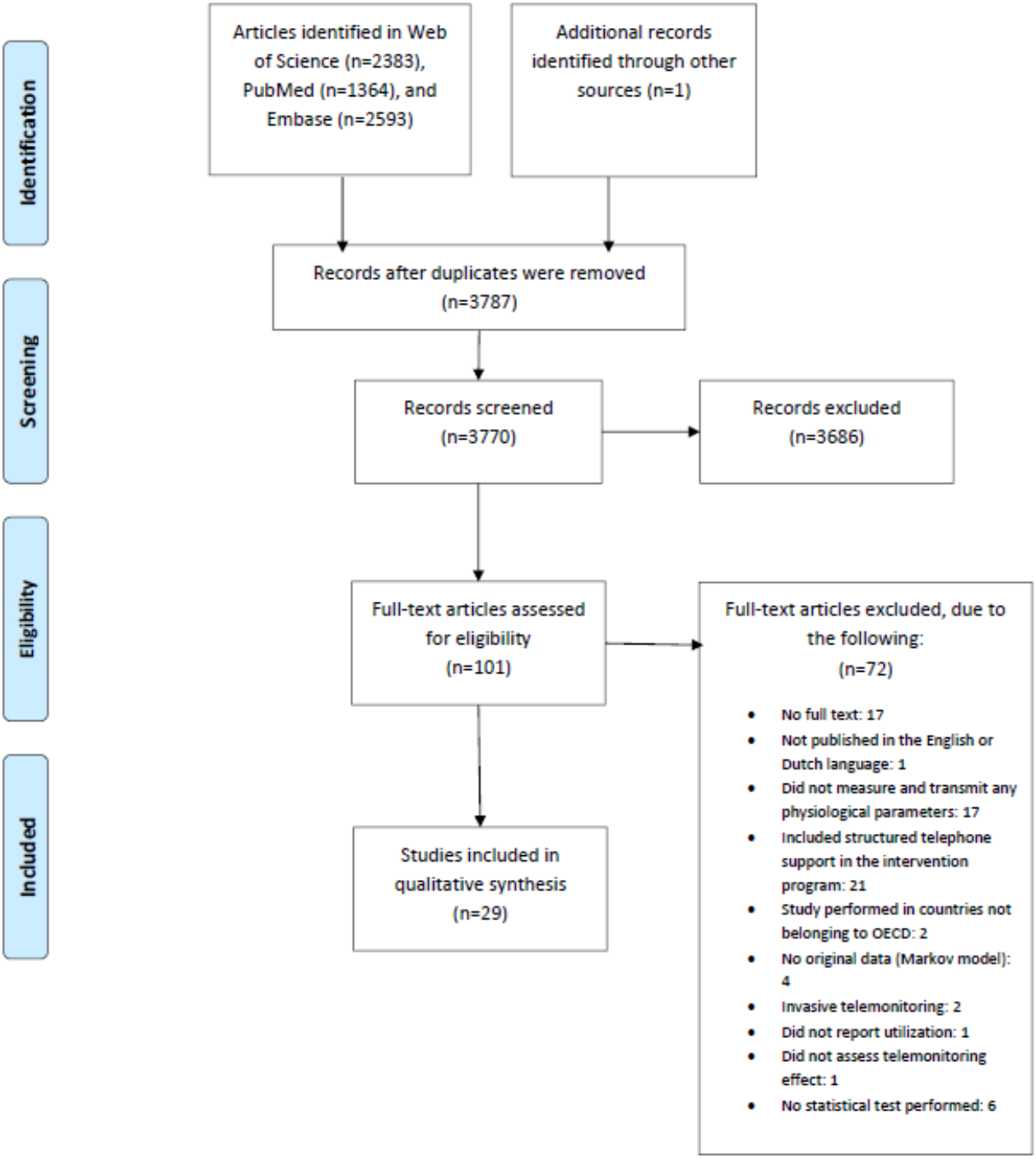
The risk of bias assessment was performed by SLA and TEPR. For the risk of bias assessment, we used the Cochrane risk-of-bias tool for randomized trials [18] and Risk of Bias in Non-randomized Studies of Interventions tools (Cochrane) [19] for RCTs and all remaining studies, respectively. The studies were assessed in duplicate and independently. Subsequently, the individual assessments were compared for possible discrepancies. When discrepancies were observed, the studies and corresponding assessments were discussed until a consensus was reached.

Results

Overview

Our search strategy identified 3770 unique studies, which were then screened in the title and abstract. This resulted in 99 remaining studies. After the full-text screening, 29 studies were included. Figure 1 shows the flow diagram. The main reason for exclusion after full-text screening was because STS was a part of the intervention (19/70, 27%), followed by a lack of measuring and transmitting physiological parameters in the program (17/70, 24%). Most studies (11/29, 38%) originated in the United States.

Figure 1. PRISMA (Preferred Reported Items for Systematic Reviews and Meta-Analyses) flowchart. OECD: Organization for Economic Co-operation and Development.



Characteristics of Included Studies

Table 1 presents an overview of study and program characteristics. A detailed description of the studies and their characteristics are presented in Table 2. Telemonitoring programs showed a high degree of heterogeneity in terms of physiological variables measured as shown in Multimedia Appendix 2 [4,20-47]. All studies included weight as a parameter, whereas only 4 included electrocardiography measures as a physiological parameter [20-23]. The studies used 10 unique combinations of physiological parameters. In addition, 2 studies included other measures such as lung fluid [24] and body composition [25], which were acquired by noninvasive means. Most programs (n=16) used telemonitoring in addition to usual care; regular check-up visits were still scheduled [4,22,24,26-38]. One study [20] stated that telemonitoring was used as a substitute for these visits, and 12 studies did not elaborate on this matter. The follow-up period of the studies ranged between 1 and 89 months, with most studies having a follow-up period of 6 months. Follow-up was consistent with the duration of the telemonitoring program itself, with the exception of 2 studies that also included a follow-up after the telemonitoring program was completed [24,39].

Table 1. Overview of the characteristics of included studies (N=29).

Study characteristics	Studies, n (%)
Country	
Belgium	2 (7)
Canada	2 (7)
Denmark	1 (3)
Finland	1 (3)
Israel	2 (7)
Italy	3 (10)
Japan	1 (3)
Netherlands	3 (10)
Spain	1 (3)
Sweden	1 (3)
United Kingdom	1 (3)
United States	11 (38)
Physiological parameter assessed	
Weight	29 (100)
Blood pressure	23 (79)
Pulse oximetry	7 (24)
Heart rate	20 (69)
Symptom questions	14 (48)
ECG[a]	4 (14)
Number of patients enrolled in telemonitoring program	
<50	8 (28)
50-100	9 (31)
100-200	9 (31)
>200	3 (10)
Follow-up period (months)	
1	1 (3)
2	1 (3)
3	4 (14)
6	10 (34)
9	1 (3)
12	8 (28)
>12	4 (14)

[a]ECG: electrocardiography.

Table 2. Detailed characteristics of the included studies (N=29).

Study	Study characteristics						
	Country	Study design	Patient inclusion criteria	NYHA[a] class	Follow-up period (months)	Intervention group (n)	Control group, n
Amir [24]	United States	Prospective	Stage C CHF[b] with hospital admission for acute HF[c]	—[d]	6	50	N/A[e]
Bakhshi [42]	Israel	Comparison of group with telemonitoring versus group without telemonitoring	Primary diagnosis CHF index hospitalization or patients seen by emergency room, internal medicine, and cardiology specialists	—	6	44	16
Delaney [43]	United States	RCT[f]	Recently discharged from home care, eligible, and previously hospitalized for HF	III: 95 (95%)	3	50	50
Dendale [44]	Belgium	RCT	Hospitalized for fluid overload due to HF	Mean NYHA class: 3.0	6	80	80
Domingo [26]	Spain	RCT	Outpatients attended in the HF unit with NYHA classes II-IV	II: 75 (82%) and III: 17 (18%)	12	44	48
Eilat-Tsanani [45]	Israel	Before-after	NYHA classes II-IV with at least 3 admissions within a 6-month period	II:74 (52.5%); III: 55 (39%); and IV: 12 (8.5%)	12	141	N/A
Frederix [39]	Belgium	RCT	Patients participated in the TEMA-HF[g] study by Dendale et al [44]	Mean NYHA class: 3.0	89	80	80
Hoban [27]	United States	RCT	HF diagnosis care from home health care agency	—	3	40	40
Kotooka [25]	Japan	RCT	Discharge following admission for acute HF or decompensated chronic HF	II: 142 (78.5%); III: 39 (21.5%)	31	90	91
Koulaouzidis [28]	United Kingdom	Retrospective comparison with patients that declined	Newly diagnosed (outpatient) patients with HF	II: 238 (52.5%) and III: 215 (47.5%)	12	124	329
Kraai [20]	Netherlands	Multicenter RCT	Admitted to intensive care or cardiology ward or visited outpatient HF- clinic and in need of treatment	II:39 (22%); III:100 (58.5%) and IV:33 (18.6%)	9	94	83
Lyngå [29]	Sweden	RCT	Hospitalized with NYHA classes III-IV	III: 309 (96.9%); IV: 10 (3.1%)	12	166	153
Maeng [30]	United States	Retrospective within patients	Have a diagnosis of CHF	—	70	541	N/A
Olivari [21]	Italy	RCT	Discharge from hospital after acute HF within the past 3 months	II: 163 (48.1%); III: 159 (46.9%); IV: 17 (5%)	12	229	119
Park [40]	United States	Prospective comparison between hospital and national readmission rates	Admitted for a diagnosis of acute HF	—	1	58	N/A

Study	Study characteristics						
	Country	Study design	Patient inclusion criteria	NYHA[a] class	Follow-up period (months)	Intervention group (n)	Control group, n
Pedone [31]	Italy	RCT	Diagnosis of HF and aged >64 years	II: 29 (32%); III: 51 (57%); IV: 10 (11%)	6	50	46
Riley [41]	United States	Prospective comparison with matched controls	HF admission diagnosis in the electronic health record	—	6	45	45
Seto [22]	Canada	RCT	Ambulatory patients diagnosed with HF	II: 43 (43%); II-III: 11 (11%); III: 42 (42%); IV: 4 (4%)	6	50	50
Soran [32]	United States	RCT	Diagnosis of HF in Medicare data and hospitalized for HF within 6 months	II: 183 (58.1%); III: 132 (41.9%)	6	160	155
Tompkins [33]	United States	RCT	Diagnosis of HF	—	6	193	197
Van der Burg [34]	Netherlands	Retrospective before-after study	NYHA classes III-IV	—	36	177	N/A
Veenstra [35]	Netherlands	Prospective prestudy and poststudy without controls	Discharged after admission for HF or outpatient visit and NYHA classes II-IV	II: 28 (27.5%); III: 57 (55.9%); IV: 17 (16.7%)	12	102	N/A
Vestergaard [46]	Denmark	RCT with economic analysis	Diagnosed with HF according to national guidelines and NYHA classes II-IV	Missing:13 (4.7%); II: 147 (53.6%); III: 92 (33.6%); IV: 22 (8%)	12	134	140
Villani [23]	Italy	RCT	NYHA classes III or IV during hospital stay and high risk of early rehospitalization at discharge	Mean NYHA: 3.01	12	40	40
Vuorinen [47]	Finland	RCT	Left ventricular ejection fraction<35%, NYHA ≥2, and needed regular follow-up	II: 36 (38%); III: 55 (59%); IV: 3 (3%)	6	47	47
Ware [36]	Canada	Before-after study	Diagnosed with HF and visited the HF clinic	≤II: 143 (45.4%)	6	315	N/A
White-Williams [4]	United States	Retrospective study with controls	Discharged from hospital stay and admitted to a home health care agency	—	3	29	17
Williams [37]	United States	Retrospective study with matched controls	Medicare eligible, had a diagnosis of HF, and experienced a recent (2 days) discharge from hospital	—	2	105	105

Study	Study characteristics						
	Country	Study design	Patient inclusion criteria	NYHA[a] class	Follow-up period (months)	Intervention group (n)	Control group, n
Zan [38]	United States	Prospective study with matched controls	Participants recruited from the outpatient clinic. Patients admitted to hospital program were excluded	Intervention group: I: 5 (24%); II: 9 (43%); III: 7 (33%)	3	21	20

[a]NYHA: New York Heart Association.

[b]CHF: chronic heart failure.

[c]HF: heart failure.

[d]Data not reported.

[e]N/A: not applicable.

[f]RCT: randomized controlled trial.

[g]TEMA-HF: Telemonitoring in the Management of Heart Failure.

Patient inclusion criteria for previous hospitalizations differed among the studies. If a hospital admission was required for inclusion, this was most often (n=10) within 7 days of discharge [4,23,24,29,35,37,39-42]. Inclusion was started within 1 month of index hospitalization for 2 studies [20,25], and 5 studies [21,32,43-45] included patients within 6 months. Most studies (n=12) [22,26-28,30,31,33,34,36,38,46,47] did not specify requirements regarding previous admissions for CHF or included patients from outpatient clinics. New York Heart Association (NYHA) Class II was most often reported (n=9) as the NYHA Classification with most patients [21,22,25,26,28,32,38,45,46]. NYHA Class III was reported in 8 studies as the dominant class [20,23,29,31,35,43,44,47].

Risk of Bias Assessment

We found that most RCTs (12/16, 75%) showed at least some bias concerns, with 2 RCTs having a high risk of bias [21,27]. The main reasons for *some concerns* in the RCTs were related to the randomization process. In addition, 75% (9/12) of nonrandomized studies were assessed as having a *serious* or *critical* risk of bias. No nonrandomized studies were assessed as having a *low risk* of bias, which indicates that the strength of evidence from none of the included nonrandomized studies is equivalent to a well-performed RCT [19]. *Bias due to confounding* was the domain that most often resulted in studies being assessed with a *serious* or *critical* risk for bias. Table 3 shows the consolidated overall bias assessments. A detailed overview of the risk of bias assessment of all studies can be found in Multimedia Appendix 3 [20-23,25-27,29,31-33,39,43,44,46,47] and Multimedia Appendix 4 [4,24,28,30,34-38,40-42,45].

Table 3. Summary of risk of bias assessment for randomized controlled trials and nonrandomized studies.

Study	Result of bias risk assessment
Randomized controlled trials[a]	
Delaney (2013) [43]	Some concerns
Dendale (2014) [44]	Some concerns
Domingo (2011) [26]	Some concerns
Hoban (2013) [27]	High risk
Kotooka (2018) [25]	Low risk
Kraai (2016) [20]	Low risk
Lyngå (2012) [29]	Some concerns
Frederix (2018) [39]	Low risk
Olivari (2018) [21]	High risk
Pedone (2015) [31]	Some concerns
Seto (2012) [22]	Some concerns
Soran (2010) [32]	Some concerns
Tompkins (2012) [33]	Some concerns
Vestergaard (2020) [46]	Low risk
Villani (2014) [23]	Some concerns
Vuorinen (2014) [47]	Some concerns
Nonrandomized studies[b]	
Amir (2017) [24]	Moderate risk
Bakhshi (2011) [42]	Serious risk
Eilat-Tsanani (2015) [45]	Critical risk
Koulaouzidis (2019) [28]	Serious risk
Maeng (2014) [30]	Critical risk
Park (2019) [40]	Critical risk
Riley (2015) [41]	Moderate risk
Van der Burg (2020) [34]	Serious risk
Veenstra (2015) [35]	Critical risk
Ware (2020) [36]	Serious risk
White-Williams (2015) [4]	Serious risk
Williams (2016) [37]	Moderate risk
Zan (2015) [38]	Serious risk

[a]Assessed using Cochrane risk-of-bias tool for randomized trials.

[b]Assessed using Risk of Bias in Non-randomized Studies of Interventions tool.

Effects of Telemonitoring on Health Care Utilization

Overview

Telemonitoring studies used many different outcome measures, the most common of which were hospitalizations, ED visits, and health care costs. Whereas some studies report the effects of telemonitoring on specific CHF outcomes, for example, heart failure readmissions, other studies only report the effect on all-cause health care utilization. The type of analysis and effect measures differed widely among studies, with hazard ratios, odds ratios, 2-tailed *t* tests, and incidence rates being reported. Table 4 shows the (statistically significant) effects of telemonitoring on health care utilization.

Table 4. Effect of telemonitoring programs on health care utilization.

Study	Outcome measure											
	Hospitalization		Number of emergency department visits		Length of stay		Days of hospitalization		Visits[a]		Health care costs	
	HF[b]	AC[c]	HF	AC	HF	AC	HF	AC	HF	AC	HF	AC
Amir (2017) [24]	—[d]	+[e]	—	—	—	—	—	—	—	—	—	—
Bakhshi (2011) [42]	—	—	—	—	—	=[f]	—	—	—	—	—	—
Delaney (2013) [43]	=	+	—	—	—	—	—	—	—	—	—	—
Dendale (2014) [44]	=	=	—	—	—	—	—	=	—	—	—	=
Domingo (2011)[g] [26]	=	—	—	—	—	—	=	—	×[h]	—	—	—
Eilat-Tsanani (2015) [45]	=	+	—	=	—	—	—	+	—	=	—	—
Frederix (2018) [39]	—	—	—	—	—	—	+	=	—	—	=	=
Hoban (2013) [27]	—	=	—	—	—	—	—	—	—	—	—	—
Kotooka (2018) [25]	=	=	—	—	—	—	—	—	—	—	—	—
Koulaouzidis (2019) [28]	=	=	—	—	—	—	+	=	—	—	—	—
Kraai (2016) [20]	=	=	—	—	—	—	—	—	+	=	—	=
Lyngå (2012) [29]	=	=	—	—	—	—	=	—	—	—	—	—
Maeng (2014) [30]	—	+	—	—	—	—	—	—	—	—	—	+
Olivari (2018)[i] [21]	=	=	—	=	=	=	—	—	—	—	—	—
Park (2019) [40]	—	=	—	—	—	—	—	—	—	—	—	—
Pedone (2015) [31]	=	+	—	—	—	—	—	—	—	—	—	—
Riley (2015)[j] [41]	—	=	—	—	—	—	—	=	—	—	—	—
Seto (2012) [22]	—	=	—	=	—	—	—	=	×	—	—	—
Soran (2010) [32]	—	—	—	—	—	—	—	—	—	—	—	×
Tompkins (2012) [33]	—	=	—	=	—	=	—	=	—	×	—	=
Van der Burg (2020) [34]	—	+	—	—	—	—	—	+	—	—	—	+
Veenstra (2015) [35]	+	—	—	—	—	—	—	—	—	—	—	—
Vestergaard (2020) [46]	—	—	—	—	—	—	—	—	—	—	—	+
Villani (2014) [23]	+	—	—	+	—	—	—	—	—	×	—	×
Vuorinen (2014) [47]	—	—	—	—	—	—	=	—	×	×	—	—
Ware (2020) [36]	+	+	=	=	=	=	—	—	=	—	—	—
White-Williams (2015) [4]	—	=	—	=	—	—	—	—	—	—	—	—
Williams (2016) [37]	—	=	—	—	—	—	—	—	—	×	—	×
Zan (2015) [38]	—	=	—	=	—	—	—	—	—	—	—	—

[a]Visits include all nonemergency department outpatient visits or consultations with health care provider contacts.

[b]HF: heart failure–related.

[c]AC: all-cause.

[d]Data not reported.

[e]Indicates statistically significant (P<.05) reduction in outcome measure.

[f]Indicates not statistically significant (P>.05) on outcome measure.

[g]Analysis based on the comparison between Motiva and Motiva plus.

[h]Indicates statistically significant (P<.05) increase in outcome measure.

[i]Analysis based on intention-to-treat analysis.

[j]Analysis based on matched-cohort analysis.

Hospitalization

Hospitalization was the most commonly reported outcome (24/29, 83% of studies). Out of 24 studies, 9 (38%) showed a statistically significant reduction in hospitalizations associated with telemonitoring. The remaining studies (15/24, 63%) found no effect on hospitalization. When excluding all studies with a high, critical, or serious risk of bias (n=11), 31% (4/13) of studies reported a statistically significant reduction in hospitalizations [23,24,31,43]. Although the number of studies reporting a statistically significant effect was limited, the effects found in these studies were clinically relevant. Amir et al [24] reported a temporary reduction of hospitalizations of 87% during the intervention. However, hospitalization increased by 79% when the intervention ended. Delaney et al [43] found that hospitalization rates were 19% and 38% in the intervention and control groups, respectively. Pedone et al [31] reported an incidence rate of 0.30 (95% CI 0.12-0.67), favoring the intervention group. Villani et al [23] found that the control group had almost twice the number of hospitalizations compared with the intervention group. We did not find a parameter in the study characteristics that consistently differed from the other studies that showed no effect of telemonitoring.

ED Visits

In total, 8 studies reported results related to the number of ED visits [4,21-23,33,36,38,45]; 7 (88%) reported no statistically significant effects of telemonitoring and 1 (13%) found a significant effect, reporting 17 ED visits in the control group (n=40) compared with 6 ED visits in the intervention group (n=40) [23].

Length of Stay

The outcome measure length of stay was reported in 4 studies [21,33,36,42]. None of the included studies found a statistically significant effect of telemonitoring on the length of stay.

Days of Hospitalization

Days of hospitalization were reported in 10 studies, of which 8 (80%) reported on all-cause days of hospitalization and 5 (50%) on days of hospitalization specifically for heart failure. There were 2 studies that showed a reduction in all-cause days of hospitalization [34,45]. The 2 studies showing a decrease in heart failure–related days of hospitalization did not result in a statistically significant decrease in all-cause days of hospitalization [28,39].

Visits

A total of 9 studies reported outcome measures related to non-ED health care visits. These visits comprised outpatient visits [20,22,23,26,36,45,47], primary care visits [33,45], home visits by nurse [37,45], urgent care visits [33], and telephone contacts [47]. Most (6/9, 67%) of these studies found that their telemonitoring intervention was associated with a significant increase in non-ED health care visits [22,23,26,33,37,47]. A total of 2 studies found no difference [36,45], and 1 study [20] found a decrease in non-ED visits for heart failure specifically. However, this reduction of heart failure rates did not result in a statistically significant decrease in the number of visits. Two studies were considered as having a critical [45] or serious risk of bias [36]. Both of these studies reported no effect of telemonitoring on non-ED visits.

Health Care Costs

A comparison of health care costs between the intervention and control groups was performed in 10 studies. Out of these 10, 3 (30%) studies showed a decrease in costs [30,34,46], 4 (40%) studies showed no statistically significant difference [20,33,39,44], and 3 (30%) studies showed a negative impact of telemonitoring on health care costs [23,32,37]. Two studies, which we assessed as having a serious or critical risk of bias, reported statistically significant decreases in health care costs [30,34]. When a decrease in costs was realized by the telemonitoring program, this was mainly due to a reduction in hospitalizations or rehospitalizations [30,34]. In addition to a reduction in hospitalizations, Vestergaard et al [46] found a reduction in outpatient care costs in the telemonitoring group. Furthermore, 3 studies finding a reduction of health care costs reported 11% [30], 35% [46], and 90% [34] reductions in health care costs. When telemonitoring was associated with an increase in costs, this was due to an increase in outpatient visits [37], an overall increase in health care utilization [32], or the costs of the telemonitoring program itself [23].

Six studies included and explicitly mentioned the costs of (the development of) the intervention in their analyses [20,23,30,32,37,46]. These studies reported large differences in the costs of the intervention, mostly due to cointerventions and specific assumptions regarding program and development costs. Multimedia Appendix 5 [20,23,30,32,37,46] provides a description of the costs. In the study by Villani et al [23], a reduction in hospitalization costs could not offset the additional costs of the intervention. This indicates a substantial influence of program costs in determining the cost-effectiveness of telemonitoring programs. An important factor for cost differences was the high cost of developing a telemonitoring program for a selected population.

Discussion

Principal Findings

The goal of this review was to identify the broad impact and various aspects of telemonitoring programs for patients with CHF and their effects on health care utilization. Most studies showed no statistically significant effects of telemonitoring on overall hospitalizations (14/21, 67%), ED visits (7/8, 88%), length of stay (4/4, 100%), and days of hospitalization (6/8, 75%). However, the remaining studies showed reductions in health care utilization for these measures. Overall, non-ED outpatient visits and health care costs were increased in 67% (6/9) of studies and 30% (3/10) of studies, respectively. The most ambiguously reported health care utilization measure was health care costs, which showed increases in 30% (3/10) of the studies, decreases in 30% (3/10) of the studies, and no differences in 40% (4/10) of the studies. Heart failure–specific health care utilization measures showed similar results as the overall health care utilization measures.

We found a high degree of clinical diversity among the interventions, in terms of physiological parameters as well as

the targeted populations. Although most telemonitoring programs were targeted at patients in NYHA Classes II-III, a variety of additional inclusion criteria were used. In addition to clinical diversity, methodological heterogeneity was also high, as can be observed by the varying risk of bias assessments.

Some of the studies included in this review suggested that telemonitoring may result in lower hospital admission rates, but most studies did not report such reductions. This mixed effect is consistent with the current literature [10,48]. In addition, we found no clear associations between the extracted study characteristics and the effect of telemonitoring and a reduction in hospitalizations. However, we found that 67% (6/9) of studies reported an increase in non-ED outpatient visits when telemonitoring was used.

Thus far, the effect of telemonitoring on non-ED visits has received little attention. However, the early warning designs of telemonitoring might also attract additional use of frontline services and false-positive alarms. An overview of systematic reviews [11] showed that 2 of 19 systematic reviews included non-ED visits in their outcomes [49,50]. These systematic reviews included 5 studies, 1 [51] of which can be classified as telemonitoring based on our definition. This RCT showed a substantial increase in non-ED visits and phone calls for telemonitoring compared with usual care. Our results suggest that this may be the case for most telemonitoring programs, as we found that 67% (6/9) of studies showed an increase in non-ED visits. One study found a decrease in heart failure–related non-ED visits. This study [20] explicitly stated that "patients allocated to the intervention group were only allowed to visit the cardiologist or HF-nurse in case of an absolute need for intervention." In addition, 83% (5/6) of studies, which found an increase in visits, had telemonitoring as an additional component in addition to usual care [22,26,37,47], and 17% (1/6) did not state this explicitly [33]. This stresses the need to treat telemonitoring as a substitute for regular care, rather than an addition to regular care, that is, if a reduction in health care utilization is the primary aim. Five studies reported an increase in non-ED visits and hospitalizations as outcome measures. Villani et al [23] found an increase in non-ED visits simultaneously with a reduction in hospitalizations. The other 4 studies [22,26,33,37] reported an increase in non-ED visits and found no effect on hospitalization. This indicates that additional visits can be a side effect of telemonitoring rather than a prerequisite for effective telemonitoring programs in terms of reduced health care utilization. Therefore, telemonitoring programs may become more cost-effective if they pay more attention to decreasing these visits.

The effect of telemonitoring on health care costs has been inconsistent across studies. Health care costs were severely affected by the costs of telemonitoring programs, which showed large differences. These differences were attributable to both assumptions regarding the development costs per patient and the actual cost differences. These cost differences can have a detrimental effect on the cost-effectiveness and financial viability of the program. No studies included indirect cost savings or expenses from a patient perspective, such as a reduction or increase in travel costs.

Strength and Limitations

A major strength of our study is the inclusion of various study designs and outcome measures for health care utilization. Therefore, this review offers a broader scope than other reviews [11]. The inherent weakness of including nonrandomized studies was the introduction of biases because of case mixes. This was demonstrated by the results of the Risk of Bias in Non-randomized Studies of Interventions tool, which showed a high degree of bias in the domain of baseline confounding. The exclusion of studies also using STS within their telemedicine program resulted in the exclusion of a significant number of studies. However, the exclusion of these studies also increases the relevance of our study for those interested in the stand-alone effect of telemonitoring on health care utilization.

The main limitation of this review is the lack of a meta-analysis and thus a limited ability to draw strong conclusions regarding the effect of telemonitoring on health care utilization. However, this review clearly shows why meta-analyses are difficult to perform on this subject and, if performed, should be interpreted with caution. We observed high heterogeneity within the study populations, telemonitoring programs, and outcomes. Therefore, a meta-analysis was not appropriate for this study. In addition, the results of meta-analyses of other studies may only be applicable to small subsets of populations, interventions, and outcomes and thus may not represent the true effect of telemonitoring on health care utilization because of limited external validity.

Implications for Practice

This study has several practical implications. The finding that telemonitoring often increases non-ED visits has consequences for the workload of outpatient clinics. These additional non-ED visits may occur because of a variety of reasons such as false positives, true positives, equipment malfunction, and whether telemonitoring is used as additional or substitute care. Detecting true positives is the primary aim, as it prevents more expensive health care utilization and improves the quality of life. Although false-positive alerts and equipment malfunctions may be reduced by improving technology and algorithms [52], our results suggest that addition instead of substitution is likely to remain, resulting in additional non-ED visits. Health care providers must be aware of this increase and adopt organizational structure of the outpatient clinic or find other ways to mitigate this increase, such as using telemonitoring as a substitute rather than additional care [20] or outsourcing technical difficulties experienced by patients to medical service centers [34].

This review showed that telemonitoring might shift health care utilization to outpatient settings, as opposed to only reducing inpatient admissions. This may complicate adoption, as the benefits may not be attributable to the same stakeholder as the costs. This is especially the case for health care systems where outpatient care is delivered by organizations other than inpatient care, such as that of Germany. In such cases, conflicting interests are to be expected, and health care payers have to come to an agreement with health care providers to overcome these issues [53]. Health care regulators can facilitate this process by creating and supporting new payment models such as shared savings and lump-sum payment models.

Finally, the high costs of developing telemonitoring programs for a selected population can diminish and even cancel future monetary gains of reduced health care consumption [23]. One way of suppressing the costs of telemonitoring can be achieved through the use and development on a larger scale. Villani et al [23] showed that developing a telemonitoring system for 40 patients was not financially viable. In contrast, Vestergaard et al [46] reduced the costs of the program per patient by targeting a larger population, namely a whole region of North Denmark constituting 6700 patients. Health care providers should either use the scale realized by a third party or codevelop telemonitoring systems with other health care providers to realize possible scale advantages. In addition, in general, substitutive programs are expected to achieve higher financial savings.

Future Research

Future studies should consider clinical diversity by including subgroup meta-analyses or performing meta-regression, as opposed to pooled meta-analyses [54]. As mentioned previously, we found a high degree of clinical and methodological diversity [54]. Despite this, some meta-analyses [48,55] have found a low to moderate effect for certain health care utilization measures. The presence of high clinical diversity and low statistical heterogeneity may be due to minimal marginal effects of components in the telemonitoring program; a part of the intervention did not affect the outcome measure. If this is the case, certain (combinations of) telemonitoring components do not add value to the intervention. Yun et al [56] and Kotb et al [57] performed such analyses on all-cause mortality and hospitalization in patients with CHF. Certain program characteristics, such as having 3 or more physiological parameters [56] or including an electrocardiograph [57], were statistically associated with a reduction in all-cause mortality and hospitalization, respectively. Similarly, essential and effective parts of telemonitoring for reducing other health care utilization can be identified, thereby supporting the development of (cost-) effective telemonitoring programs. Clear descriptions of the intervention and context are needed to perform such analyses. This review can be used as guidance for forming subgroups or variables of interest for meta-regressions. As there are many more factors that may affect the impact of telemonitoring on health care utilization, qualitative research may be used to develop hypotheses and guide meta-regression protocols.

Conclusions

This review investigated the effects of telemonitoring programs on different aspects of health care utilization. Telemonitoring has the potential to reduce hospitalization rates. However, this was not achieved in most studies, as the number of non-ED visits increased in the majority of studies. The effect of telemonitoring on health care costs is highly ambiguous and depends on the effectiveness of the intervention in reducing health care utilization as well as on the costs of the telemonitoring program itself. Health care providers and payers should be aware that the majority of current telemonitoring programs do not result in a reduction in health care utilization and may even increase health care utilization by increasing the number of non-ED visits. Possible payer strategies should be focused at increasing the scale to reduce program costs and implement telemonitoring as a substitute to reduce possible increases in outpatient visits. Nevertheless, more focus is needed to determine the essential factors of telemonitoring programs that reduce health care utilization.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search queries for PubMed, Web of Science, and Embase.
[PDF File (Adobe PDF File), 67 KB - jmir_v23i9e26744_app1.pdf]

Multimedia Appendix 2
Detailed characteristics of telemonitoring programs.
[PDF File (Adobe PDF File), 62 KB - jmir_v23i9e26744_app2.pdf]

Multimedia Appendix 3
Risk of bias assessment: Cochrane Risk of Bias tool for randomized trials.
[PDF File (Adobe PDF File), 55 KB - jmir_v23i9e26744_app3.pdf]

Multimedia Appendix 4
Risk of bias assessment: Risk of Bias in Non-randomized Studies of Interventions tool for non-randomized studies.
[PDF File (Adobe PDF File), 117 KB - jmir_v23i9e26744_app4.pdf]

Multimedia Appendix 5
Costs of telemonitoring programs.
[PDF File (Adobe PDF File), 62 KB - jmir_v23i9e26744_app5.pdf]

References

1. Berry C, Murdoch DR, McMurray JJ. Economics of chronic heart failure. Eur J Heart Fail 2001 Jun;3(3):283-291. [Medline: 11377998]
2. Cook C, Cole G, Asaria P, Jabbour R, Francis DP. The annual global economic burden of heart failure. Int J Cardiol 2014 Feb 15;171(3):368-376. [doi: 10.1016/j.ijcard.2013.12.028] [Medline: 24398230]
3. Wammes JJ, Auener S, van der Wees PJ, Tanke MA, Bellersen L, Westert GP, et al. Characteristics and health care utilization among patients with chronic heart failure: a longitudinal claim database analysis. ESC Heart Fail 2019 Dec;6(6):1243-1251 [FREE Full text] [doi: 10.1002/ehf2.12512] [Medline: 31556246]
4. White-Williams C, Unruh L, Ward K. Hospital utilization after a telemonitoring program: a pilot study. Home Health Care Serv Q 2015;34(1):1-13. [doi: 10.1080/01621424.2014.995256] [Medline: 25517540]
5. Gensini GF, Alderighi C, Rasoini R, Mazzanti M, Casolo G. Value of telemonitoring and telemedicine in heart failure management. Card Fail Rev 2017 Nov;3(2):116-121 [FREE Full text] [doi: 10.15420/cfr.2017:6:2] [Medline: 29387464]
6. Chaudhry SI, Phillips CO, Stewart SS, Riegel B, Mattera JA, Jerant AF, et al. Telemonitoring for patients with chronic heart failure: a systematic review. J Card Fail 2007 Feb;13(1):56-62 [FREE Full text] [doi: 10.1016/j.cardfail.2006.09.001] [Medline: 17339004]
7. Kulshreshtha A, Kvedar JC, Goyal A, Halpern EF, Watson AJ. Use of remote monitoring to improve outcomes in patients with heart failure: a pilot trial. Int J Telemed Appl 2010;2010:870959 [FREE Full text] [doi: 10.1155/2010/870959] [Medline: 20508741]
8. Burdese E, Testa M, Raucci P, Ferreri C, Giovannini G, Lombardo E, et al. Usefulness of a telemedicine program in refractory older congestive heart failure patients. Diseases 2018 Jan 20;6(1):10 [FREE Full text] [doi: 10.3390/diseases6010010] [Medline: 29361704]
9. Martín-Lesende I, Orruño E, Cairo C, Bilbao A, Asua J, Romo MI, et al. Assessment of a primary care-based telemonitoring intervention for home care patients with heart failure and chronic lung disease. The TELBIL study. BMC Health Serv Res 2011 Mar 08;11:56 [FREE Full text] [doi: 10.1186/1472-6963-11-56] [Medline: 21385401]
10. Inglis SC, Clark RA, Dierckx R, Prieto-Merino D, Cleland JGF. Structured telephone support or non-invasive telemonitoring for patients with heart failure. Cochrane Database Syst Rev 2015(10):CD007228. [doi: 10.1002/14651858.CD007228.pub3] [Medline: 26517969]
11. Bashi N, Karunanithi M, Fatehi F, Ding H, Walters D. Remote monitoring of patients with heart failure: an overview of systematic reviews. J Med Internet Res 2017 Jan 20;19(1):e18 [FREE Full text] [doi: 10.2196/jmir.6571] [Medline: 28108430]
12. Ligthelm RJ, Borzì V, Gumprecht J, Kawamori R, Wenying Y, Valensi P. Importance of observational studies in clinical practice. Clin Ther 2007 Jun;29(6 Pt 1):1284-1292. [doi: 10.1016/j.clinthera.2007.07.004] [Medline: 18036390]
13. Sinnige J, Korevaar JC, Westert GP, Spreeuwenberg P, Schellevis FG, Braspenning JC. Multimorbidity patterns in a primary care population aged 55 years and over. Fam Pract 2015 Oct;32(5):505-513 [FREE Full text] [doi: 10.1093/fampra/cmv037] [Medline: 26040310]
14. Bramer W, Bain P. Updating search strategies for systematic reviews using EndNote. J Med Libr Assoc 2017 Jul;105(3):285-289 [FREE Full text] [doi: 10.5195/jmla.2017.183] [Medline: 28670219]
15. Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. De-duplication of database search results for systematic reviews in EndNote. J Med Libr Assoc 2016 Jul;104(3):240-243 [FREE Full text] [doi: 10.3163/1536-5050.104.3.014] [Medline: 27366130]
16. Field MJ. Telemedicine: a guide to assessing telecommunications in healthcare. J Digit Imaging 1997 Aug;10(3 Suppl 1):28 [FREE Full text] [doi: 10.1007/BF03168648] [Medline: 9268830]
17. Faragli A, Abawi D, Quinn C, Cvetkovic M, Schlabs T, Tahirovic E, et al. The role of non-invasive devices for the telemonitoring of heart failure patients. Heart Fail Rev 2020 Apr 27:1063-1080. [doi: 10.1007/s10741-020-09963-7] [Medline: 32338334]
18. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Br Med J 2011;343:d5928 [FREE Full text] [Medline: 22008217]
19. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. Br Med J 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
20. Kraai I, de Vries A, Vermeulen K, van Deursen V, van der Wal M, de Jong R, et al. The value of telemonitoring and ICT-guided disease management in heart failure: Results from the IN TOUCH study. Int J Med Inform 2016 Jan;85(1):53-60 [FREE Full text] [doi: 10.1016/j.ijmedinf.2015.10.001] [Medline: 26514079]
21. Olivari Z, Giacomelli S, Gubian L, Mancin S, Visentin E, Di Francesco V, et al. The effectiveness of remote monitoring of elderly patients after hospitalisation for heart failure: the renewing health European project. Int J Cardiol 2018 Apr 15;257:137-142. [doi: 10.1016/j.ijcard.2017.10.099] [Medline: 29506685]
22. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Mobile phone-based telemonitoring for heart failure management: a randomized controlled trial. J Med Internet Res 2012;14(1):e31 [FREE Full text] [doi: 10.2196/jmir.1909] [Medline: 22356799]

23. Villani A, Malfatto G, Compare A, Della RF, Bellardita L, Branzi G, et al. Clinical and psychological telemonitoring and telecare of high risk heart failure patients. J Telemed Telecare 2014 Dec;20(8):468-475. [doi: 10.1177/1357633X14555644] [Medline: 25339632]
24. Amir O, Ben-Gal T, Weinstein JM, Schliamser J, Burkhoff D, Abbo A, et al. Evaluation of remote dielectric sensing (ReDS) technology-guided therapy for decreasing heart failure re-hospitalizations. Int J Cardiol 2017 Aug 01;240:279-284 [FREE Full text] [doi: 10.1016/j.ijcard.2017.02.120] [Medline: 28341372]
25. Kotooka N, Kitakaze M, Nagashima K, Asaka M, Kinugasa Y, Nochioka K, HOMES-HF study investigators. The first multicenter, randomized, controlled trial of home telemonitoring for Japanese patients with heart failure: home telemonitoring study for patients with heart failure (HOMES-HF). Heart Vessels 2018 Aug;33(8):866-876. [doi: 10.1007/s00380-018-1133-5] [Medline: 29450689]
26. Domingo M, Lupón J, González B, Crespo E, López R, Ramos A, et al. [Noninvasive remote telemonitoring for ambulatory patients with heart failure: effect on number of hospitalizations, days in hospital, and quality of life. CARME (CAtalan Remote Management Evaluation) study]. Rev Esp Cardiol 2011 Apr;64(4):277-285 [FREE Full text] [doi: 10.1016/j.recesp.2010.10.032] [Medline: 21411210]
27. Hoban MB, Fedor M, Reeder S, Chernick M. The effect of telemonitoring at home on quality of life and self-care behaviors of patients with heart failure. Home Healthc Nurse 2013;31(7):368-377. [doi: 10.1097/NHH.0b013e318291fd56] [Medline: 23632274]
28. Koulaouzidis G, Barrett D, Mohee K, Clark AL. Telemonitoring in subjects with newly diagnosed heart failure with reduced ejection fraction: From clinical research to everyday practice. J Telemed Telecare 2018 Jan 01:1357633X17751004. [doi: 10.1177/1357633X17751004] [Medline: 29419343]
29. Lyngå P, Persson H, Hägg-Martinell A, Hägglund E, Hagerman I, Langius-Eklöf A, et al. Weight monitoring in patients with severe heart failure (WISH). A randomized controlled trial. Eur J Heart Fail 2012 Apr;14(4):438-444 [FREE Full text] [doi: 10.1093/eurjhf/hfs023] [Medline: 22371525]
30. Maeng DD, Starr AE, Tomcavage JF, Sciandra J, Salek D, Griffith D. Can telemonitoring reduce hospitalization and cost of care? A health plan's experience in managing patients with heart failure. Popul Health Manag 2014 Dec;17(6):340-344. [doi: 10.1089/pop.2013.0107] [Medline: 24865986]
31. Pedone C, Rossi FF, Cecere A, Costanzo L, Antonelli IR. Efficacy of a physician-led multiparametric telemonitoring system in very old adults with heart failure. J Am Geriatr Soc 2015 Jun;63(6):1175-1180. [doi: 10.1111/jgs.13432] [Medline: 26031737]
32. Soran OZ, Feldman AM, Piña IL, Lamas GA, Kelsey SF, Selzer F, et al. Cost of medical services in older patients with heart failure: those receiving enhanced monitoring using a computer-based telephonic monitoring system compared with those in usual care: the Heart Failure Home Care trial. J Card Fail 2010 Nov;16(11):859-866. [doi: 10.1016/j.cardfail.2010.05.028] [Medline: 21055649]
33. Tompkins C, Orwat J. A randomized trial of telemonitoring heart failure patients. J Healthc Manag 2010;55(5):312-22; discussion 322. [Medline: 21077581]
34. van der Burg JM, Aziz NA, Kaptein MC, Breteler MJ, Janssen JH, van Vliet L, et al. Long-term effects of telemonitoring on healthcare usage in patients with heart failure or COPD. Clinical eHealth 2020;3:40-48. [doi: 10.1016/j.ceh.2020.05.001]
35. Veenstra W, Op den Buijs J, Pauws S, Westerterp M, Nagelsmit M. Clinical effects of an optimised care program with telehealth in heart failure patients in a community hospital in the Netherlands. Neth Heart J 2015 Jun;23(6):334-340 [FREE Full text] [doi: 10.1007/s12471-015-0692-7] [Medline: 25947078]
36. Ware P, Ross HJ, Cafazzo JA, Boodoo C, Munnery M, Seto E. Outcomes of a heart failure telemonitoring program implemented as the standard of care in an outpatient heart function clinic: pretest-posttest pragmatic study. J Med Internet Res 2020 Feb 08;22(2):e16538 [FREE Full text] [doi: 10.2196/16538] [Medline: 32027309]
37. Williams C, Wan TT. A cost analysis of remote monitoring in a heart failure program. Home Health Care Serv Q 2016;35(3-4):112-122. [doi: 10.1080/01621424.2016.1227009] [Medline: 27552654]
38. Zan S, Agboola S, Moore SA, Parks KA, Kvedar JC, Jethwani K. Patient engagement with a mobile web-based telemonitoring system for heart failure self-management: a pilot study. JMIR Mhealth Uhealth 2015;3(2):e33 [FREE Full text] [doi: 10.2196/mhealth.3789] [Medline: 25842282]
39. Frederix I, Vanderlinden L, Verboven A, Welten M, Wouters D, De Keulenaer G, et al. Long-term impact of a six-month telemedical care programme on mortality, heart failure readmissions and healthcare costs in patients with chronic heart failure. J Telemed Telecare 2019 Jun;25(5):286-293. [doi: 10.1177/1357633X18774632] [Medline: 29742959]
40. Park C, Otobo E, Ullman J, Rogers J, Fasihuddin F, Garg S, et al. Impact on readmission reduction among heart failure patients using digital health monitoring: feasibility and adoptability study. JMIR Med Inform 2019 Nov 15;7(4):e13353 [FREE Full text] [doi: 10.2196/13353] [Medline: 31730039]
41. Riley WT, Keberlein P, Sorenson G, Mohler S, Tye B, Ramirez AS, et al. Program evaluation of remote heart failure monitoring: healthcare utilization analysis in a rural regional medical center. Telemed J E Health 2015 Mar;21(3):157-162 [FREE Full text] [doi: 10.1089/tmj.2014.0093] [Medline: 25025239]

42. Bakhshi S, Li X, Semenov N, Apodaca-Madrid J, Mahoor MH, Newman KE, et al. Congestive heart failure home monitoring pilot study in urban Denver. Annu Int Conf IEEE Eng Med Biol Soc 2011;2011:3150-3153. [doi: 10.1109/IEMBS.2011.6090859] [Medline: 22255008]
43. Delaney C, Apostolidis B, Bartos S, Morrison H, Smith L, Fortinsky R. A randomized trial of telemonitoring and self-care education in heart failure patients following home care discharge. Home Health Care Manag Pract 2013 Feb 20;25(5):187-195. [doi: 10.1177/1084822312475137]
44. Dendale P, De Keulenaer G, Troisfontaines P, Weytjens C, Mullens W, Elegeert I, et al. Effect of a telemonitoring-facilitated collaboration between general practitioner and heart failure clinic on mortality and rehospitalization rates in severe heart failure: the TEMA-HF 1 (TElemonitoring in the MAnagement of Heart Failure) study. Eur J Heart Fail 2012 Mar;14(3):333-340 [FREE Full text] [doi: 10.1093/eurjhf/hfr144] [Medline: 22045925]
45. Eilat-Tsanani S, Golovner M, Marcus O, Dayan M, Sade Z, Iktelat A, et al. Evaluation of telehealth service for patients with congestive heart failure in the north of Israel. Eur J Cardiovasc Nurs 2016 Apr;15(3):e78-e84. [doi: 10.1177/1474515115602677] [Medline: 26311654]
46. Vestergaard AS, Hansen L, Sørensen SS, Jensen MB, Ehlers LH. Is telehealthcare for heart failure patients cost-effective? An economic evaluation alongside the Danish TeleCare North heart failure trial. BMJ Open 2020 Jan 27;10(1):e031670 [FREE Full text] [doi: 10.1136/bmjopen-2019-031670] [Medline: 31992604]
47. Vuorinen A, Leppänen J, Kaijanranta H, Kulju M, Heliö T, van Gils M, et al. Use of home telemonitoring to support multidisciplinary care of heart failure patients in Finland: randomized controlled trial. J Med Internet Res 2014;16(12):e282 [FREE Full text] [doi: 10.2196/jmir.3651] [Medline: 25498992]
48. Pekmezaris R, Tortez L, Williams M, Patel V, Makaryus A, Zeltser R, et al. Home telemonitoring in heart failure: a systematic review and meta-analysis. Health Aff (Millwood) 2018 Dec;37(12):1983-1989. [doi: 10.1377/hlthaff.2018.05087] [Medline: 30633680]
49. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K, et al. Home telemonitoring for congestive heart failure: a systematic review and meta-analysis. J Telemed Telecare 2010;16(2):68-76. [doi: 10.1258/jtt.2009.090406] [Medline: 20008054]
50. Dang S, Dimmick S, Kelkar G. Evaluating the evidence base for the use of home telehealth remote monitoring in elderly with heart failure. Telemed J E Health 2009 Oct;15(8):783-796. [doi: 10.1089/tmj.2009.0028] [Medline: 19831704]
51. Cleland JG, Louis AA, Rigby AS, Janssens U, Balk AH, TEN- H. Noninvasive home telemonitoring for patients with heart failure at high risk of recurrent admission and death: the Trans-European Network-Home-Care Management System (TEN-HMS) study. J Am Coll Cardiol 2005 May 17;45(10):1654-1664 [FREE Full text] [doi: 10.1016/j.jacc.2005.01.050] [Medline: 15893183]
52. Brons M, Koudstaal S, Asselbergs FW. Algorithms used in telemonitoring programmes for patients with chronic heart failure: a systematic review. Eur J Cardiovasc Nurs 2018 Oct;17(7):580-588 [FREE Full text] [doi: 10.1177/1474515118786838] [Medline: 29954184]
53. Desveaux L, Soobiah C, Bhatia RS, Shaw J. Identifying and overcoming policy-level barriers to the implementation of digital health innovation: qualitative study. J Med Internet Res 2019 Dec 20;21(12):e14994 [FREE Full text] [doi: 10.2196/14994] [Medline: 31859679]
54. Deeks J, Higgins J, Altman D. Chapter 10: Analysing data and undertaking meta-analyses. In: Cochrane Handbook for Systematic Reviews of Interventions. Chichester (UK): John Wiley & Sons; 2019.
55. Klersy C, De Silvestri A, Gabutti G, Regoli F, Auricchio A. A meta-analysis of remote monitoring of heart failure patients. J Am Coll Cardiol 2009 Oct 27;54(18):1683-1694 [FREE Full text] [doi: 10.1016/j.jacc.2009.08.017] [Medline: 19850208]
56. Yun JE, Park J, Park H, Lee H, Park D. Comparative effectiveness of telemonitoring versus usual care for heart failure: a systematic review and meta-analysis. J Card Fail 2018 Jan;24(1):19-28. [doi: 10.1016/j.cardfail.2017.09.006] [Medline: 28939459]
57. Kotb A, Cameron C, Hsieh S, Wells G. Comparative effectiveness of different forms of telemedicine for individuals with heart failure (HF): a systematic review and network meta-analysis. PLoS One 2015;10(2):e0118681 [FREE Full text] [doi: 10.1371/journal.pone.0118681] [Medline: 25714962]

Abbreviations

CHF: chronic heart failure
ED: emergency department
NYHA: New York Heart Association
RCT: randomized controlled trial
STS: structured telephone support

Edited by R Kukafka; submitted 23.12.20; peer-reviewed by T Wan, W Bramer, A Beleigoli; comments to author 04.03.21; revised version received 18.03.21; accepted 30.06.21; published 29.09.21.

Please cite as:
Auener SL, Remers TEP, van Dulmen SA, Westert GP, Kool RB, Jeurissen PPT
The Effect of Noninvasive Telemonitoring for Chronic Heart Failure on Health Care Utilization: Systematic Review
J Med Internet Res 2021;23(9):e26744
URL: https://www.jmir.org/2021/9/e26744
doi:10.2196/26744
PMID:34586072

Original Paper

Stepwise-Hierarchical Pooled Analysis for Synergistic Interpretation of Meta-analyses Involving Randomized and Observational Studies: Methodology Development

In-Soo Shin[1], PhD; Chai Hong Rim[2], MD, PhD

[1]Graduate School of Education, Dongguk University, Seoul, Republic of Korea

[2]Department of Radiation Oncology, Ansan Hospital, Korea University, Gyeonggido, Republic of Korea

Corresponding Author:
Chai Hong Rim, MD, PhD
Department of Radiation Oncology
Ansan Hospital
Korea University
Jeokgeumro 123
Danwon Gu
Gyeonggido, 15355
Republic of Korea
Phone: 82 314126850
Fax: 82 314126851
Email: crusion3@naver.com

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/9/e33534

Abstract

Background: The necessity of including observational studies in meta-analyses has been discussed in the literature, but a synergistic analysis method for combining randomized and observational studies has not been reported. Observational studies differ in validity depending on the degree of the confounders' influence. Combining interpretations may be challenging, especially if the statistical directions are similar but the magnitude of the pooled results are different between randomized and observational studies (the "gray zone").

Objective: To overcome these hindrances, in this study, we aim to introduce a logical method for clinical interpretation of randomized and observational studies.

Methods: We designed a stepwise-hierarchical pooled analysis method to analyze both distribution trends and individual pooled results by dividing the included studies into at least three stages (eg, all studies, balanced studies, and randomized studies).

Results: According to the model, the validity of a hypothesis is mostly based on the pooled results of randomized studies (the highest stage). Ascending patterns in which effect size and statistical significance increase gradually with stage strengthen the validity of the hypothesis; in this case, the effect size of the observational studies is lower than that of the true effect (eg, because of the uncontrolled effect of negative confounders). Descending patterns in which decreasing effect size and statistical significance gradually weaken the validity of the hypothesis suggest that the effect size and statistical significance of the observational studies is larger than the true effect (eg, because of researchers' bias).

Conclusions: We recommend using the stepwise-hierarchical pooled analysis approach for meta-analyses involving randomized and observational studies.

(J Med Internet Res 2021;23(9):e29642) doi:10.2196/29642

KEYWORDS

meta-analysis; observational study; randomized study; interpretation; combination; statistics; synergy; methodology; interpretation; hypothesis; validity

Introduction

In the literature, the number of meta-analyses that include observational studies has steadily increased in recent decades [1]. Nevertheless, controversy persists regarding the validity and utility of these meta-analyses. Some researchers are reluctant to assess the validity of the pooled results from studies of a heterogeneous nature and with less robust data. It is clear that compared to the pooled results from randomized studies, the results derived from observational studies may be less representative. Substantial publication of low-quality meta-analyses or those that investigate duplicated topics, which may be empowered by recruiting observational studies, has also been disparaged [2,3].

The abovementioned criticism raises the fundamental question of whether meta-analyses including observational studies should exist in the clinical field. It is clear that randomized studies form the basis of clinical research and have the greatest influence on therapeutic advances and clinical decisions. However, not all decisions in actual clinical practice can be supported only by robust evidence obtained from randomized studies [4]. In particular, it is inevitable that clinical decisions will be made based on observational studies in fields where patients with rare diseases or intractable status are commonly encountered for which there are few available known standard modalities to apply [5]. From a practical perspective, conducting randomized studies requires abundant support, and this support is not available in all medical disciplines. For example, although vendors are willing to support the design of large randomized studies to develop new drugs, if the application of a certain modality has less benefit for the vendor, the driving force for designing a high-quality study will be low.

As an example, in the treatment of liver cancer [6], there is a drug that has demonstrated mild survival gain with little local effect (ie, sorafenib: response rate of ~3%) in the treatment of inoperable cases [7,8]. This drug was studied in phase 3 randomized trials that only proved the survival benefit of the drug for unresectable liver cancers. Although radiotherapy has a significant local effect, with a response rate of over 50%, no phase 3 randomized study has demonstrated a survival gain [9]. Despite this, in a surveillance study on 161 liver cancer clinicians, 86% of physicians stated that they would apply radiotherapy for unresectable liver cancer with major vascular involvement, compared to 66% who would prescribe sorafenib [10]. How were these clinical decisions reached? Clinicians in practice inevitably rely on case series or small observational studies, especially when facing intractable situations in which randomized studies alone cannot support all clinical decisions. In other words, clinicians must perform a self–meta-analysis in their own way involving studies with various designs, commonly including observational studies. The justification for performing meta-analyses that include observational studies can be demonstrated by the necessity to optimize such self–meta-analyses.

In this study, we identify points that require improvement during the process of planning and conducting meta-analyses, and we suggest a method to synergistically interpret results from both nonrandomized and randomized studies.

Methods

Identifying Limitations to Overcome

Meta-analyses are performed to aid clinical decision-making in intractable oncologic situations in which a single standard modality has not been established. These meta-analyses must inevitably include observational and randomized studies. The limitations that we recognized must be overcome are as follows.

Confounders in Observational Studies

When comparing intervention and control groups, the randomization of participants has the advantage of evenly distributing both known and unpredictable confounders [11]. Because of these advantages, randomized studies can allow robust conclusions to be drawn with respect to determining clinical decisions. The main limitation of observational studies is the difficulty of controlling for these confounders. Furthermore, no established method has been presented to quantitatively and objectively measure how such confounders affect pooled estimates. The risk of bias is difficult to control in advance owing to the lack of availability of a protocol [12]. Therefore, the validity of the results is relatively low compared to that of randomized studies. The *Cochrane Handbook for Systematic Reviews of Interventions* states that only observational studies with at least moderate or low risk of bias should be selected in systematic reviews [12]. The Grading of Recommendations, Assessment, Development and Evaluation (GRADE) handbook explains the limitations of observational studies across several categories; it addresses structural issues such as flawed measurement of outcomes and exposure, short follow-up, and inappropriate eligibility criteria, along with inadequate control of confounding. If such a risk exists, it is recommended to downgrade the quality of the study by 1-2 grades according to the degree [13].

On the other hand, as methods of disease assessment advance, more factors are being identified that influence a patient's prognosis. Recent well-designed observational studies were designed to control a variable number of confounders in the study. A study in which clinical confounders were controlled using methods such as propensity matching and multiple regression analysis [14,15] in a sufficient number of patients should not be analyzed at the same level as studies in which such methods were not used.

In addition, when analyzing observational studies, consideration should be given to how the treatment decisions have been established. Consider two studies that verified the effectiveness of adjuvant radiotherapy after biliary tract cancer surgery (adjuvant radiotherapy for biliary tract cancer has the effect of reducing recurrence, but an increase in survival has not been sufficiently shown) [16]. All related data were obtained from nonrandomized studies. In the first study, the institution decided at a multidisciplinary meeting whether to apply adjuvant radiotherapy; the results were based on comparative data from a single center (adjuvant radiotherapy arm vs no radiotherapy arm). On the other hand, the second observational study

compared the results obtained at two independent institutions. The first institution actively performed radiotherapy to maximize treatment efficiency and hospital profits. The second institution did not have a radiation oncology department; therefore, it was necessary to make radiotherapy referrals externally. This situation made surgeons reluctant to recommend radiotherapy, and the patients rarely received adjuvant radiotherapy. In the first study, patients who received radiotherapy were more likely to have prognostic factors related to recurrence. In the second study, the clinical profiles of patients were likely to be evenly distributed between the intervention and control groups.

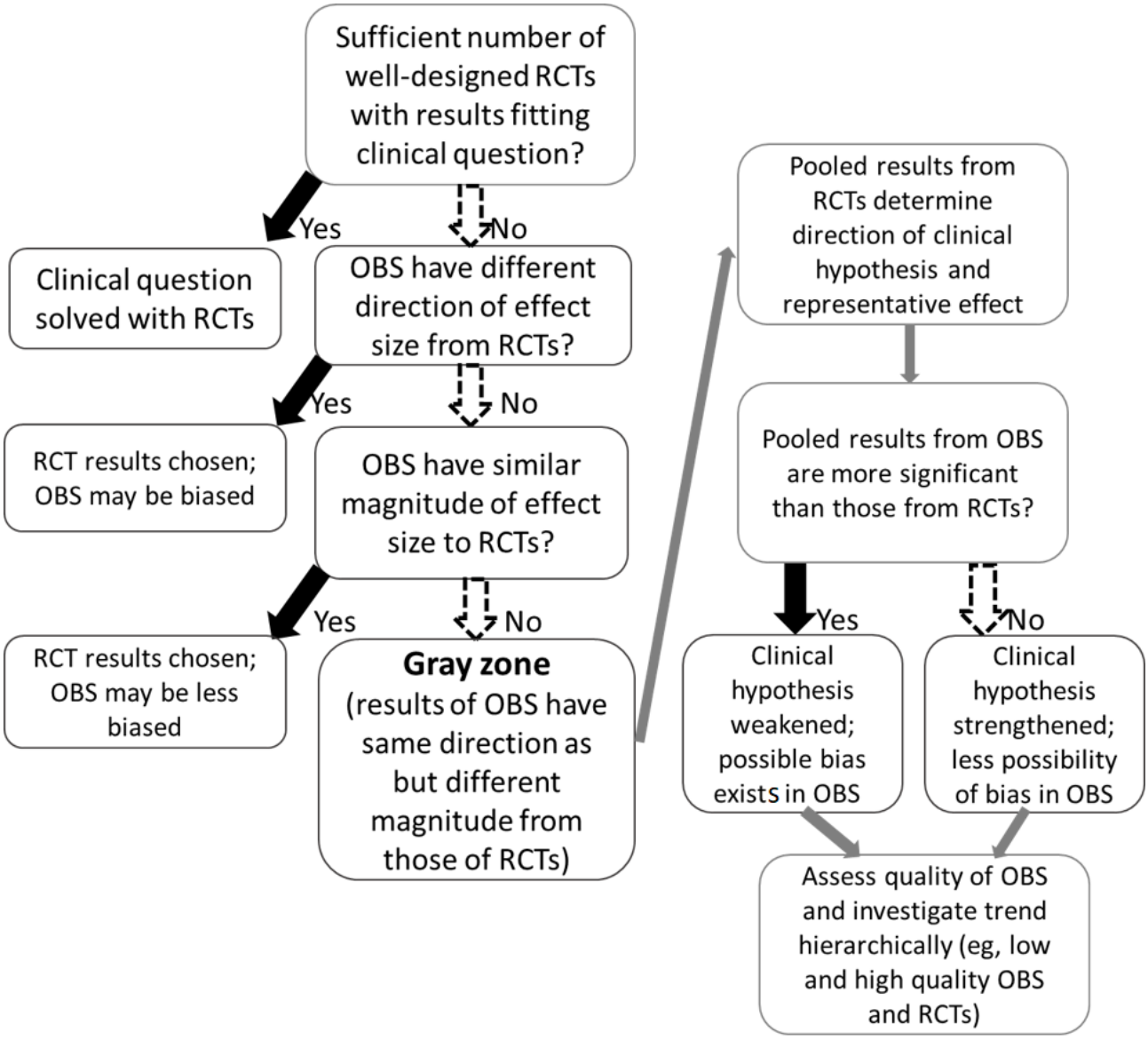
In summary, observational studies are more likely to be affected by confounders than randomized studies. However, observational studies need to be weighted differently according to their design as well as the degree of control for confounders.

The "Gray Zone": Necessity of Combining Interpretations of Randomized and Observational Studies

If there are enough well-designed randomized studies on a subject to be analyzed, there is little need for a meta-analysis including nonrandomized studies. However, the more detailed the clinical topic to be studied and the more incurable the disease, the more difficult it is to make a therapeutic decision using only data from randomized studies. In their randomized sampling analysis of Cochrane reviews, Shrier et al [11] reported that 6 of 16 reviews included 0, 1, or 2 randomized trials. Furthermore, 158 of 183 analyses in 7 additional studies included 2 or fewer randomized studies [11,17]. The reason that randomized studies can control for known and unknown confounders is based on the assumption that the number of participants included in the study is infinite [18]. However, in practice, many randomized studies have difficulty recruiting a sufficient number of patients. Additionally, a blinding process is needed to control for the placebo effect and optimize the design of randomized studies [11]. However, this design is not possible in randomized studies comparing different types of treatment (eg, comparing the effectiveness of lobectomy and radiosurgery in early lung cancer). Limitations such as these necessitate the identification of clinical reasoning, complemented by meta-analyses involving observational studies [19,20].

In a meta-analysis that includes both randomized and observational studies, if the pooled results of randomized studies and observational studies have similar effect sizes in the same direction without a notable difference in statistical significance, there will be little disagreement in the interpretation of these results. In contrast, if the directions of the two results clearly contradict each other, the majority of scholars will agree to adopt the results of the randomized studies and reject the results of the observational studies, under the assumption that the randomized studies lack significant design flaws. However, there is a "gray zone" where the results of studies with different designs (randomized vs observational studies) have the same direction, but the magnitude of the effect size differs (Figure 1); no clear standard method has been established for combining and interpreting such results. In these situations, the role of pooled results from observational studies may be rather auxiliary if a sufficient number of randomized studies with sufficient validity are recruited. However, as described above, when treating rare diseases or intractable diseases in the clinical field, information from observational studies is necessary for clinical decisions. In other words, it is necessary to complement clinical reasoning based on pooled results of observational studies when the number of randomized studies and the numbers of patients recruited in said studies are insufficient.

Figure 1. Process by which the "gray zone" is explored and clinical logic flow in the gray zone. OBS: observational studies; RCTs: randomized controlled trials.



Clinical Logic Flow in the Gray Zone

Physicians should make clinical decisions by using studies with different designs in gray zone situations. Many clinicians review literature found through subjective searches and tend to rely more on research published by authoritative institutions. However, this unsystematic method should be avoided, and a recommended clinical logic flow of interpretation may be as follows:

The pooled results from RCTs determine the direction of the clinical hypothesis and the representative effect size. In the gray zone, complementation from data synthesized from observational studies may be necessary. If the pooled results from observational studies are more significant than those from RCTs, the clinical hypothesis could be weakened and confounding bias could be present among the observational studies. In other words, the clinical hypothesis seems more meaningful in clinical studies with a possibility of bias and a low evidence grade, but it has less significant results than previously expected in high-grade studies such as randomized studies. If the pooled effect of observational studies is less significant, the clinical hypothesis can be strengthened, and there is less possibility of bias. This suggests that clinical hypotheses are less meaningful in clinical studies with possible bias and a low grade of evidence, but more meaningful results are produced in high-grade studies such as randomized studies. Finally, the quality of observational studies can be assessed, and trends of pooled effects according to study design (high- and low-quality observational studies and randomized studies) can be investigated (Figure 1). This process will typically categorize three or more groups, and hierarchical trends can be used to complement clinical hypotheses.

This clinical logical flow will be set as a model and is introduced in detail below.

Rationale of Stepwise-Hierarchical Pooled Analysis

Stepwise-hierarchical pooled analysis is a method of interpreting the pooled results of studies categorized according to their

design and validity. In general, the studies included in a meta-analysis are analyzed by dividing them into at least three groups, and then the individual results of each group and the changing trends among groups are analyzed. In the first level, all studies are analyzed, and in the second level, balanced studies in which major confounders are controlled for are analyzed. Balanced studies are generally defined as those in which major clinical factors are evenly distributed, either based on the study design or statistical method, with additional consideration of the treatment strategy of affiliated institutions whenever possible (discussed in the previous section). Randomized studies can also be included at this stage in the analysis as balanced studies, especially when the number of nonrandomized and balanced studies is small. The final step is to analyze randomized studies. Randomized studies can be analyzed at one level lower if the design is suboptimal (eg, the main clinical factors are not evenly distributed between the intervention and control groups or the randomization method is not reliable) or the number of included patients is too small.

Briefly, the interpretation of stepwise-hierarchical pooled analysis is as follows: The pooled results and statistical significance of the randomized study mainly determine the validity of the hypothesis. When proceeding from an analysis that includes all studies with a low evidence grade to an analysis of more selected studies, this trend further supports the validity of the hypothesis if it is a pattern in which the magnitude and statistical significance of the result increase. However, a decreasing pattern may weaken the validity of the hypothesis, suggesting that there may be biases in the design and results of studies with low evidence ratings.

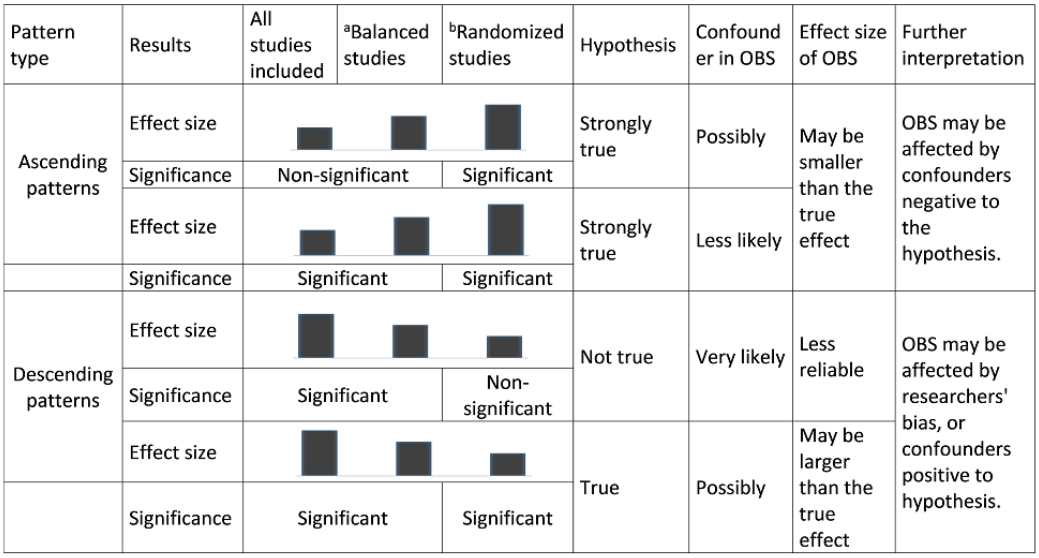
Results

Descriptive Interpretation

The descriptive interpretation of the four representative patterns (Figure 2) is as follows:

1. The effect size and statistical significance increase gradually: The results of the randomized study analysis are statistically significant, and the effect size gradually increases, strengthening the support for the hypothesis. Therefore, the probability is high that the hypothesis is true and strongly positive. The effect size in the observational studies will be lower than the true effect, and if confounders are controlled for, the effect size can be increased. The results of the pooled analyses of observational studies with confounders may not be statistically significant.
2. The effect size gradually increases and the results are statistically significant at all stages: The results of the randomized study analysis are statistically significant, and the pattern of increasing effect size gradually strengthens the reliability of the hypothesis. Therefore, the probability is high that the hypothesis is true and strongly positive. The effect size of observational studies is lower than that of the true effect. Confounders may have a negative effect on the results of observational studies, but because they show statistically significant results, this effect is assumed to be smaller than that in pattern 1.
3. The effect size and statistical significance decrease gradually: The target hypothesis is rejected because the results of the randomized study analysis are not statistically significant. The effect size and statistical significance of the observational studies are not trustworthy. Observational studies are likely to be affected by confounders and researchers' bias.
4. The effect size gradually decreases and the results are statistically significant at all stages: The target hypothesis is judged to be true because the results of the randomized study analysis are statistically significant. However, the pattern of the effect size gradually decreases, which lowers the reliability of the hypothesis. The effect size of observational studies is larger than the true effect. Observational studies are likely to be affected by confounders and researchers' bias. Once again, out of the above patterns, the hypothesis is true if the effect sizes are similar in the pooled analyses of both randomized and observational studies, and both analyses are statistically significant. In contrast, if the results of the randomized and observational studies contradict each other, the pooled results of the randomized studies should be weighted more heavily and further investigation of this contradiction should be performed. The stepwise-hierarchical method may not be highly necessary for these situations.

Figure 2. Interpretation of the four representative patterns of stepwise-hierarchical pooled analysis. OBS: observational studies.



Examples of Clinical Interpretation

Our team recently published two meta-analyses that used the stepwise-hierarchical method [21,22]. The main results of the two studies showed typical features of the ascending and descending patterns. Therefore, the clinical interpretations of the main results of each paper are discussed as examples.

The first study focused on adjuvant radiochemotherapy versus chemotherapy after surgery for gastric cancer. In general, the role of additional radiotherapy has not been accepted widely after D2 gastrectomy, including extensive lymphatic dissection [23]. This is because the result for the primary endpoint (disease-free survival) of the only phase 3 randomized study on the subject was marginally nonsignificant [24]. However, several observational studies and small randomized trials have reported the oncologic benefit of radiotherapy [25,26]. Therefore, our team conducted a meta-analysis including randomized and nonrandomized comparative studies to evaluate the disease-free survival benefit of adjuvant radiochemotherapy [21]. As shown in Figure 3, the effect size in the pooled analysis for all studies was 1.264 (95% CI 0.997-1.603), and the *P* value was marginally nonsignificant at *P*=.053. The effect size in the pooled analysis of balanced studies (ie, studies in which major clinical indicators are similarly distributed between arms) was 1.417 (95% CI 1.171-1.715), and the *P* value was highly significant at *P*<.001. The effect size in the pooled analysis of only randomized studies was 1.440 (95% CI 1.110-1.867; *P*=.006), which was also highly significant. The trend of these results correlates with the first of the four typical patterns described above. In other words, the hypothesis of this meta-analysis (radiochemotherapy is significant in reducing disease-free survival after D2 gastrectomy) is strongly supported. The trend in which the effect size increases from considering all studies to considering only balanced or randomized studies strengthened the validity of the hypothesis. The results of observational studies may have underestimated the effect size relative to the true effect due to the influence of confounders (eg, patients assumed to have greater risk of recurrence underwent radiochemotherapy). Furthermore, the low heterogeneity in the analyses of balanced and randomized studies suggests that the pooled results of those studies are reliable and well designed, and they are less affected by possible confounders.

The second study was on the benefit of local treatment for oligometastases. Oligometastases refer to a disease state with ≤3 or ≤5 metastatic lesions (definitions differ between studies) [27]. In the recent literature, it was proposed that local treatment for oligometastatic foci could prolong cancer survival [28,29]. Several randomized studies have been published, but the number of patients recruited is generally insufficient [30]. In addition, because the studies in the literature were published according to the type of primary cancer, it was difficult to comprehensively analyze the oncologic benefit of local treatment on general oligometastases. Therefore, we attempted to prove the hypothesis that local treatment for oligometastases will increase overall survival in a meta-analysis [22]. In the analysis of all studies, the pooled effect size was 3.039 (95% CI 2.272-4.064) and the *P* value was significant (*P*<.001). In the analysis of balanced studies, the pooled effect size was 2.560 (95% CI 1.791-3.659), and the *P* value was also highly significant (*P*<.001). In the final analysis of randomized studies, the *P* value was significant (*P*=.04); however, the pooled effect size was 1.406 (95% CI 1.015-1.949), which was smaller in magnitude than that in the previous analyses. The trend of these results correlates with the fourth of the four typical patterns (Figure 4). In other words, the hypothesis of this meta-analysis is true, referring to the analysis results of randomized studies. However, unlike the pattern seen in the meta-analysis of gastric

cancer, the change in the effect size or *P* value does not increase the validity of the hypothesis. Observational studies may have been affected by a confounder, and the results may have been larger than the true effect size. Of note, in many studies, local treatment arms had a lower number of metastatic foci than control arms, although the difference was not statistically significant. Unlike the low heterogeneity in the pooled analysis of randomized studies, the high heterogeneity among observational studies suggests the possible effects of confounders. However, such a pattern does not necessarily indicate that the result is weak and not useful. We also found that the benefit of local treatment was higher in certain cancer types (eg, lung cancer, colorectal cancer) and with higher metastatic burden (studies with <5 metastases compared to those with <3 metastases) in further subgroup analyses. Therefore, the authors concluded that although local treatment for oligometastases is beneficial, patients must be carefully selected with consideration of the type of disease or metastatic burden, and the design of future observational studies needs to be improved.

Figure 3. A clinical meta-analysis example of the ascending pattern in the stepwise-hierarchical method based on our previous meta-analysis evaluating the benefits of adjuvant radiochemotherapy after D2 gastrectomy as compared to chemotherapy alone [21]. The forest plots are newly drawn from the raw data obtained by the authors. ES: effect size; CRT: chemoradiotherapy.

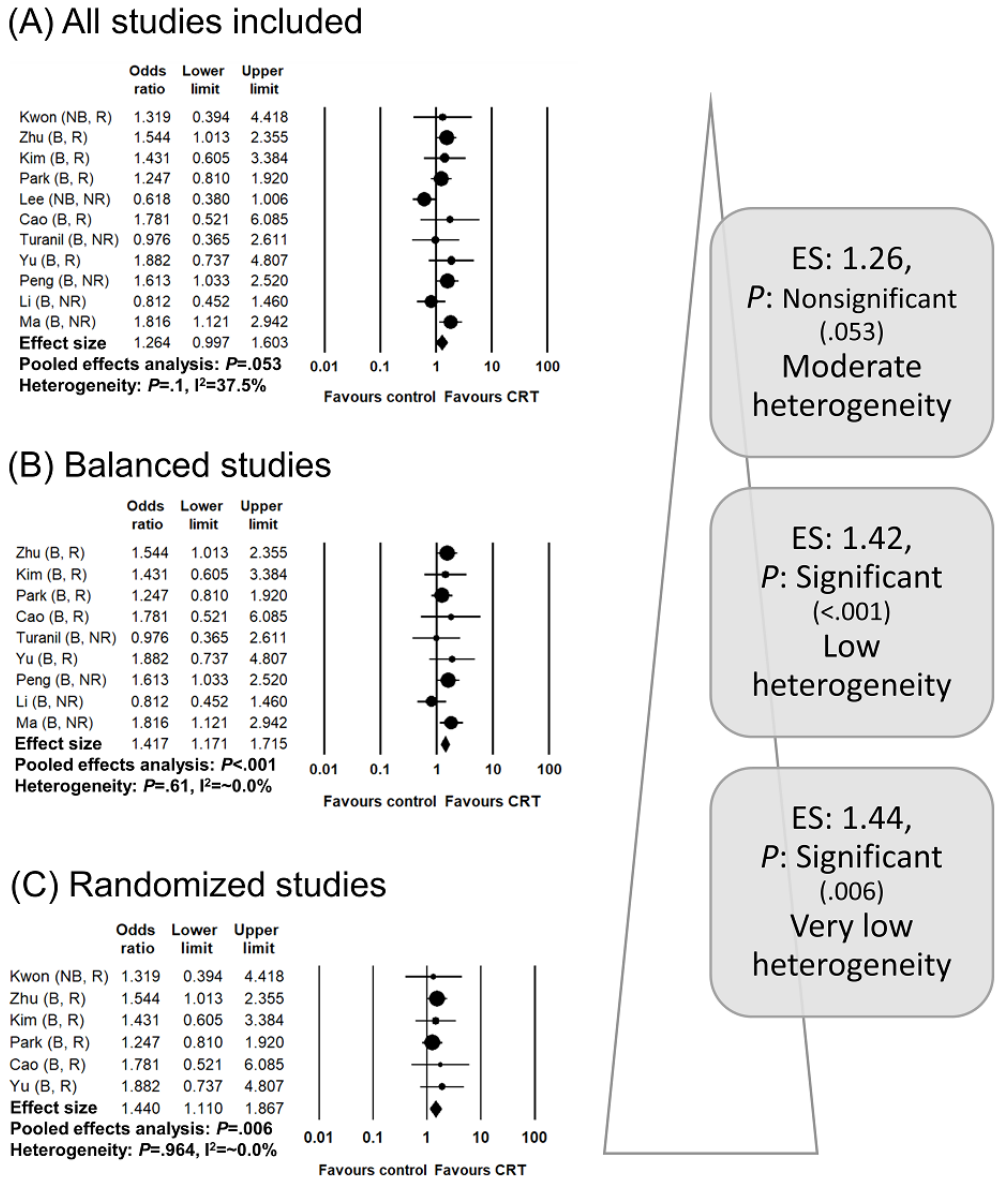
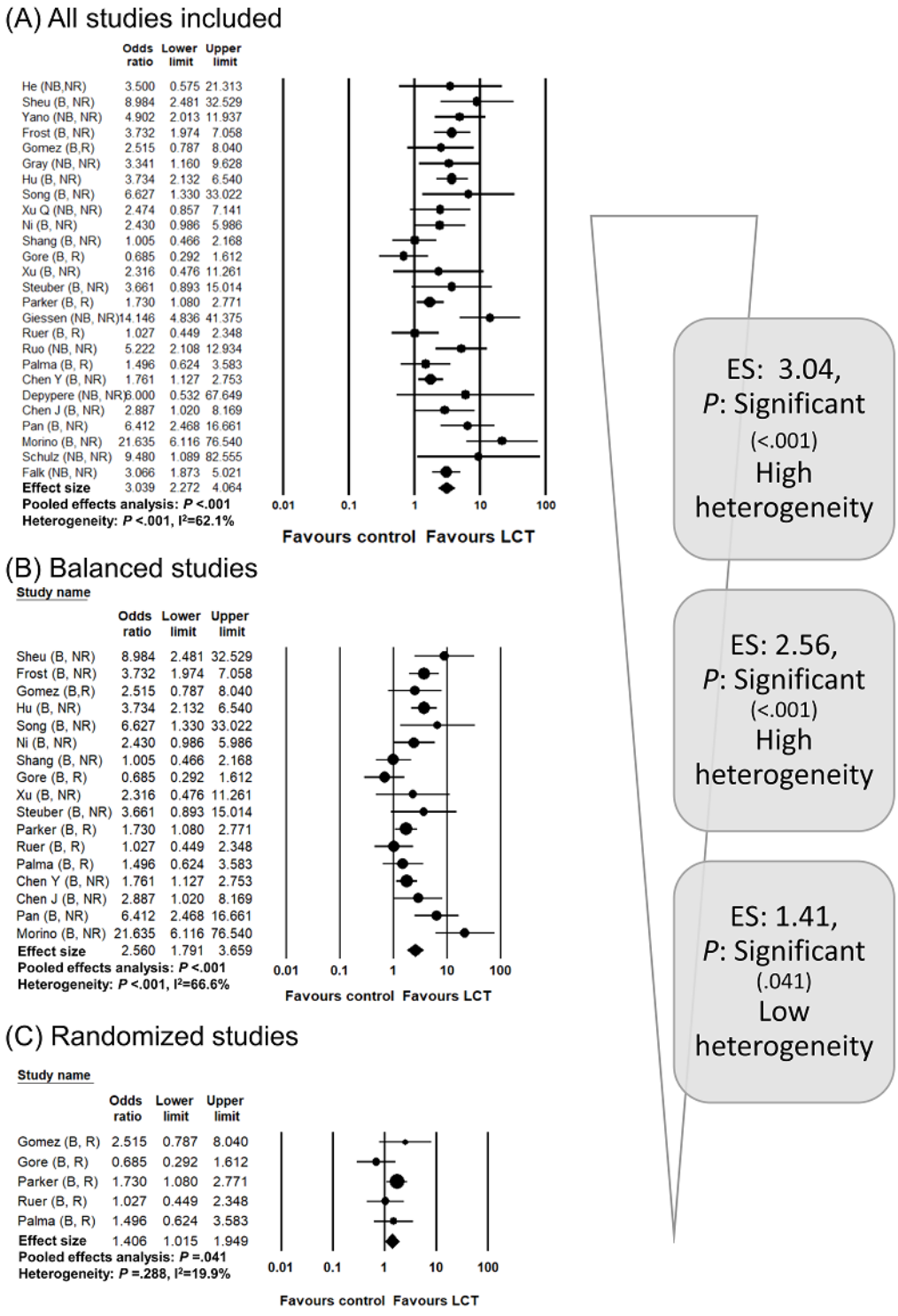


Figure 4. A clinical meta-analysis example of the descending pattern in the stepwise-hierarchical method based on our previous meta-analysis evaluating the benefits of local treatment on oligometastatic disease [22]. The forest plots are newly drawn from the raw data obtained by the authors. ES: effect size; LCT: local consolidative treatment.



Discussion

Principal Considerations

The number of meta-analyses in the literature that include observational studies has been steadily increasing [1]. In actual clinical fields, the decisions that can be fully supported by blinded, randomized studies are limited. It is difficult to assemble a sufficient number of patients free from ethical considerations when the benefits of an intervention are expected to be significant due to observational studies [31]. The treatment methods applied to the intervention and control groups should be of the same type in terms of what the patient perceives. As the understanding of a disease increases and treatment options diversify, it will become increasingly necessary to obtain

assistance for therapeutic decisions from meta-analyses of observational studies [17]. Shrier et al [11] described the clinical necessity of meta-analyses including nonrandomized studies; they discussed the practical limitations of randomized studies and explained that well-designed observational studies obtained similar results to those of randomized studies. Vandenbroucke [32] suggested that the reliability of the results can be improved by meta-analyzing observational studies selected in terms of the subject, design, and analysis method.

Other previous publications have discussed the justification for including observational studies in meta-analyses or how to select studies with valid qualities. The *Cochrane Handbook for Systematic Reviews of Interventions* [12], which previously had a conservative perspective on including nonrandomized studies in systematic reviews, added a chapter in the most recent edition on how to assess and interpret these studies in a meta-analysis. Of note, the handbook asserted that only observational studies without a high risk of bias should be included in the meta-analysis. It was also pointed out that there is still no established model that can evaluate how bias or confounders of observational studies affect estimates. However, little is known about how observational and randomized studies should be integrated and analyzed to yield actual clinical decisions.

Limitations of observational studies are categorized and explained in the GRADE handbook [13]. They include fundamental flaws such as inappropriate eligibility criteria, flawed measurement of exposure, inadequate follow-up, and inadequate control of confounders. In the presence of these limitations, it is suggested that the evidence grade should be lowered by one or two steps. Although it is agreeable to evaluate the validity of observational studies in stages, a practical methodology for integrating randomized studies with low- and high-grade observational studies into a formal meta-analysis has not been sufficiently introduced. Indeed, many clinical practice guidelines use GRADE to analyze the grade of evidence and recommendations; those analyses, including observational studies, often rely on narrative reviews. In summary, the necessity to include observational studies in systematic reviews and evaluate their quality has been highlighted in recent literature analytics. However, obtaining clinically useful information by complementing the results of randomized studies with information from observational studies has not been sufficiently suggested.

Recently, the integration of different studies into designs in the field of network meta-analysis has been discussed. In a network meta-analysis, direct and indirect evidence should be analyzed and integrated. A methodology integrating randomized and observational studies has also been studied in the process of synthesizing evidence with different levels of validity [33,34]. Efthimiou et al [35] classified the proposed integrated analysis methods in the literature to date into three categories. These are design-adjusted analyses, in which all trials included in the network meta-analysis involve estimates adjusted according to possible bias and overprecision (based on expert opinions); using informative priors, in which meta-analysis of randomized trials is performed based on priors formulated from meta-analyzing observational studies (Bayesian approach); and three categorical models, in which a meta-analysis is performed for each design, and consequently, the overall effect is acquired by synthesizing all design-specific estimates. Although these approaches have been suggested, according to the scoping review by Zhang et al [36], the vast majority (74%) of network meta-analyses used naïve pooling without specific consideration.

The methods suggested in the field of network meta-analysis and the method of the present study are similar in principle. That is, the results are integrated into a differential consideration of the validity of the evidence. On the other hand, the model of this study is distinct from those suggested in network meta-analysis, in that it is a clinically logical model that analyzes the trend of the synthesized results after differential analysis by considering study quality. In addition, the model proposed in this study is less difficult to apply because it does not require additional statistical analysis or software use. It also has the advantage that clinical interpretation is easy and intuitive, even for physicians without mathematical expertise, because it is based on clinical logical flow. These distinctive features and practical merits provide a summary of the significance of the stepwise hierarchical model, which is a novel method suggested for integration of nonrandomized and randomized studies in frequentist (or classical) meta-analyses.

Limitations

The limitations of this study are as follows. The four typical patterns described in this study cannot explain all possible patterns and their variations. For a detailed interpretation of clinical decisions, indicators of heterogeneity and publication bias should be interpreted as well. Researchers who are accustomed to making bidirectional decisions based on a specific *P* value of .05 will find the process of analyzing trends in changes in statistical significance unfamiliar [37]. Therefore, quantitative and qualitative interpretation are necessary. Cooperation between a clinician and a biostatistician with sufficient experience in meta-analysis is recommended to successfully use our model. The conclusion empowered by the main results as well as the subgroup results of our second example study can serve as a reference of cooperative interpretation. We expect future meta-analysis studies to use our model and interpret their results, including diverse variations to strengthen the utility of the model and resolve current limitations.

Conclusions

We recommend using the stepwise-hierarchical pooled analysis approach as a model for interpreting meta-analyses involving randomized and observational studies in a synergistic manner.

Acknowledgments

This study was supported by the National Research Fund of Korea (NRF-2018R1D1A1B07046998 and NRF-2019M2D2A1A01031560). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

None declared.

References

1. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000 Apr 19;283(15):2008-2012. [doi: 10.1001/jama.283.15.2008] [Medline: 10789670]
2. Gurevitch J, Koricheva J, Nakagawa S, Stewart G. Meta-analysis and the science of research synthesis. Nature 2018 Mar 07;555(7695):175-182. [doi: 10.1038/nature25753] [Medline: 29517004]
3. Siontis KC, Hernandez-Boussard T, Ioannidis JPA. Overlapping meta-analyses on the same topic: survey of published studies. BMJ 2013 Jul 19;347(jul19 1):f4501-f4501 [FREE Full text] [doi: 10.1136/bmj.f4501] [Medline: 23873947]
4. Frieden TR. Evidence for health decision making - beyond randomized, controlled trials. N Engl J Med 2017 Aug 03;377(5):465-475. [doi: 10.1056/NEJMra1614394] [Medline: 28767357]
5. Shin I, Rim CH. Updating perspectives on meta-analyses in the field of radiation oncology. Medicina (Kaunas) 2021 Jan 28;57(2):117 [FREE Full text] [doi: 10.3390/medicina57020117] [Medline: 33525358]
6. Hong S, Won Y, Park YR, Jung K, Kong H, Lee ES, Community of Population-Based Regional Cancer Registries. Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2017. Cancer Res Treat 2020 Apr;52(2):335-350 [FREE Full text] [doi: 10.4143/crt.2020.206] [Medline: 32178489]
7. Cheng A, Kang Y, Chen Z, Tsao C, Qin S, Kim JS, et al. Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. The Lancet Oncology 2009 Jan;10(1):25-34. [doi: 10.1016/s1470-2045(08)70285-7]
8. Llovet JM, Ricci S, Mazzaferro V, Hilgard P, Gane E, Blanc J, et al. Sorafenib in advanced hepatocellular carcinoma. N Engl J Med 2008 Jul 24;359(4):378-390. [doi: 10.1056/nejmoa0708857]
9. Rim CH, Kim CY, Yang DS, Yoon WS. Comparison of radiation therapy modalities for hepatocellular carcinoma with portal vein thrombosis: a meta-analysis and systematic review. Radiother Oncol 2018 Oct;129(1):112-122. [doi: 10.1016/j.radonc.2017.11.013] [Medline: 29233562]
10. Lee JS, Kim BK, Kim SU, Park JY, Ahn SH, Seong JS, et al. A survey on transarterial chemoembolization refractoriness and a real-world treatment pattern for hepatocellular carcinoma in Korea. Clin Mol Hepatol 2020 Jan;26(1):24-32 [FREE Full text] [doi: 10.3350/cmh.2018.0065] [Medline: 31104456]
11. Shrier I, Boivin J, Steele RJ, Platt RW, Furlan A, Kakuma R, et al. Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. Am J Epidemiol 2007 Nov 15;166(10):1203-1209. [doi: 10.1093/aje/kwm189] [Medline: 17712019]
12. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M. Ch 24.5: Assessing risk of bias in non-randomized studies. In: Cochrane Handbook for Systematic Reviews of Interventions. New York, NY: John Wiley & Sons; 2019.
13. Schünemann H, Brożek J, Guyatt G, Oxman A. GRADE handbook.: Cochrane Collaboration; 2013. URL: https://gdt.gradepro.org/app/handbook/handbook.html#h.t09m6hxd5iz [accessed 2021-08-23]
14. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology 1999;10(1):37-48. [doi: 10.1097/00001648-199901000-00008]
15. Hernán MA. A definition of causal effect for epidemiological research. J Epidemiol Community Health 2004 Apr 01;58(4):265-271 [FREE Full text] [doi: 10.1136/jech.2002.006361] [Medline: 15026432]
16. Ren B, Guo Q, Yang Y, Liu L, Wei S, Chen W, et al. A meta-analysis of the efficacy of postoperative adjuvant radiotherapy versus no radiotherapy for extrahepatic cholangiocarcinoma and gallbladder carcinoma. Radiat Oncol 2020 Jan 15;15(1):15 [FREE Full text] [doi: 10.1186/s13014-020-1459-x] [Medline: 31941520]
17. Shrier I. Cochrane Reviews: new blocks on the kids. Br J Sports Med 2003 Dec 01;37(6):473-474 [FREE Full text] [doi: 10.1136/bjsm.37.6.473] [Medline: 14665580]
18. Agresti A. Categorical Data Analysis. New York, NY: John Wiley & Sons; 2003.
19. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
20. Melia R, Francis K, Duggan J, Bogue J, O'Sullivan M, Chambers D, et al. Mobile health technology interventions for suicide prevention: protocol for a systematic review and meta-analysis. JMIR Res Protoc 2018 Jan 26;7(1):e28. [doi: 10.2196/resprot.8635]

21. Rim CH, Shin I, Lee HY, Yoon WS, Park S. Oncologic benefit of adjuvant chemoradiation after D2 gastrectomy: a stepwise hierarchical pooled analysis and systematic review. Cancers (Basel) 2020 Jul 31;12(8):2125 [FREE Full text] [doi: 10.3390/cancers12082125] [Medline: 32751879]
22. Rim CH, Shin I, Park S, Lee HY. Benefits of local consolidative treatment in oligometastases of solid cancers: a stepwise-hierarchical pooled analysis and systematic review. NPJ Precis Oncol 2021 Jan 21;5(1):2 [FREE Full text] [doi: 10.1038/s41698-020-00141-4] [Medline: 33479481]
23. Chang JS, Koom WS, Lee Y, Yoon HI, Lee HS. Postoperative adjuvant chemoradiotherapy in D2-dissected gastric cancer: is radiotherapy necessary after D2-dissection? World J Gastroenterol 2014 Sep 28;20(36):12900-12907 [FREE Full text] [doi: 10.3748/wjg.v20.i36.12900] [Medline: 25278687]
24. Lee J, Lim D, Kim S, Park S, Park J, Park Y. Phase III trial comparing capecitabine plus cisplatin versus capecitabine plus cisplatin with concurrent capecitabine radiotherapy in completely resected gastric cancer with D2 lymph node dissection: the ARTIST trial. J Clin Oncol 2012;30(3):268-273. [doi: 10.3410/f.14208981.15726093]
25. Xu J, Zhu J, Wei Q. Adjuvant radiochemotherapy versus chemotherapy alone for gastric cancer: implications for target definition. J Cancer 2019;10(2):458-466 [FREE Full text] [doi: 10.7150/jca.27335] [Medline: 30719140]
26. Zhou M, Kang M, Li G, Guo X, Zhang Z. Postoperative chemoradiotherapy versus chemotherapy for R0 resected gastric cancer with D2 lymph node dissection: an up-to-date meta-analysis. World J Surg Oncol 2016 Aug 08;14(1):209 [FREE Full text] [doi: 10.1186/s12957-016-0957-7] [Medline: 27502921]
27. Guckenberger M, Lievens Y, Bouma AB, Collette L, Dekker A, deSouza NM, et al. Characterisation and classification of oligometastatic disease: a European Society for Radiotherapy and Oncology and European Organisation for Research and Treatment of Cancer consensus recommendation. Lancet Oncol 2020 Jan;21(1):e18-e28. [doi: 10.1016/s1470-2045(19)30718-1]
28. Palma DA, Louie AV, Rodrigues GB. New strategies in stereotactic radiotherapy for oligometastases. Clin Cancer Res 2015 Nov 30;21(23):5198-5204. [doi: 10.1158/1078-0432.ccr-15-0822]
29. Weichselbaum RR, Hellman S. Oligometastases revisited. Nat Rev Clin Oncol 2011 Jun 22;8(6):378-382. [doi: 10.1038/nrclinonc.2011.44] [Medline: 21423255]
30. Laba JM, Senan S, Schellenberg D, Harrow S, Mulroy L, Senthi S, et al. Identifying barriers to accrual in radiation oncology randomized trials. Curr Oncol 2017 Dec 01;24(6):e524-e530 [FREE Full text] [doi: 10.3747/co.24.3662] [Medline: 29270062]
31. Nicholls SG, Carroll K, Zwarenstein M, Brehaut JC, Weijer C, Hey SP, Ethics of Pragmatic Trials project. The ethical challenges raised in the design and conduct of pragmatic trials: an interview study with key stakeholders. Trials 2019 Dec 23;20(1):765 [FREE Full text] [doi: 10.1186/s13063-019-3899-x] [Medline: 31870433]
32. Vandenbroucke JP. When are observational studies as credible as randomised trials? Lancet 2004 May;363(9422):1728-1731. [doi: 10.1016/s0140-6736(04)16261-2]
33. Chaimani A, Caldwell DM, Li T, Higgins JPT, Salanti G. Chapter 11: Undertaking network meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions, Second Edition. New York, NY: John Wiley & Sons; 2019:285-320.
34. Slattery BW, Haugh S, O'Connor L, Francis K, Dwyer CP, O'Higgins S, et al. An evaluation of the effectiveness of the modalities used to deliver electronic health interventions for chronic pain: systematic review with network meta-analysis. J Med Internet Res 2019 Jul 17;21(7):e11086 [FREE Full text] [doi: 10.2196/11086] [Medline: 31317869]
35. Efthimiou O, Mavridis D, Debray TPA, Samara M, Belger M, Siontis GCM, GetReal Work Package 4. Combining randomized and non-randomized evidence in network meta-analysis. Stat Med 2017 Apr 15;36(8):1210-1226. [doi: 10.1002/sim.7223] [Medline: 28083901]
36. Zhang K, Arora P, Sati N, Béliveau A, Troke N, Veroniki AA, et al. Characteristics and methods of incorporating randomized and nonrandomized evidence in network meta-analyses: a scoping review. J Clin Epidemiol 2019 Sep;113:1-10. [doi: 10.1016/j.jclinepi.2019.04.023] [Medline: 31059803]
37. Wasserstein R, Lazar N. The ASA statement on p-values: context, process, and purpose. Am Stat 2016 Jun 09;70(2):129-133 [FREE Full text] [doi: 10.1080/00031305.2016.1154108]

Abbreviations

GRADE: Grading of Recommendations, Assessment, Development and Evaluation

Edited by G Eysenbach; submitted 14.04.21; peer-reviewed by M Behzadifar; comments to author 20.05.21; revised version received 09.06.21; accepted 27.07.21; published 02.09.21.

Please cite as:
Shin IS, Rim CH
Stepwise-Hierarchical Pooled Analysis for Synergistic Interpretation of Meta-analyses Involving Randomized and Observational Studies: Methodology Development
J Med Internet Res 2021;23(9):e29642
URL: https://www.jmir.org/2021/9/e29642
doi:10.2196/29642
PMID:34315697

Original Paper

Understanding Pediatric Surgery Cancellation: Geospatial Analysis

Lei Liu[1,2], BSc; Yizhao Ni[1,3], PhD; Andrew F Beck[3,4,5], MPH, MD; Cole Brokamp[3,6], PhD; Ryan C Ramphul[7], PhD; Linda D Highfield[8,9], PhD; Megha Karkera Kanjia[10,11], MD; J "Nick" Pratap[3,12], MB BChir

[1]Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

[2]Department of Biomedical Informatics, University of Cincinnati, Cincinnati, OH, United States

[3]Department of Pediatrics, University of Cincinnati, Cincinnati, OH, United States

[4]Division of General and Community Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

[5]Division of Hospital Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

[6]Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

[7]Department of Government Relations and Community Benefits, Texas Children's Hospital, Houston, TX, United States

[8]Department of Management, Policy & Community Health, University of Texas Health Science Center School of Public Health, Houston, TX, United States

[9]Department of Epidemiology, Human Genetics & Environmental Sciences, University of Texas Health Science Center School of Public Health, Houston, TX, United States

[10]Department of Pediatric Anesthesiology and Pain Management, Texas Children's Hospital, Houston, TX, United States

[11]Department of Anesthesiology, Baylor College of Medicine, Houston, TX, United States

[12]Department of Anesthesia, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

Corresponding Author:
J "Nick" Pratap, MB BChir
Department of Anesthesia
Cincinnati Children's Hospital Medical Center
MLC 2001
3333 Burnet Avenue
Cincinnati, OH, 45229-3039
United States
Phone: 1 513 636 4408
Email: jnpratap@pratap.co.uk

Abstract

Background: Day-of-surgery cancellation (DoSC) represents a substantial wastage of hospital resources and can cause significant inconvenience to patients and families. Cancellation is reported to impact between 2% and 20% of the 50 million procedures performed annually in American hospitals. Up to 85% of cancellations may be amenable to the modification of patients' and families' behaviors. However, the factors underlying DoSC and the barriers experienced by families are not well understood.

Objective: This study aims to conduct a geospatial analysis of patient-specific variables from electronic health records (EHRs) of Cincinnati Children's Hospital Medical Center (CCHMC) and of Texas Children's Hospital (TCH), as well as linked socioeconomic factors measured at the census tract level, to understand potential underlying contributors to disparities in DoSC rates across neighborhoods.

Methods: The study population included pediatric patients who underwent scheduled surgeries at CCHMC and TCH. A 5-year data set was extracted from the CCHMC EHR, and addresses were geocoded. An equivalent set of data >5.7 years was extracted from the TCH EHR. Case-based data related to patients' health care use were aggregated at the census tract level. Community-level variables were extracted from the American Community Survey as surrogates for patients' socioeconomic and minority status as well as markers of the surrounding context. Leveraging the selected variables, we built spatial models to understand the variation in DoSC rates across census tracts. The findings were compared to those of the nonspatial regression and deep learning models. Model performance was evaluated from the root mean squared error (RMSE) using nested 10-fold cross-validation. Feature importance was evaluated by computing the increment of the RMSE when a single variable was shuffled within the data set.

Results: Data collection yielded sets of 463 census tracts at CCHMC (DoSC rates 1.2%-12.5%) and 1024 census tracts at TCH (DoSC rates 3%-12.2%). For CCHMC, an L2-normalized generalized linear regression model achieved the best performance in predicting all-cause DoSC rate (RMSE 1.299%, 95% CI 1.21%-1.387%); however, its improvement over others was marginal.

For TCH, an L2-normalized generalized linear regression model also performed best (RMSE 1.305%, 95% CI 1.257%-1.352%). All-cause DoSC rate at CCHMC was predicted most strongly by *previous no show*. As for community-level data, the proportion of African American inhabitants per census tract was consistently an important predictor. In the Texas area, the proportion of overcrowded households was salient to DoSC rate.

Conclusions: Our findings suggest that geospatial analysis offers potential for use in targeting interventions for census tracts at a higher risk of cancellation. Our study also demonstrates the importance of home location, socioeconomic disadvantage, and racial minority status on the DoSC of children's surgery. The success of future efforts to reduce cancellation may benefit from taking social, economic, and cultural issues into account.

(J Med Internet Res 2021;23(9):e26231) doi:10.2196/26231

KEYWORDS

surgery cancellation; socioeconomic factors; spatial regression models; machine learning

Introduction

Background

Surgical interventions, along with other diagnostic and therapeutic procedures performed under anesthesia, can deliver significant health benefits; it has been estimated that 30% of the global burden of disease is treatable by surgery [1]. Unlike most drug and nonprocedural therapies that require ongoing adherence for maximal benefit, surgery is typically delivered at a single encounter with a complex multidisciplinary health care team. Therefore, barriers to compliance with surgery are likely to be different from barriers to chronic treatment compliance.

Cancellation is an important barrier to the successful delivery of surgical therapy and is reported to affect between 2% and 20% of the 50 million procedures performed annually in American hospitals [2,3]. Cancellation has become a focus of interest at children's hospitals in view of its substantial negative repercussions for patients, families, and institutions. First, if surgery is canceled, the child fails to receive therapeutic or diagnostic benefits. Taking as an example the most common surgery in childhood, insertion of *ear tubes*, confirmed by a systematic review, reduces hearing loss in children with otitis media with effusion [4]. Cancellation of, or even delay in, insertion of ear tubes may thus impair language or speech development and affect behavioral, cognitive, or quality of life outcomes. For families, surgery cancellation leads to psychological stress and increased financial burden. As an illustration, researchers at another academic tertiary children's hospital, also in the American Midwest, found an average wasted round-trip of more than 160 miles for those who come for surgery but had to cancel [5]. This resulted in one-third of accompanying family members missing a day of work, which was unpaid in half of the cases. Parents and children expressed disappointment, frustration, and anger as a result of cancellation. From an institutional perspective, expensive staff and facilities costs are not reimbursed when surgeries are canceled. Even with a low 4.1% day-of-surgery cancellation (DoSC) rate at our hospital, potential lost revenue exceeds US $3000 per hour for operating room billing alone, with more than 5 hours lost per day, costing over US $2 million per year [6].

In our previous work, we found that up to 85% of cancellations may be amenable to modification of patients' and families' behaviors. We undertook a quality improvement project that reduced cancellations by delivering interventions *across the board* to all scheduled patients [6]. In preparation for subsequent improvement efforts, we sought predictors of cancellation to gain insight into its etiology and with the aim of targeting future efforts more efficiently [7].

To date, most studies on surgery cancellation applied classical statistical techniques to demonstrate association [8-14], whereas few studies have used machine learning to predict surgical cases at risk of cancellation [7,15,16]. Most recently, we developed machine learning–based approaches to identify *individual surgery cases* at high risk of DoSC from patient-specific and contextual data from 2 distinct pediatric surgical sites of Cincinnati Children's Hospital Medical Center (CCHMC), offering the promise of targeted interventions [7]. At the conception of this study, we hypothesized that the risk of DoSC at our tertiary children's hospital varies according to the location of the patient. We sought *geographical clusters* of high and low cancellation rates and used these findings to explore the underlying social determinants.

In recent years, increasing volumes of geospatial data have become publicly available, including from censuses, cataloging crimes, and relating to a variety of social and economic processes. As, in spatial data sets, observations may not be independent (spatial autocorrelation) or the relationships between variables may vary across geographical space (spatial nonstationarity), assumptions underlying conventional statistical modeling approaches may be violated. For these reasons, specialized methodologies have been developed for geospatial analysis to explain the spatial patterns of human behavior and the underlying factors that contribute to or explain these patterns [17]. In addition, geospatial models may capture spatial patterns (eg, spatial clusters) during model construction to achieve a better fit [18-20]. Spatial autocorrelation measures quantify the correlation of a variable with itself through geographical space [21]. A mixed regressive spatial autoregressive (SAR) model combines an autoregressive structure with a conventional regression model by assuming that a regional outcome is also impacted by outcomes from its neighbors [18], whereas a spatial error model (SEM) combines a conventional regression model with spatially autoregressive disturbances [22,23]. Spatial moving average models are similar to SEMs but use a different autocorrelation structure to represent the disturbances. Spatial Durbin models extend the SAR framework by hypothesizing

that a regional outcome is additionally impacted by predictors from the neighborhoods that comprise it [18,23]. Similarly, an SAR confused model is a generalized SAR model that accommodates spatial dependence in both the outcome variable and error terms [18]. A spatial Durbin error model includes exogenous interaction effects in addition to interaction effects among error terms [23].

Geospatial modeling approaches have been applied to a wide range of public health problems, such as the estimation of mortality or of air pollution effects or identification of causal factors in disease [24-26]. However, such spatial analytical techniques have not been used to study the geography of surgery cancellation. In particular, community characteristics provide vital information that is lacking when using individual-level data in isolation. Therefore, in this study, we explore a variety of established geospatial models to identify and visualize spatial trends in cancellation rates and candidate predictors. The findings were compared with those of the generalized linear regression model (GLM) and deep learning model.

Objectives

This study presents a geospatial analysis of patient-specific variables from the electronic health records (EHRs) of CCHMC and of Texas Children's Hospital (TCH) as well as of socioeconomic factors measured at the census tract level. We use the data to understand the potential contributors to spatial variation in the cancellation rates of pediatric surgery. We hypothesize that there are marked disparities in DoSC rates across neighborhoods. To our knowledge, this is the first study to investigate the geographic variation of pediatric surgery cancellation rates. The long-term objective is to understand contributors underlying DoSC and barriers experienced by patients and their families so that support can be focused efficiently toward families who are both in need and are most likely to benefit.

Methods

Data

A 5-year geocoded data set (May 2011-May 2016) of 88,013 surgeries including 3702 (4.21%) DoSCs corresponding to patients living in the hospital's primary service area was extracted from the CCHMC EHR. The data set included 2 primary surgical sites (main and Liberty campuses) of the institution that covers 472 census tracts in Greater Cincinnati. An equivalent set of 166,533 surgeries over 5.7 years with 10,236 (6.14%) DoSCs was extracted from the TCH EHR for validation. The data set included 3 primary surgery sites (Texas Medical Center, West Campus, and the Woodlands) of the institution that serves 1065 census tracts in Greater Houston. Ethics approval for this study was provided by the CCHMC institutional review board (study ID 2018-4568). Both CCHMC and TCH are urban, pediatric academic medical centers that function as the primary pediatric facilities for their surrounding metropolitan areas and also accept many tertiary and quaternary care referrals from elsewhere. All home locations were geocoded with an in-house geographic information system to ensure that no protected health information was sent outside the institution. Owing to high address matching accuracy, 90.2% (229,600/254,546) of the locations were geocoded at the city-block level (ie, a group of buildings surrounded by streets), and a further 6.8% (17,309/254,546) were geocoded at street level (ie, center of the matched street). For all surgical activities at CCHMC, cancellations are comprehensively adjudicated to one of 10 reason codes by clinical staff at the time of cancellation, thus allowing analysis for specific causes, including acute patient illness, failure to attend surgery (*no show*), failure to comply with eating and drinking instructions (*nil per os* [*NPO*] *violation*), and refusal to undergo surgery by either patient or family. For CCHMC, rescheduled cases were defined as completed surgeries with prior cancellations of similar case length within the preceding 90 days for the same individual. At TCH, in the absence of estimated case length data, rescheduled cases were determined by the procedure name and service department. All rescheduled cases (CCHMC: 1578/88,013, 1.79%; TCH: 4077/166,533, 2.45%) were excluded from the analysis to avoid diluting the effects of cancellation predictors by subsequently completed surgeries.

EHR variables for individual surgery cases, including recent health care use, schedule-related factors, prior cancellation behaviors, and information from a preoperative telephone call, were extracted as previously described [7] and spatially aggregated at the census tract level (Table 1). The estimated driving time from patients' homes to surgical sites was calculated and categorized into 6-minute intervals (*>60* for locations farther than 1 hour away) using the DeGAUSS R package (Cole Brokamp) [27]. Socioeconomic factors were obtained from the US Census Bureau's 2011-2015 American Community Survey (ACS) 5-year estimates, which provide data at the level of individual census tracts [28]. ACS variables were selected for practical relevance to successful preparation and attendance for surgery, including those relating to poverty, home ownership, household vehicle availability, housing (vacancy, value, and crowding), marriage, educational attainment, population density, linguistic isolation, African American (*Black*) race, and Hispanic heritage (Table 2). Census tract population density was computed as the ratio of the population to the total land area.

Table 1. Case-based variables from institutional electronic health records. (N=14).

Category	Variables, n (%)	Description
Transportation	1 (7)	Driving time from home to the surgical site
Preoperative phone call	1 (7)	Number of call attempts
Recent health care use	5 (35)	Number of recent emergency room attendance (two 2 time points), number of medications taken regularly at home before surgery, office visits, and hospitalizations in the previous 6 months
Prior cancellation behaviors	5 (35)	Numbers of previous cancellations, previous *no shows*, previous other cancellations, clinic *no shows*, and previous surgeries
Surgery related factors	2 (14)	Lead time and estimated case length

Table 2. American Community Survey data from the US Census Bureau measured at the census tract level.

ACS[a] table code	Table description	Extracted data
B02001	Race	Black or African American race
B03003	Hispanic or Latino origin	Hispanic or Latino heritage
B17012	Poverty status of families by household type by number of related children aged <18 years	Families in poverty
B15002	Sex by educational attainment for the population aged ≥25 years	Population with low educational attainment
B16002	Language spoken at home and ability to speak English	Linguistic isolation
B06008	Place of birth by marital status in the United States	Adults never married
B08201	Household size by vehicles available	No car in household
B25003	Residential tenure	Rented houses
B25077	Median home value (US $)	Median home value
B19125	Median family income in the past 12 months by the presence of own children aged <18 years	Median household income
B25002	Residential occupancy status	Vacant houses
B25014	Tenure by occupants per room	Household overcrowding
B01003	Census tract total population	Total population

[a]ACS: American Community Survey.

Spatial Autocorrelation

In geospatial analysis, it is important to assess the spatial independence of variables before model construction. Spatial autocorrelation measures describe the degree of spatial dependence or patterns for a variable across a spatial area [21]. We used the global Moran I statistic to test spatial independence for the DoSC rate and extracted variables [29]. Moran I values with significant *P* values ($P<.05$) indicate that the values for a variable are either spatially clustered (positive Moran I value) or dispersed (negative Moran I value), whereas there is no spatial dependence of the variable if the *P* value is not significant.

Data Processing

DoSCs resulting from CCHMC's top four most frequent patient-related cancellation reasons (ie, patient illness, *no show*, NPO violation, and patient or family refusal) were considered as canceled cases (denoted as all-cause cancellation) [7]. Census tracts without inhabitants (eg, for CCHMC, the census tract corresponding to the Cincinnati or Northern Kentucky International Airport—GEOID 21015980100) or with less than 20 surgical cases were excluded a priori. The rate of DoSC was calculated per census tract for the primary service area of the 2 hospitals (463 for CCHMC and 1024 for TCH) with empirical Bayesian shrinkage toward a beta before lessening the influence of sparsely populated tracts with few patients [30]. The corresponding rates for common patient-related cancellation reasons were also computed for CCHMC (but not for TCH for which such categorization was not available). For each census tract, the rates of the categorized driving times were similarly computed using empirical Bayesian estimation. The most common category of driving time was used as the base category to avoid the linear dependencies induced between the features. Missing ACS values (ie, median home value: 0.9% (4/463) of missing values for CCHMC and 1.66% (17/1024) of missing values for TCH; median household income: 6.7% (31/463) of missing values for CCHMC and 2.8% (29/1024) of missing values for TCH) were imputed using grand mean and mode imputation. The median home value and the median household income were categorized based on information from the US Census Bureau website [31-34]. All variables based on percentages were rescaled based on IQR to aid the interpretation of regression models [35]. The collinearity among variables was tested using the variance inflation factor (VIF) [36].

Variables not exceeding a predefined threshold of collinearity (VIF<10) were included in the model construction. Finally, EHR and ACS variables without evidence of significant geospatial clustering were excluded from the data set.

Spatial Weights

Spatial weight matrices summarize the spatial relations between the census tracts. Neighboring tracts were determined by sharing at least one boundary edge. Inverse distance weighting was applied to compute spatial weights [37], and weight matrices were then standardized by row, such that the sum of spatial weights for each census tract equals 1.

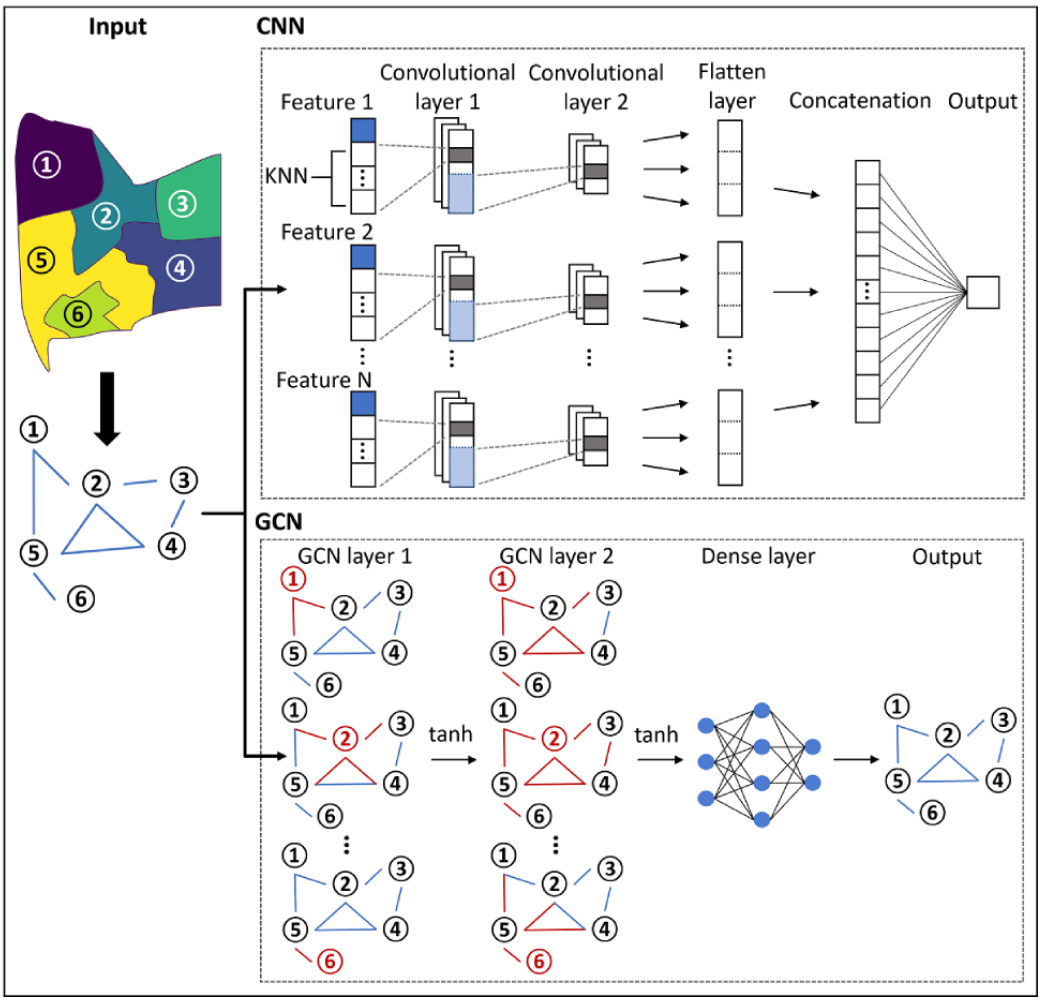
Conventional Regression Models

We modeled the prediction of the DoSC rate for each census tract as a supervised regression problem and tested both nonspatial regression models (GLM, L2-normalized GLM, support vector machine with polynomial kernels [SVM-P], and decision tree) and spatial regression models including SAR model, spatial Durbin model, SEM, spatial Durbin error model, spatial moving average, and SAR confused models [18,22,23,38-40]. Appropriate variants of the spatial regression models, such as the L2-normalized SAR models, were also implemented. We used these models to allow for the possibility of spatial impact on a census tract by neighboring tracts. Regression models were implemented using packages for the R programming language (R Foundation for Statistical Computing) [41].

Deep Learning Models

In addition to traditional regression models, we implemented convolutional neural networks (CNNs) and graph convolutional networks (GCNs) to allow for the possibility of nonlinear relationships between the variables and DoSC rates [42,43]. Figure 1 illustrates the structures of the CNN and GCN models. For CNNs, targeted census tracts with their K-nearest neighbors (K=5, 10, 15, and 20) were used to construct feature vectors, each of which was trained by a 2-layer 1D CNN. The concatenated output was then used to predict the DoSC rates for individual census tracts. For GCNs, an adjacency matrix (eg, inverse distance weighting matrix) representing the graph structure and a feature matrix were taken as inputs and a framework with two layers of GCN and 1 layer of fully connected neural network to predict DoSC rates. Deep learning models were implemented using TensorFlow (version 2.2) for Python [44,45].

Figure 1. Model development for deep learning models. CNN: convolutional neural networks; GCN: graph convolutional networks; KNN: K-nearest neighbor.



Experimental Setup

Owing to the relatively low number of census tracts (sample size) in the data sets, a nested 10-fold cross-validation (10 folds for both the outer and inner loops) was adopted. The approach randomly split the data set into 10 rotating subsets: 9 for model training and hyperparameter tuning and 1 for testing at each run. At each iteration, an inner cross-validation was applied to the 9 folds of training data to tune a model's hyperparameters with grid search parameterization [46], including (1) cost parameters for L2-normalized GLM [38], L2-normalized SAR model [18], and SVM-P [39] (screened from 10^{-6} to 10^{6}); (2) optimal degree for SVM-P (screened from 1 to 3); (3) minimum number of observations in a node (3, 5, 10, 15, and 20) and the complexity parameters (screened from 10^{-6} to 10^{-1}, 0.3, 0.5, and 0.8) for decision tree [40]; (4) learning rates of an optimizer parameter (screened from 10^{-3} to 10^{-1}) for deep learning models [42,43]; (5) filter size of kernels (2 and 3) for 1D CNN layers [42]; and (6) dimensionality of output space (40, 50, 60, 75, and 90) for 2 GCN layers [43]. The model with optimal hyperparameters was then trained on all 9 folds of data and evaluated on the hold-out subset. The process was repeated 10 times to cover all subsets, and the evaluation score was averaged across the subsets as the final performance of the model. Model selection was based on the performance in the outer loop of the nested cross-validation. For models without hyperparameters, a 10-fold cross-validation was performed using the same folds as used in the outer loop of nested cross-validation. To assess the validity of the geospatial analysis, we compared the model performances with those aggregated from individual DoSC predictions. The best-performing gradient boosted logistic regression model, with 58 EHR variables developed in our previous study, was applied to the surgical cases located within the studied census tracts to predict DoSC [7]. The individual-level predictions were then aggregated to predict the DoSC rate per census tract (denoted as the individual-prediction-aggregation model).

Evaluation Metrics

Model performance was evaluated by root mean square error (RMSE), which is a commonly used evaluation metric for numerical predictions in regression analysis [47]. The spatial autocorrelation of model residuals was assessed using the global Moran I statistic for indications that the model was misspecified. Geographically, weighted Pearson correlation was used to mark census tracts with significant local correlations between observed and predicted cancellation rates [48]. A permutation-based technique was used to calculate the feature importance scores using the DALEX (Model Agnostic Language for Exploration and Explanation) R package (Przemyslaw Biecek) [49]. The importance of every variable was measured by computing the increment of the RMSE when a single variable was shuffled within the data set.

Results

Descriptive Statistics for the Data Sets

Among 86,435 CCHMC surgical cases meeting the analysis selection criteria, the overall all-cause DoSC rate was 3.76% (3255). Patients lived in 472 different census tracts within the primary service area of the CCHMC. Of the 472 census tracts, 9 (1.9%) contributed to a few surgical cases (<20 cases) and were excluded from the analysis. Of the 463 remaining, the 97.5th percentile for cancellation rate was 9.4%, but 10 tracts (2.2%) had no cancellations. The TCH data set contained 166,533 surgical cases over 5.7 years, corresponding to 1065 different census tracts, with a DoSC rate of 6.14% (10,236/166,533). After similar preprocessing, 162,026 surgery cases and 1024 census tracts, with a median DoSC rate of 6.2%, were included in the analysis.

Of note, the 463 census tracts within the CCHMC study area had a median population of 3987 (IQR 2668, maximum 20,188), with a total population of 2 million. The 1024 Houston area census tracts encompassed a population of 6.38 million (median 5342 per tract). The relatively small population per census tract supports that, although the ACS provides aggregate statistics, these represent features of the locale and community immediately adjacent to patients' homes with a high degree of spatial granularity. In support of this assertion, the proportion of African American patients in each census tract from the CCHMC EHR data is closely associated with the equivalent proportion of the general population in the ACS data (R^2=0.89). This finding underscores the validity of using ACS variables as surrogates for individual patients' socioeconomic and minority status, in addition to their characterization of the surrounding context.

Figure 2 depicts the geospatial variation in DoSC rate by census tract of home location in the Greater Cincinnati and Houston regions. Enlarged maps are presented in Figures S1 and S2 in Multimedia Appendix 1, and interactive maps are presented in Multimedia Appendix 2 [50]. For CCHMC, tracts with increased cancellation risk clustered mainly in the most populous urban areas. Tracts with lower cancellation risk were located in suburban and rural locations. However, in the Houston area (Figure 2), high- and low-canceling census tracts were more geographically dispersed. These visual impressions are supported by the global Moran I as a measure of spatial autocorrelation (Figure 3).

Figure 2. Geospatial distributions of day-of-surgery cancellation rate in the primary service areas of (A) Cincinnati Children's Hospital Medical Center (2011-2016) and (B) Texas Children's Hospital (2012-2017). CCHMC: Cincinnati Children's Hospital Medical Center; TCH: Texas Children's Hospital.

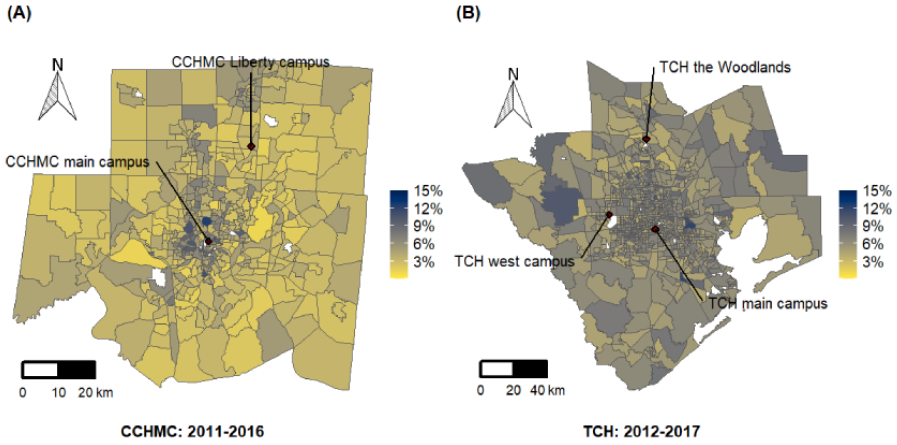
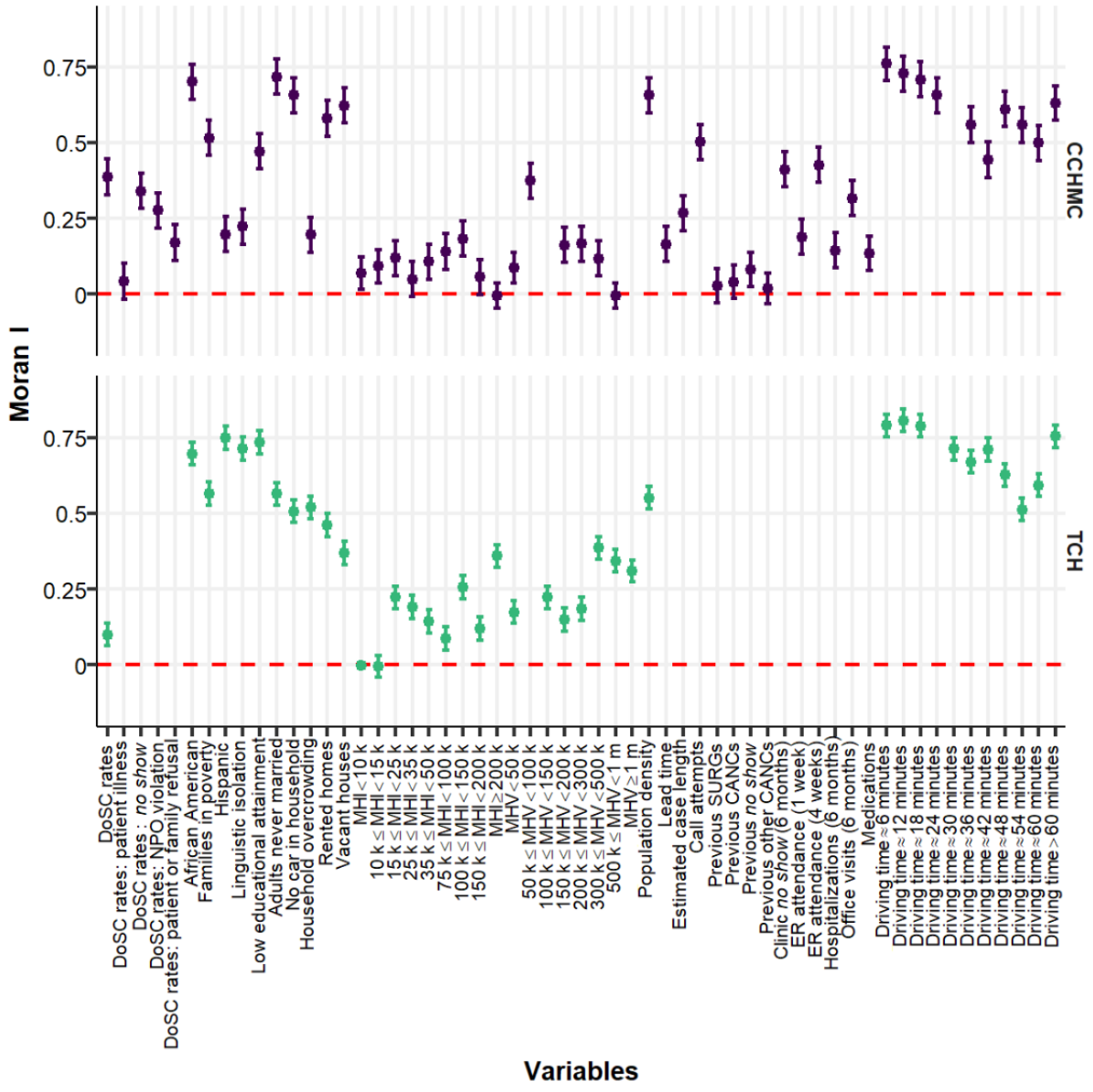


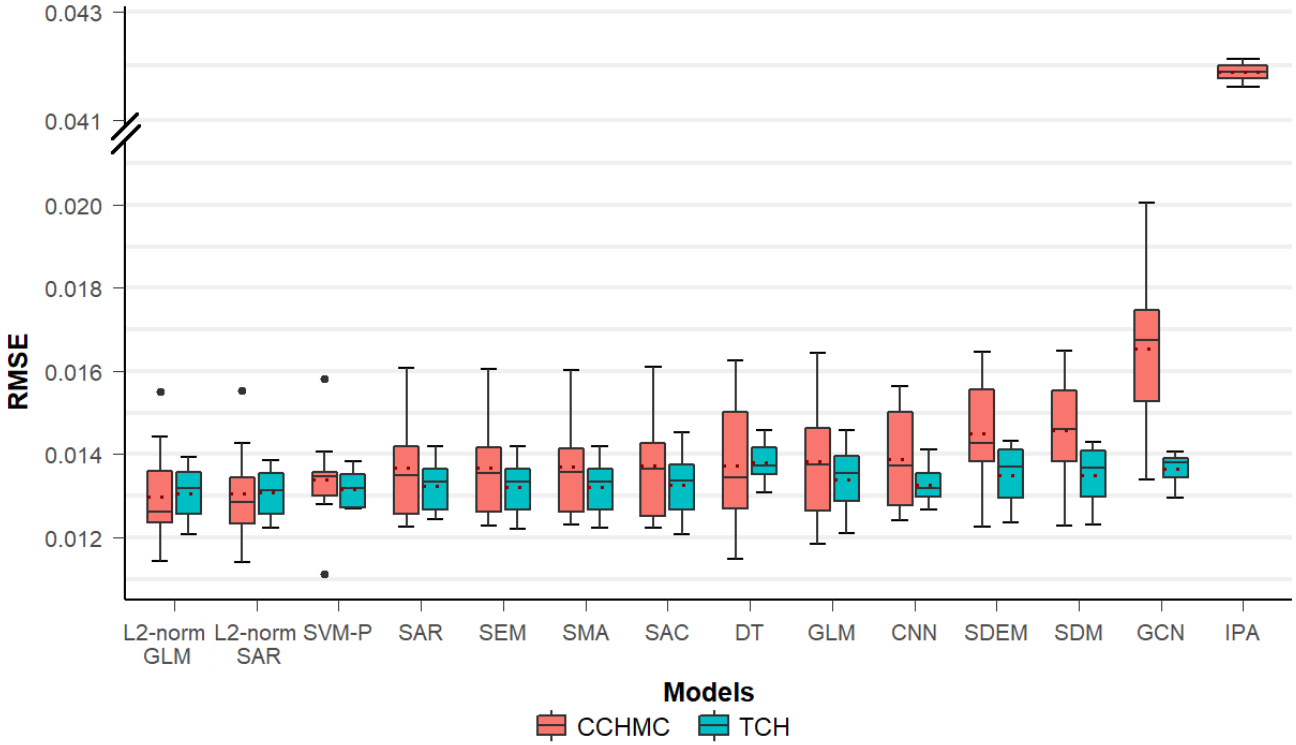
Figure 3. Spatial autocorrelation of day-of-surgery cancellation rate, case-based electronic health record variables and American Community Survey variables for Cincinnati Children's Hospital Medical Center and Texas Children's Hospital measured by global Moran I with 95% CI. CANC: cancellation; CCHMC: Cincinnati Children's Hospital Medical Center; DoSC: day-of-surgery cancellation; ER: emergency room; MHI: median household income; MHV: median house value; NPO violation: failure to comply with eating and drinking instructions; SURG: surgery; TCH: Texas Children's Hospital.



Predicting DoSC Rates With Conventional Regression and Deep Learning Models

In preparation for model development, we tested for collinearity among case-based EHR variables and ACS independent variables using VIF. None of the variables exceeded a predefined threshold of collinearity (VIF<10), which supports their independence. Certain variables were excluded from model development because they had insignificant Moran I values in the spatial autocorrelation analysis (Figure 3). Specifically, these variables excluded for low spatial variation were numbers of previous surgeries (*P*=.13), previous cancellations (*P*=.06), previous non–*no show* cancellations (*P*=.19), and patient illness (*P*=.07). The exclusion resulted in 46 and 37 numerical variables for model construction for CCHMC and TCH, respectively. Figure 4 presents the performance of the regression and deep learning models in predicting the all-cause DoSC rates. The lowest (best) RMSE was generated by the L2-normalized GLM at 1.299% (95% CI 1.21%-1.387%) for the CCHMC data set. All models outperformed the individual-prediction-aggregation model (RMSE 4.189%, 95% CI 4.178%-4.201%). This finding was statistically significant (*P*<.001). For the TCH data set, L2-normalized GLM also achieved the best performance, yielding an RMSE of 1.305% (95% CI 1.257%-1.352%).

Figure 4. Model performance for predicting day-of-surgery cancellation rates at census tract level for Cincinnati Children's Hospital Medical Center and Texas Children's Hospital. Dashed line in each box represents the mean value of root mean squared error. CCHMC: Cincinnati Children's Hospital Medical Center; CNN: convolutional neural networks; DT: decision tree; GCN: graph convolutional networks; GLM: generalized linear regression model; IPA: individual-prediction-aggregation model; RMSE: root mean square error; SAC: spatial autoregressive confused model; SAR: spatial autoregressive model; SDEM: spatial Durbin error model; SDM: Spatial Durbin models; SEM: spatial error model; SMA: spatial moving average; SVM-P: support vector machine with polynomial kernels; TCH: Texas Children's Hospital.



To augment our understanding of potential cancellation causes, we used DALEX [49] to identify the most important predictors in the best-performing L2-normalized GLMs (Figures 5 and 6). All-cause cancellation at CCHMC was predicted most strongly by the variable *previous no show* (Figure 5). Turning to community-level variables, the proportion of African American inhabitants per census tract is important. However, among community-level factors in the Houston area, the proportion of overcrowded households showed the strongest association with surgery cancellation rate (Figure 6). The median household income was also predictive, whereas spatial models highlighted the importance of clustered neighborhoods with low educational attainment (Figure S3 in Multimedia Appendix 1).

Figure 5. Feature importance generated from the best-performing L2-normalized generalized linear regression model for the Cincinnati Children's Hospital Medical Center data set. Variables were ranked in descending order based on their importance in predicting all-cause day-of-surgery cancellation rates. ER: emergency room; MHI: median household income; MHV: median house value; NPO: nil per os; RMSE: root mean squared error.

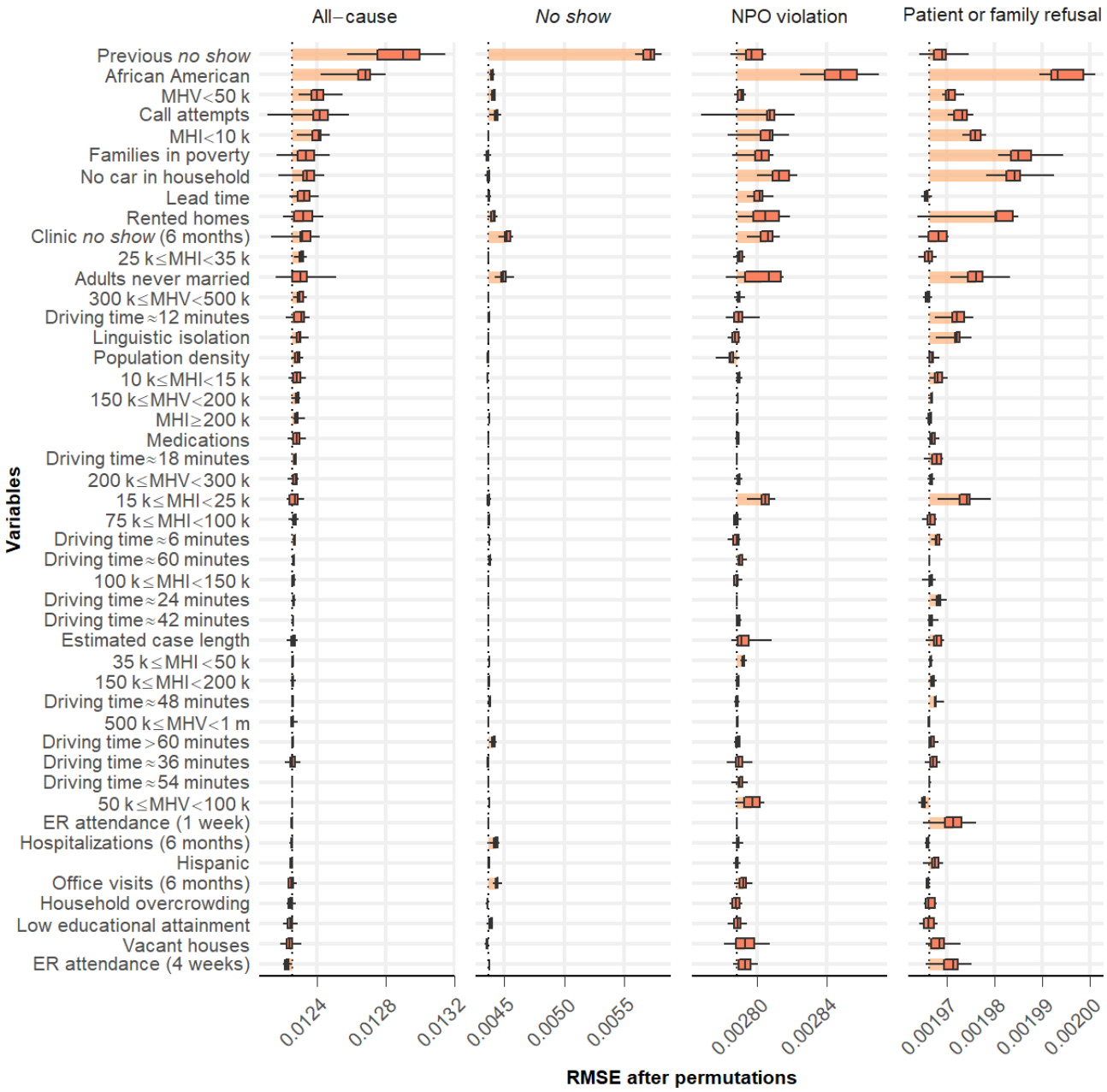
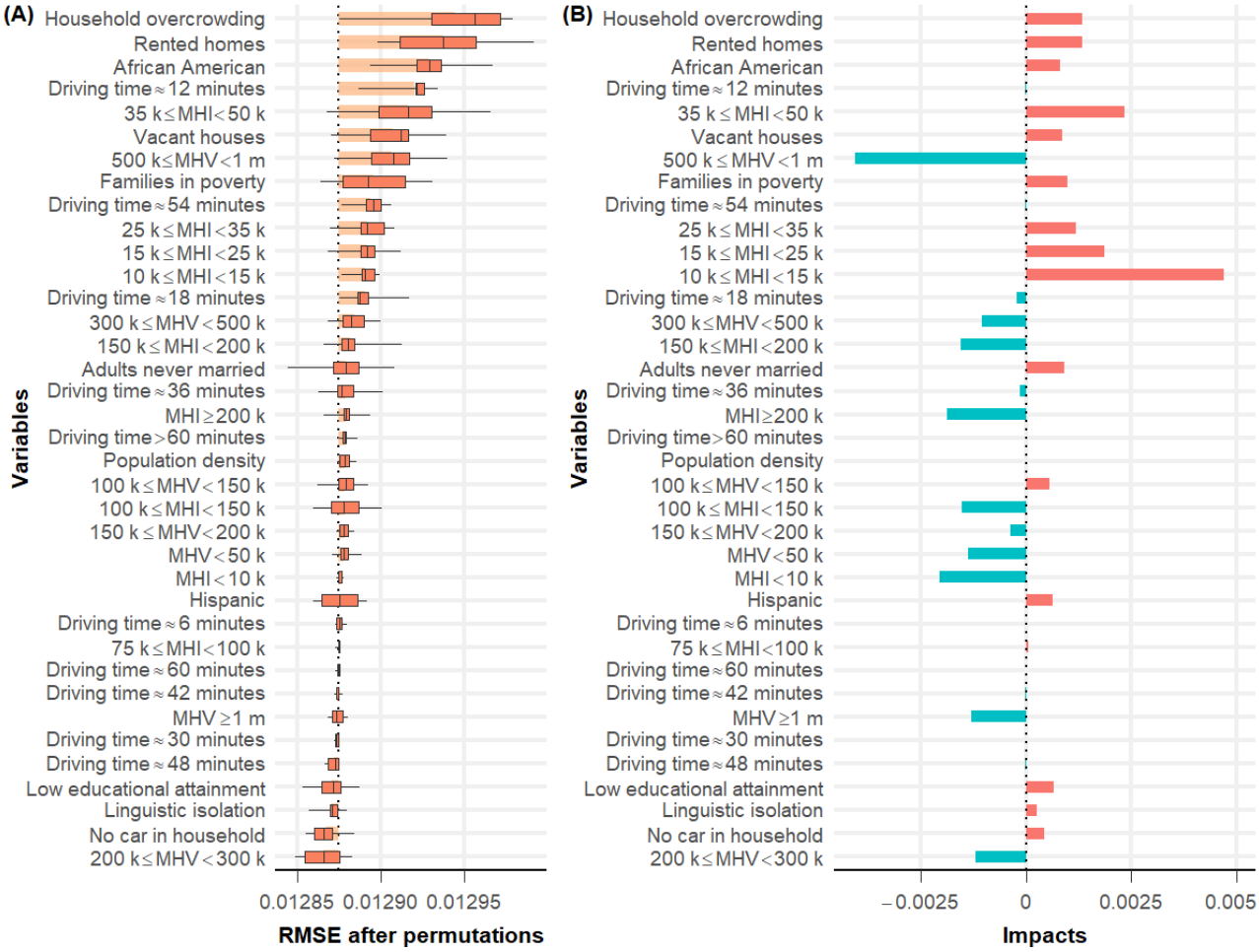
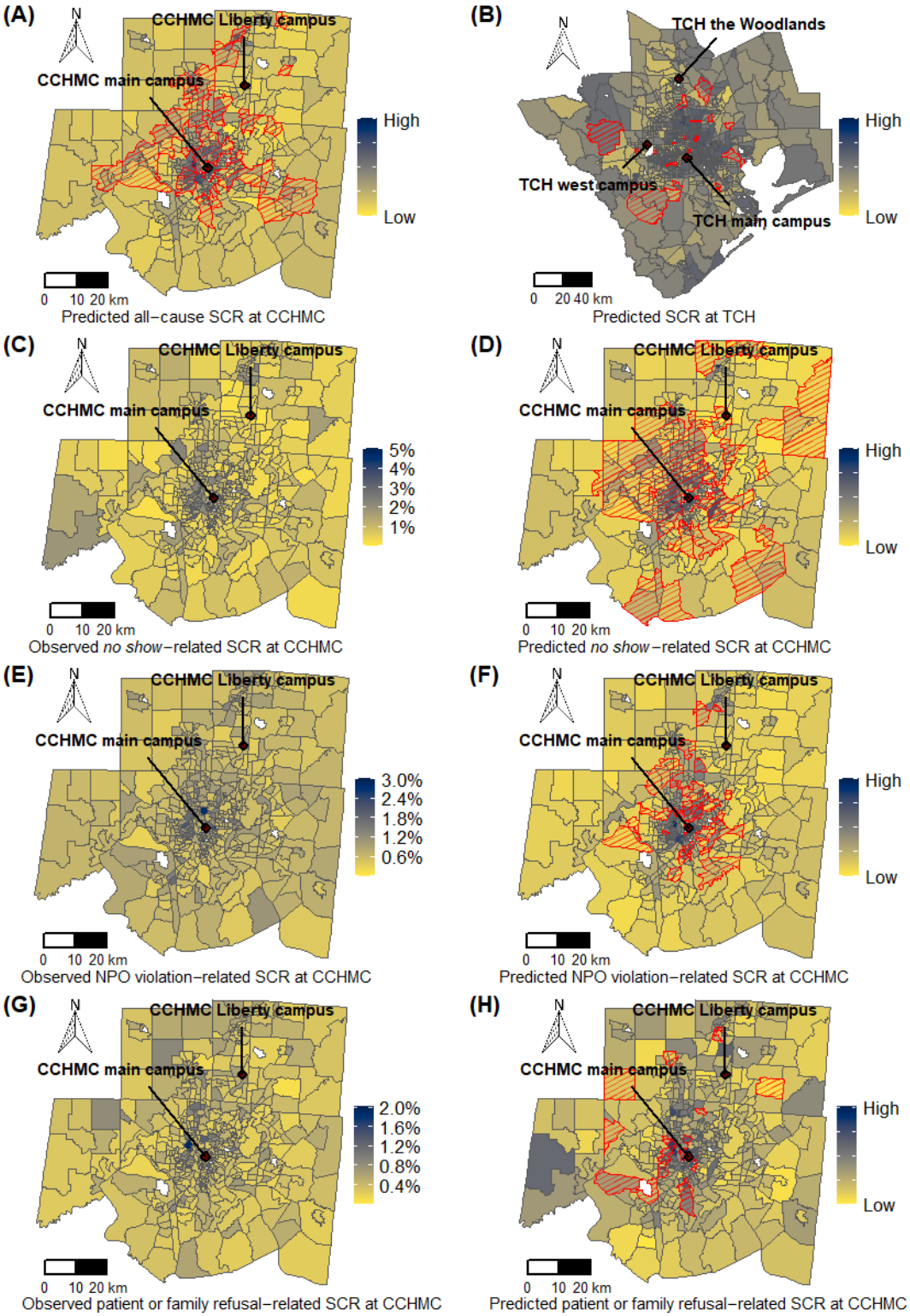


Figure 6. (A) Feature importance and (B) impacts generated from the best-performing L2-normalized generalized linear regression model for the Texas Children's Hospital data set. Variables were ranked in descending order based on their importance in predicting day-of-surgery cancellation rates. MHI: median household income; MHV: median house value; RMSE: root mean squared error.



The predicted DoSC rates at the census tract level from the best-performing models for CCHMC and TCH are compared with the actual data in the maps presented in Figure 7. Areas in proximity to hospital locations and with larger populations showed a higher correlation between the observed and predicted DoSC rates. Figure 7 shows the observed and predicted DoSC rates for specific cancellation causes, as detailed in the next section (enlarged maps in Figures S4-S11 in Multimedia Appendix 1). The interactive map versions of Figure 7 are presented in Multimedia Appendix 2 [50].

Figure 7. Comparison of surgery cancellation rates at census tract level predicted by the best-performing model to actual data. Census tracts with significant local correlation between the observed and predicted values are marked with red borders and cross-hatching. (A) Predicted all-cause surgery cancellation rate at Cincinnati Children's Hospital Medical Center; (B) predicted surgery cancellation rate at Texas Children's Hospital; (C) observed no show–related surgery cancellation rate at Cincinnati Children's Hospital Medical Center; (D) predicted no show–related surgery cancellation rate at Cincinnati Children's Hospital Medical Center; (E) observed nil per os violation-related surgery cancellation rate at Cincinnati Children's Hospital Medical Center; (F) predicted nil per os violation-related surgery cancellation rate at Cincinnati Children's Hospital Medical Center; (G) observed patient or family refusal–related surgery cancellation rate at Cincinnati Children's Hospital Medical Center; (H) predicted patient or family refusal–related surgery cancellation rate at Cincinnati Children's Hospital Medical Center. CCHMC: Cincinnati Children's Hospital Medical Center; NPO: nil per os; TCH: Texas Children's Hospital; SCR: surgery cancellation rate.



Predicting Specific Cancellation Codes

Cancellations were coded by reason in the CCHMC data set. The four most frequent DoSC reasons account for more than 87% of the cases. These are patient illness (40%), nonattendance (*no show*; 20%), concern for aspiration risk because of noncompliance with preoperative *NPO* instructions (NPO violation; 18%), and patient or family refusing surgery after scheduling (10%) [6]. Cancellation because of patient illness was not analyzed using a spatial approach because there was no evidence of significant geospatial clustering (Figure 3). Spatial risk analyses and predicted cancellation rates generated by the best-performing models (Figure S12 in Multimedia Appendix 1) for the other individual cancellation codes are presented in Figure 7. Models using EHR data and ACS data aggregated at census tract level achieved better performance than individual-prediction-aggregation models (RMSEs of 1.124%, 0.868%, and 0.517% for *no show*–related, *NPO violation*–related, and *patient or family refusal*–related surgery cancellations, respectively). Of the 3 individual cancellation causes examined in this way, *no show* showed better geographically weighted Pearson correlation between observed and predicted DoSC rates than NPO violation or patient or family refusal (Figure 7). The key features for the specific cancellation reasons are shown in Figure 5. Prior cancellation behaviors, including the number of previous *no show* cancellations and clinic appointment *no shows* in the previous 6 months were predictive for day-of-surgery *no show* cancellation rates. Among community-level factors, the proportion of never married adults showed an association with *no show* cancellation rates. For both NPO violation–related and patient or family refusal–related cancellation rates, the proportion of African Americans per census tract, as well as ACS markers of poverty (including the proportions of households without a car, families in poverty or renting homes, median household income, and median house value) were salient. Linguistic isolation was predictive of the rates of cancellation because of patient or family refusal.

Discussion

Principal Findings

This study aims to understand the potential contributors to disparities in DoSC rates across neighborhoods. At 2 different tertiary children's hospitals, we found marked geographic variation, particularly for cancellations coded as unrelated to patient illness. To understand this spatial variation, we developed models using case-based EHR data and ACS data aggregated at the census tract level. For the CCHMC data set, an L2-normalized GLM achieved the best performance in predicting the all-cause DoSC rate, but its improvement over the other regression models was marginal. The L2-normalized SAR model showed a comparable performance. The L2-normalized GLM performed better in urban areas around the CCHMC (Figure 7), possibly because of the larger number of surgical cases in these tracts. Interestingly, deep learning models did not offer improved predictive power, suggesting that geospatial impacts on DoSCs could be more regional and linear. A similar performance trend was observed for the TCH data set, suggesting the generalizability of our approach. Looking into the specific causes of cancellation at CCHMC, we found that patient illness (the most frequently recorded cause) did not show marked geographic variation. Of the 3 most frequent individual causes that showed spatial variation, *no show* was predicted better than NPO violation and patient or family refusal by the models.

Our geospatial analysis was helpful in identifying key factors, including potentially actionable predictors and underlying DoSCs at the census tract level. In the CCHMC data set (where the categorization of cancellation reasons was available from the EHR, unlike for TCH), the all-cause DoSC rate is composed of the top four most frequent patient-related cancellation causes (ie, patient illness, *no show*, NPO violation, and patient or family refusal). The key factors for all-cause cancellation reflect the average consensus for specific cancellation causes. The rate of prior *no show* cancellations by patients in a census tract best predicted the all-cause DoSC rate (Figure 5). The use of ACS data also provided granular and relevant information on community social and economic factors that were not available from the EHR, expanding the view of contextual factors likely influencing a family's preparations for their child's surgery [51]. We found that the proportion of African American inhabitants per census tract was predictive of geographic variation in the all-cause cancellation rate; that is, tracts with a higher proportion of African American inhabitants had higher DoSC rates. Important predictors for the 3 individual cancellation causes are discussed in detail below.

For the best model to predict *no show* cancellation rate, prior cancellation behaviors were of prime importance, including the number of previous *no show* cancellations and clinic *no show* in the previous 6 months (Figure 5). Both the GLM and SAR model suggested that every unit increase in the number of previous *no shows* was associated with a ~30% increase in the *no show* DoSC rate (Figures S13 and S14 in Multimedia Appendix 1). These patient- or family-level factors were complemented by a series of contextual variables that proved relevant. Indeed, for *no show* DoSCs, we found a similar link with the census tract proportion of adults who had never married, a potential surrogate for single parent–headed households. We speculate that such households experience more difficulty navigating day-to-day schedules. Working single parents may have trouble taking time off work. Those with multiple children may be challenged to ensure adequate coverage for their care. Obtaining appointments can be difficult; they may be especially difficult for those without a robust support structure [52].

Turning to NPO violation–related and patient or family refusal–related cancellations, we found commonality in key predictors. The proportion of African American inhabitants per census tract was predictive of both NPO violation–related and patient or family refusal–related DoSC rates. There are a variety of reasons that could underlie this finding—the reasons built atop analogous findings that highlight racial disparities across a range of health outcomes [53-56]. Racial segregation is strongly correlated with socioeconomic segregation in the United States. We found that, in addition to the proportion of African Americans within a tract, median household income, median house value, proportion of households without a car, and the proportion of families in poverty or renting homes were also

influential [57,58]. Thus, it is possible that the link between race and DoSC was largely driven by structural racism and the concomitant challenges that accompany the disproportionate economic disadvantage experienced by racial minorities (eg, lack of trust in the health care system, inflexible work schedules, transportation barriers, and competing priorities) [59-61]. Moreover, the proportion of adults who were never married was salient to both. Linguistic isolation was associated with higher rates of cancellation because of patient or family refusal (Figure 5). We, therefore, hypothesize that CCHMC's communications with urban, poor, minority, and non–English-speaking families leave room for improvement. Similar predictors (in this case for all-cause cancellation) were also identified using the TCH data set, including the proportion of African Americans per census tract and ACS markers of poverty (proportion of overcrowded households, families renting homes, low median household income, and low median house value; Figure 6).

Our analysis is novel for demonstrating that surgery cancellation is a source of inequity in surgical care and also for applying geospatial analysis to investigate barriers to care delivery. Furthermore, unlike the majority of studies in the literature on surgery cancellation, our study investigates patient- and family-related factors in the community using a geographic data set and offers insight into the underlying spatial risk factors and barriers experienced by families.

The findings of this study offer encouragement that geospatial analysis could appropriately be used to target interventions for patients living in census tracts with a higher rate of cancellations. In this way, support can be focused efficiently on families who are both in need and are most likely to benefit. Moreover, the specific predictors identified for individual cancellation codes may inform the design of interventions to address specific failure modes. For example, the association of cancellation risk with linguistic isolation argues for delivering preoperative communications using clear and simple language and for ready availability of interpreting services.

Arguably, our results uniquely reflect the characteristics of the Greater Cincinnati conurbation and its surrounding area, the pattern of referrals for surgery at CCHMC, and preoperative processes at this hospital. Thus, the findings may not be directly applicable to other locations. However, we hypothesize that similar social factors may determine cancellation patterns in other hospitals offering surgery for children, as evidenced by the similar performance and predictors observed in the TCH data set. In particular, the methodology used is likely to be transferable to other locales and institutions and to disparate aspects of health care delivery. With the easy availability of high-quality commercial or open-sourced geocoding software, our approach will be relatively easy to translate.

Limitations

We acknowledge that our study is limited in several ways. First, as an observational study, *exposures* to socioeconomic disadvantage or to racial minority status are not the only potential explanations for observed differences in cancellation rates. Second, the study relied on extracting patients' home addresses from the EHRs, which may be inaccurate (eg, outdated) or incomplete. In addition, cancellation of children's surgery likely depends on individual circumstances and perhaps seasonal factors. Finally, the relatively low number of census tracts (sample size) in the data sets might limit the application of complex models, such as deep learning.

Conclusions

This study aimed to conduct a geospatial analysis of patient-specific variables from EHRs and linked socioeconomic factors to understand the underlying contributors to disparities in DoSC rates across neighborhoods. Our findings demonstrate the importance of home location, socioeconomic disadvantage, and racial minority status on the last-minute cancellation of children's surgery. The success of future efforts to reduce cancellation may benefit from taking social, economic, and cultural issues into account. Although the original aim of this study was to drive improvement efforts, our results add further evidence of the importance of social determinants in children's health, including increased incidence and frequency of illness, barriers to accessing health care, and readmissions [62,63].

Authors' Contributions

LL developed the algorithms, ran the experiments, analyzed the results, created the tables and figures, and wrote the manuscript. YN assisted with the design of the study, provided suggestions for algorithm development, analyzed the results, and contributed to the manuscript. AFB and CB reviewed the errors and provided suggestions for the result analysis. RCR, LDH, and MKK coordinated the data extraction and provided suggestions for the analysis of the results. JNP conceptualized the study, coordinated the data extraction, preprocessed the data, analyzed the results, and wrote the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary figures of enlarged maps, feature importance, and model performance.
[DOCX File , 2832 KB - jmir_v23i9e26231_app1.docx]

Multimedia Appendix 2

Interactive maps.
[TXT File , 0 KB - jmir_v23i9e26231_app2.txt]

References

1. Shrime MG, Bickler SW, Alkire BC, Mock C. Global burden of surgical disease: an estimation from the provider perspective. Lancet Glob Health 2015 Apr;3:8-9. [doi: 10.1016/s2214-109x(14)70384-5]
2. Boudreau SA, Gibson MJ. Surgical cancellations: a review of elective surgery cancellations in a tertiary care pediatric institution. J Perianesth Nurs 2011 Oct;26(5):315-322. [doi: 10.1016/j.jopan.2011.05.003] [Medline: 21939884]
3. FastStats 2013. Centers for Disease Control and Prevention. URL: http://www.cdc.gov/nchs/fastats/inpatient-surgery.htm [accessed 2016-07-20]
4. Browning G, Rovers M, Williamson I, Lous J, Burton M. Grommets (ventilation tubes) for hearing loss associated with otitis media with effusion in children. Cochrane Database Syst Rev 2010:1801. [doi: 10.1002/14651858.cd001801.pub3]
5. Tait AR, Voepel-Lewis T, Munro HM, Gutstein HB, Reynolds PI. Cancellation of pediatric outpatient surgery: economic and emotional implications for patients and their families. J Clin Anesth 1997 May;9(3):213-219. [doi: 10.1016/s0952-8180(97)00032-9]
6. Pratap JN, Varughese AM, Mercurio P, Lynch T, Lonnemann T, Ellis A, et al. Reducing cancelations on the day of scheduled surgery at a children's hospital. Pediatrics 2015 May 13;135(5):1292-1299. [doi: 10.1542/peds.2014-2418] [Medline: 25869374]
7. Liu L, Ni Y, Zhang N, Pratap JN. Mining patient-specific and contextual data with machine learning technologies to predict cancellation of children's surgery. Int J Med Inform 2019 Sep;129:234-241 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.06.007] [Medline: 31445261]
8. Schuster M, Neumann C, Neumann K, Braun J, Geldner G, Martin J, et al. The effect of hospital size and surgical service on case cancellation in elective surgery. Anesth Analg 2013:1. [doi: 10.1213/ane.0b013e318222be4d]
9. Hand R, Levin P, Stanziola A. The causes of cancelled elective surgery. Qual Assur Util Rev 1990 Feb 03;5(1):2-6. [doi: 10.1177/0885713x9000500102] [Medline: 2136658]
10. Lacqua M, Evans J. Cancelled elective surgery: an evaluation. Am Surg 1994 Nov;60(11):809-811. [Medline: 7978670]
11. Argo JL, Vick CC, Graham LA, Itani KM, Bishop MJ, Hawn MT. Elective surgical case cancellation in the Veterans Health Administration system: identifying areas for improvement. Am J Surg 2009 Nov;198(5):600-606. [doi: 10.1016/j.amjsurg.2009.07.005] [Medline: 19887185]
12. Seim AR, Fagerhaug T, Ryen SM, Curran P, Saether OD, Myhre HO, et al. Causes of cancellations on the day of surgery at two major university hospitals. Surg Innov 2009 Jun 21;16(2):173-180. [doi: 10.1177/1553350609335035] [Medline: 19460816]
13. Al Talalwah N, McIltrot KH. Cancellation of surgeries: integrative review. J Perianesth Nurs 2019 Feb;34(1):86-96. [doi: 10.1016/j.jopan.2017.09.012] [Medline: 29678319]
14. Caesar U, Karlsson J, Olsson L, Samuelsson K, Hansson-Olofsson E. Incidence and root causes of cancellations for elective orthopaedic procedures: a single center experience of 17,625 consecutive cases. Patient Saf Surg 2014;8(1):24 [FREE Full text] [doi: 10.1186/1754-9493-8-24] [Medline: 24955115]
15. Luo L, Zhang F, Yao Y, Gong R, Fu M, Xiao J. Machine learning for identification of surgeries with high risks of cancellation. Health Informatics J 2018 Dec 05;26(1):141-155 [FREE Full text] [doi: 10.1177/1460458218813602] [Medline: 30518275]
16. Zhang F, Cui X, Gong R, Zhang C, Liao Z. Key experimental factors of machine learning-based identification of surgery cancellations. J Healthc Eng 2021 Feb 20;2021:6247652. [doi: 10.1155/2021/6247652] [Medline: 33688420]
17. Illian J, Penttinen A, Stoyan H, Stoyan D. Statistical Analysis and Modelling of Spatial Point Patterns. Chichester: John Wiley & Sons Ltd; 2008.
18. Haining R. Spatial Data Analysis in the Social and Environmental Sciences. Cambridge: Cambridge University Press; 1990.
19. Bivand R. Spatial econometrics functions in R: classes and methods. J Geograph Syst 2002 Dec 1;4(4):405-421. [doi: 10.1007/s101090300096]
20. Cressie N. Statistics for Spatial Data. New York: John Wiley & Sons, Inc; 1993.
21. Moran PA. Notes on continuous stochastic phenomena. Biometrika 1950;37(1-2):17-23. [doi: 10.1093/biomet/37.1-2.17]
22. Anselin L, Bera A. Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Handbook of Applied Economic Statistics. Boca Raton: CRC Press; 1998:237-289.
23. Anselin L. Spatial Econometrics: Methods and Models. Dordrecht: Springer Science & Business Media; 1988.
24. Lee D, Shaddick G. Spatial modeling of air pollution in studies of its short-term health effects. Biometrics 2010 Dec;66(4):1238-1246. [doi: 10.1111/j.1541-0420.2009.01376.x] [Medline: 20070295]
25. Owusu-Edusei K, Bohm MK, Chesson HW, Kent CK. Chlamydia screening and pelvic inflammatory disease: insights from exploratory time-series analyses. Am J Prev Med 2010 Jun;38(6):652-657. [doi: 10.1016/j.amepre.2010.02.008] [Medline: 20494242]
26. Sparks PJ, Sparks CS. An application of spatially autoregressive models to the study of US county mortality rates. Popul Space Place 2009 Jun 15;16(6):465-481. [doi: 10.1002/psp.564]

27. Brokamp C. DeGAUSS: Decentralized geomarker assessment for multi-site studies. J Open Source Softw 2018 Oct;3(30):812. [doi: 10.21105/joss.00812]
28. 2011-2015 American Community Survey 5-year estimates. U.S. Census Bureau. 2016. URL: https://www.census.gov/programs-surveys/acs/technical-documentation/table-and-geography-changes/2015/5-year.html [accessed 2021-03-25]
29. Cliff A, Ord J. Spatial Processes: Models and Applications. London: Pion; 1980:1-260.
30. Robbins H. An empirical bayes approach to statistics. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability. 1956 Presented at: The Third Berkeley Symposium on Mathematical Statistics and Probability; 1956; Berkeley p. 157-163. [doi: 10.1525/9780520313880-015]
31. 2011-2015 American Community Survey: Income in the past 12 months (in 2015 inflation-adjusted dollars) for Ohio. U.S. Census Bureau. URL: https://data.census.gov/cedsci/table?t=Income%20%28Households,%20Families,%20Individuals%29&g=0400000US48&tid=ACSST5Y2015.S1901&hidePreview=false [accessed 2021-03-25]
32. 2011-2015 American Community Survey 5-year estimates: Income in the past 12 months (in 2015 inflation-adjusted dollars) for Texas. U.S. Census Bureau. 2016. URL: https://data.census.gov/cedsci/table?q=TEXAS&t=Income%20%28Households,%20Families,%20Individuals%29&g=0400000US39&tid=ACSST5Y2015.S1901&hidePreview=false [accessed 2021-03-25]
33. 2011-2015 American Community Survey 5-year estimates: Selected housing characteristics for Ohio. U.S. Census Bureau. 2016. URL: https://data.census.gov/cedsci/table?t=Housing&g=0400000US39&tid=ACSDP5Y2015.DP04&hidePreview=false
34. 2011-2015 American Community Survey 5-year estimates: Selected housing characteristics for Texas. U.S. Census Bureau. 2016. URL: https://data.census.gov/cedsci/table?t=Housing&g=0400000US48&tid=ACSDP5Y2015.DP04&hidePreview=false [accessed 2021-03-25]
35. Upton G, Cook I. Understanding Statistics. Oxford: Oxford University Press; 1996.
36. Miles J. Tolerance and variance inflation factor. Wiley StatsRef: Statistics Reference. New York: John Wiley & Sons, Ltd; 2014. URL: https://onlinelibrary.wiley.com/doi/10.1002/9781118445112.stat06593 [accessed 2021-08-16]
37. Shepard D. A two-dimensional interpolation function for irregularly-spaced data. In: Proceedings of the 23rd ACM national conference: Association for Computing Machinery. 1968 Presented at: 23rd ACM national conference: Association for Computing Machinery; August 27 - 29, 1968; New York p. 517-524. [doi: 10.1145/800186.810616]
38. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Sta. Soft 2010;33(1):1-22. [doi: 10.18637/jss.v033.i01]
39. Shawe-Taylor J, Cristianini N. Kernel Methods for Pattern Analysis. Cambridge: Cambridge University Press; 2004:1-478.
40. Quinlan JR. Induction of decision trees. Mach Learn 1986 Mar;1(1):81-106. [doi: 10.1007/bf00116251]
41. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2019. URL: http://www.R-project.org [accessed 2021-03-25]
42. Lawrence S, Giles C, Tsoi AC, Back A. Face recognition: a convolutional neural-network approach. IEEE Trans Neural Netw 1997 Jan;8(1):98-113. [doi: 10.1109/72.554195] [Medline: 18255614]
43. Kipf T, Welling M. Semi-supervised classification with graph convolutional networks. ArXiv. 2017. URL: https://arxiv.org/abs/1609.02907 [accessed 2021-08-16]
44. van Rossum G, Drake Jr F. The Python Language Reference Manual. Devon, UK: Network Theory Ltd; 2011.
45. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: Large-scale machine learning on heterogeneous distributed systems. ArXiv. 2016. URL: https://arxiv.org/abs/1603.04467 [accessed 2021-08-16]
46. Krstajic D, Buturovic LJ, Leahy DE, Thomas S. Cross-validation pitfalls when selecting and assessing regression and classification models. J Cheminform 2014 Mar 29;6(1):10 [FREE Full text] [doi: 10.1186/1758-2946-6-10] [Medline: 24678909]
47. Barnston AG. Correspondence among the correlation, RMSE, and heidke forecast verification measures; refinement of the heidke score. Weather Forecast 1992 Dec;7(4):699-709. [doi: 10.1175/1520-0434(1992)007<0699:catcra>2.0.co;2]
48. Brunsdon C, Fotheringham A, Charlton M. Geographically weighted summary statistics — a framework for localised exploratory data analysis. Comput Environ Urban Syst 2002 Nov;26(6):501-524. [doi: 10.1016/s0198-9715(01)00009-6]
49. Biecek P. DALEX: Explainers for complex predictive models in R. J Mach Learn Res 2018;19(1):3245-3249 [FREE Full text]
50. Liu L. Supplementary materials: interactive maps. GitHub. 2020. URL: https://leiliu-uc.github.io/geospatial/InteractiveMaps.html [accessed 2021-03-25]
51. Beck AF, Sandel MT, Ryan PH, Kahn RS. Mapping neighborhood health geomarkers to clinical care decisions to promote equity in child health. Health Aff (Millwood) 2017 Jun 01;36(6):999-1005 [FREE Full text] [doi: 10.1377/hlthaff.2016.1425] [Medline: 28583957]
52. Moncrief T, Beck AF, Simmons JM, Huang B, Kahn RS. Single parent households and increased child asthma morbidity. J Asthma 2014 Apr 09;51(3):260-266 [FREE Full text] [doi: 10.3109/02770903.2013.873806] [Medline: 24320709]
53. Beck AF, Huang B, Simmons JM, Moncrief T, Sauers HS, Chen C, et al. Role of financial and social hardships in asthma racial disparities. Pediatrics 2014 Mar 02;133(3):431-439 [FREE Full text] [doi: 10.1542/peds.2013-2437] [Medline: 24488745]

54. Beck AF, Riley CL, Taylor SC, Brokamp C, Kahn RS. Pervasive income-based disparities in inpatient bed-day rates across conditions and subspecialties. Health Aff (Millwood) 2018 Apr;37(4):551-559 [FREE Full text] [doi: 10.1377/hlthaff.2017.1280] [Medline: 29608357]
55. Beck AF, Huang B, Auger KA, Ryan PH, Chen C, Kahn RS. Explaining racial disparities in child asthma readmission using a causal inference approach. JAMA Pediatr 2016 Jul 01;170(7):695-703. [doi: 10.1001/jamapediatrics.2016.0269] [Medline: 27182793]
56. Mujahid MS, Roux AV, Cooper RC, Shea S, Williams DR. Neighborhood stressors and race/ethnic differences in hypertension prevalence (the Multi-Ethnic Study of Atherosclerosis). Am J Hypertens 2011 Feb 01;24(2):187-193 [FREE Full text] [doi: 10.1038/ajh.2010.200] [Medline: 20847728]
57. Crowder K, Pais J, South SJ. Neighborhood diversity, metropolitan constraints, and household migration. Am Sociol Rev 2012 Jun 30;77(3):325-353 [FREE Full text] [doi: 10.1177/0003122412441791] [Medline: 22753955]
58. Williams D, Mohammed S, Leavell J, Collins C. Race, socioeconomic status, and health: complexities, ongoing challenges, and research opportunities. Ann N Y Acad Sci 2010 Feb;1186:69-101 [FREE Full text] [doi: 10.1111/j.1749-6632.2009.05339.x] [Medline: 20201869]
59. Smith LA, Bokhour B, Hohman KH, Miroshnik I, Kleinman KP, Cohn E, et al. Modifiable risk factors for suboptimal control and controller medication underuse among children with asthma. Pediatrics 2008 Oct 01;122(4):760-769. [doi: 10.1542/peds.2007-2750] [Medline: 18829799]
60. Boyd RL, Weeks L, McLemore M. On racism: a new standard for publishing on racial health inequities. Health Affairs Blog. 2020. URL: https://www.healthaffairs.org/do/10.1377/hblog20200630.939347/full/ [accessed 2021-03-25]
61. Jones CP, Jones CY, Perry GS, Barclay G, Jones CA. Addressing the social determinants of children's health: a cliff analogy. J Health Care Poor Underserved 2009;20(4 Suppl):1-12. [doi: 10.1353/hpu.0.0228] [Medline: 20168027]
62. Sills MR, Hall M, Colvin JD, Macy ML, Cutler GJ, Bettenhausen JL, et al. Association of social determinants with children's hospitals' preventable readmissions performance. JAMA Pediatr 2016 Apr 01;170(4):350-358. [doi: 10.1001/jamapediatrics.2015.4440] [Medline: 26881387]
63. Beck AF, Simmons JM, Huang B, Kahn RS. Geomedicine: area-based socioeconomic measures for assessing risk of hospital reutilization among children admitted for asthma. Am J Public Health 2012 Dec;102(12):2308-2314. [doi: 10.2105/ajph.2012.300806]

Abbreviations

ACS: American Community Survey
CCHMC: Cincinnati Children's Hospital Medical Center
CNN: convolutional neural network
DALEX: Model Agnostic Language for Exploration and Explanation
DoSC: day-of-surgery cancellation
EHR: electronic health record
GCN: graph convolutional network
GLM: generalized linear regression model
NPO: nil per os
RMSE: root mean square error
SAR: spatial autoregressive
SEM: spatial error model
SVM-P: support vector machine with polynomial kernels
TCH: Texas Children's Hospital
VIF: variance inflation factor

Edited by R Kukafka; submitted 04.12.20; peer-reviewed by L Gottlieb, Z Ren, Y Chu, C Son; comments to author 01.02.21; revised version received 28.03.21; accepted 05.07.21; published 10.09.21.

Please cite as:
Liu L, Ni Y, Beck AF, Brokamp C, Ramphul RC, Highfield LD, Kanjia MK, Pratap J"
Understanding Pediatric Surgery Cancellation: Geospatial Analysis
J Med Internet Res 2021;23(9):e26231
URL: https://www.jmir.org/2021/9/e26231
doi:10.2196/26231
PMID:34505837

Original Paper

Assessing Patient Needs During Natural Disasters: Mixed Methods Analysis of Portal Messages Sent During Hurricane Harvey

Juha Baek[1*], MPSA, PhD; Bridget Simon-Friedt[1*], MS, PhD; Adriana Lopez[1], PhD; Jacob M Kolman[1], MA; Juan Nicolas[1], BS; Stephen L Jones[1,2,3], MD, MSHI; Robert A Phillips[1,4,5], MD; Terri Menser[1,2,3,6], MBA, PhD

[1]Center for Outcomes Research, Houston Methodist, Houston, TX, United States

[2]Department of Surgery, Houston Methodist, Houston, TX, United States

[3]Department of Surgery, Weill Cornell Medical College, New York, NY, United States

[4]Department of Cardiology, Houston Methodist, Houston, TX, United States

[5]Department of Medicine, Weill Cornell Medical College, New York, NY, United States

[6]Department of Population Health Sciences, Weill Cornell Medical College, New York, NY, United States

[*]these authors contributed equally

Corresponding Author:
Terri Menser, MBA, PhD
Center for Outcomes Research
Houston Methodist
6550 Fannin St
Houston, TX
United States
Phone: 1 2706252749
Email: tmenser@houstonmethodist.org

Abstract

Background: Patient portals play an important role in connecting patients with their medical care team, which improves patient engagement in treatment plans, decreases unnecessary visits, and reduces costs. During natural disasters, patients' needs increase, whereas available resources, specifically access to care, become limited.

Objective: This study aims to examine patients' health needs during a natural crisis by analyzing the electronic messages sent during Hurricane Harvey to guide future disaster planning efforts.

Methods: We explored patient portal use data from a large Greater Houston area health care system focusing on the initial week of the Hurricane Harvey disaster, beginning with the date of landfall, August 25, 2017, to August 31, 2017. A mixed methods approach was used to assess patients' immediate health needs and concerns during the disruption of access to routine and emergent medical care. Quantitative analysis used logistic regression models to assess the predictive characteristics of patients using the portal during Hurricane Harvey. This study also included encounters by type (emergency, inpatient, observation, outpatient, and outpatient surgery) and time (before, during, and after Hurricane Harvey). For qualitative analysis, the content of these messages was examined using the constant comparative method to identify emerging themes found within the message texts.

Results: Out of a total of 557,024 patients, 4079 (0.73%) sent a message during Hurricane Harvey, whereas 31,737 (5.69%) used the portal. Age, sex, race, and ethnicity were predictive factors for using the portal and sending a message during the natural disaster. We found that prior use of the patient portal increased the likelihood of portal use during Hurricane Harvey (odds ratio 13.688, 95% CI 12.929-14.491) and of sending a portal message during the disaster (odds ratio 14.172, 95% CI 11.879-16.907). Having an encounter 4 weeks before or after Hurricane Harvey was positively associated with increased use of the portal and sending a portal message. Patients with encounters during the main Hurricane Harvey week had a higher increased likelihood of portal use across all five encounter types. Qualitative themes included: access, prescription requests, medical advice (chronic conditions, acute care, urgent needs, and Hurricane Harvey–related injuries), mental health, technical difficulties, and provider constraints.

Conclusions: The patient portal can be a useful tool for communication between patients and providers to address the urgent needs and concerns of patients as a natural disaster unfolds. This was the first known study to include encounter data to understand portal use compared with care provisioning. Prior use was predictive of both portal use and message sending during Hurricane

Harvey. These findings could inform the types of demands that may arise in future disaster situations and can serve as the first step in intentionally optimizing patient portal usability for emergency health care management during natural disasters.

(J Med Internet Res 2021;23(9):e31264) doi:10.2196/31264

KEYWORDS

Hurricane Harvey; natural disasters; patient portals; electronic messages; emergency preparedness; disaster medicine

Introduction

The prevalence of natural disasters such as storms and severe flooding has increased over the last century, both nationally and globally, leading to substantial human and financial loss [1-3]. In the United States, the average number of disasters per year has significantly risen from 1.63 events per year (from 1917 to 1966) to 14 events per year (from 1967 to 2016) [3]. Since the early 1990s, the average annual number of disasters has dramatically escalated to 19.7 events [3]. Of the 10 worst natural catastrophes in US history, 9 occurred after 2000.

Hurricane Harvey was a Category 4 hurricane that made landfall along the Texas coast on August 25, 2017. This storm deposited more than 60 inches of rain onto the city of Houston and surrounding areas and caused more than 100 deaths because of direct or indirect effects [4]. It also adversely affected approximately 200,000 homes, causing an estimated US $125 billion in damage, making it the second most financially devastating storm after Hurricane Katrina in 2005 [5]. The resultant flooding and damage caused by Harvey brought health care access challenges for area residents across the Greater Houston community. Other recent disasters include the COVID-19 pandemic and winter storm Uri (February 2021) that shut down the Houston area for 5 days and caused loss of life because of infrastructure inadequacies. Both disasters affected access to traditional health care settings. The operational lessons learned during Hurricane Harvey have been previously described and summarize the organizational response during natural disasters [6]. During such disasters, the medical needs of patients increase, whereas available resources become limited. As a result of such emergency situations, health care infrastructures are often unable to operate normally, thus restricting patients' access to timely care and communication with their providers [7].

Digital patient portals were first introduced in the late 1990s and are defined as secure web-based tools that allow patients to access their health information and communicate with their health care team through electronic messages [8-11]. Health care organizations received government incentives under the Meaningful Use Incentive Program of the Health Information Technology for Economic and Clinical Health Act of 2009 which has increased the adoption and use of patient portals in health care [12-14]. Current statistics suggest that approximately 90% of the health care organizations provide some type of portals to their patients, and a 2016 American Hospital Association report also showed that approximately 92% of the patients are able to access their health records [15,16]. Patient portals contribute to improving patient engagement in health care, facilitating patient-centered care, enhancing satisfaction with care, reducing unnecessary visits and phone calls, and decreasing health care service use and costs [8,17-21].

Patient portals have a wide array of functionalities of potential relevance in a disaster scenario, including secure messaging between patients and health care providers; access to medical records, test results, and education resources; and medication refills [22-24]. In addition, portal access is often a prerequisite to being able to schedule and attend telehealth visits, which have increased significantly since the COVID-19 pandemic struck [25-27]. Secure messaging has been reported to be among the most commonly used features of patient portals [23,24], and qualitative assessment of patient portal message content has been conducted across various patient populations, including patients with diabetes, surgical patients, and patients in the primary care setting [10,28,29]. Prior studies have shown that most messages were related to patients' medical needs such as appointments, medical problems, and test-related content [28,29].

Although previous studies have examined portal messages sent between patients and providers across various patient populations and health care settings, these evaluations were conducted during periods of standard care. Given that patients' needs increase during emergency situations, communication between patients and providers through patient portals could be an effective tool to help patients manage their health status and improve health outcomes in a period of heightened medical need. However, to our knowledge, no study has evaluated patients' electronic messages or patient portal use during a disaster. Therefore, this study aims to examine patients' health needs during a natural crisis, which can help guide future intentional efforts with regard to using patient portals during disaster periods.

Methods

Design and Setting

This study was conducted at Houston Methodist, which is a large health system in the Greater Houston area comprising 8 hospitals, including an academic hospital. All patients can request access to the web-based patient portal, which allows them to communicate with health care providers, check laboratory test results, update medications, request appointments, and manage their care on the web. This study used a mixed methods approach to assess patients' immediate health needs during an emergency situation resulting from Hurricane Harvey, which was defined in this study as the day of hurricane landfall and the following 6 days (from August 25, 2017, to August 31, 2017).

The study population was limited to patients aged 18 to 105 years, and portal access and portal nonaccess were compared

by demographic characteristics. We included the entire electronic health record (EHR) population for 2017 and created categories for missing demographic information instead of omitting these individuals. Omitting patients for selecting the *prefer not to answer* response for race and ethnicity, specifically, could have unnecessarily decreased the total population and unintentionally reduced its diversity. In addition, all patients in the quantitative model have data for portal activity during the week of Hurricane Harvey, which is the main study objective. This study was reviewed and approved by the institutional review board of the Houston Methodist Research Institute.

Measures

Outcomes

Portal users were distinguished from the total patient population in 2017 using a binary indicator for those who had an active portal account, regardless of use. Portal use was measured using patient-level audit trail data obtained from the EHR and was coded binarily; Harvey portal use was defined as at least one log-in during the designated disaster period week from August 25, 2017, to August 31, 2017. Harvey message sending was also coded in a binary manner, with *1* assigned to all users who sent at least one message during the study period.

Covariates

The main independent variable was the binary use of the patient portal before Harvey in 2017, defined as a single log-in between January 1, 2017, and August 24, 2017. The demographic factors included age (18-44 years, 45-64 years, ≥65 years, or unknown), sex (female or male), and race (White, Black, Asian, Other, or Unknown). The Other category collapsed races with small sample sizes and included Native Hawaiian or Other Pacific Islander and American Indian or Alaska Native. The Unknown category included patients who declined to answer demographic questions or EHRs with no value for the respective category. Binary measures of a defined set of comorbidities using the Charlson Comorbidity Index [30] were created using patient-level *International Classification of Diseases, Tenth Revision*, codes based on discharge dispositions that accompany all encounter types.

In all, five types of encounter variables were constructed based on the date of appointment and included emergency department (ED) visits, inpatient stays, observations, outpatient visits, and surgeries. These encounter types were selected because they were five of the most common encounter types of a total of 23 encounter types that year, representing 87% of the total encounters. The categories for each of these encounter variables were developed by considering the patients' likelihood of or need for interaction with the health care system, which we assumed would be relative to their receipt of care related to the disaster timeline (eg, patients' portal use would likely increase just before or directly after an appointment, procedure, or other encounter). The encounter categories were defined as (1) no encounter in 2017; (2) an encounter in 2017, excluding the 4 weeks before, the week of, and the 4 weeks after Hurricane Harvey (January 1-July 27, 2017, or September 29-December 31, 2017); (3) the 4-week period both immediately before (July 28-August 24, 2017) and immediately after the week of Harvey (September 1-September 28, 2017); and (4) an encounter during the week of Harvey (August 25-August 31, 2017).

Analytic Methods

The quantitative evaluation included two main binary indicators of portal use: portal use of any kind and message sending during the defined study period. The qualitative evaluation component identified messages by date and examined all message content sent through the patient portal during the week of Hurricane Harvey. Further quantitative analyses focused on portal use were limited to the subset of patients with a portal account because access to the portal was a prerequisite.

Descriptive statistics were calculated for the total patient population in 2017 and patients with an activated portal account and compared with those of portal nonusers, in addition to the subpopulations of interest in this study (ie, portal use and message sending during Harvey), to readily allow for comparison of the patients' characteristics. The mean and SD were calculated for continuous variables, and frequencies and percentages were used for categorical variables. We used two-tailed *t* tests or Pearson chi-square tests to initially examine differences between portal users and portal nonusers but presented standardized mean differences (SMDs) given the large sample size of the study population; an absolute value greater than or equal to 0.10 of the SMD was considered statistically significant [31]. We used logistic regression models to assess predictors of portal use (model 1) and portal message use (model 2) during Hurricane Harvey. All statistical analyses were conducted using Stata version 16 (StataCorp LLC). Statistical significance was set at $P<.05$.

Qualitative Methods

All electronic messages sent by patients through the web-based patient portal during the study period were extracted from the EHR, and message IDs were used to pair individual messages with the sender and receiver. Structured query language was used to format the message threads and export them to a spreadsheet. All automated or systemwide messages were excluded, such as bulk messages regarding office closures and other disaster-related communications sent from the health system. Initially, we attempted to segregate Harvey-related messages from standard care messages, but given the subjective nature of defining a Harvey-related message, all messages from the study period were reviewed in the qualitative thematic analysis [32,33]. The first message file (4354/9316, 46.74% of messages) was coded without a priori defined themes by JB, BSF, and TM using the constant comparative method, which is an inductive data coding process [34]. After reviewing the common themes across these initial messages, the following targeted research questions were developed:

1. *What problems did patients face during Harvey that they communicated through the patient portal?*

2. *How were providers able to help patients through the patient portal?*

3. *What needs were providers unable to address through the patient portal?*

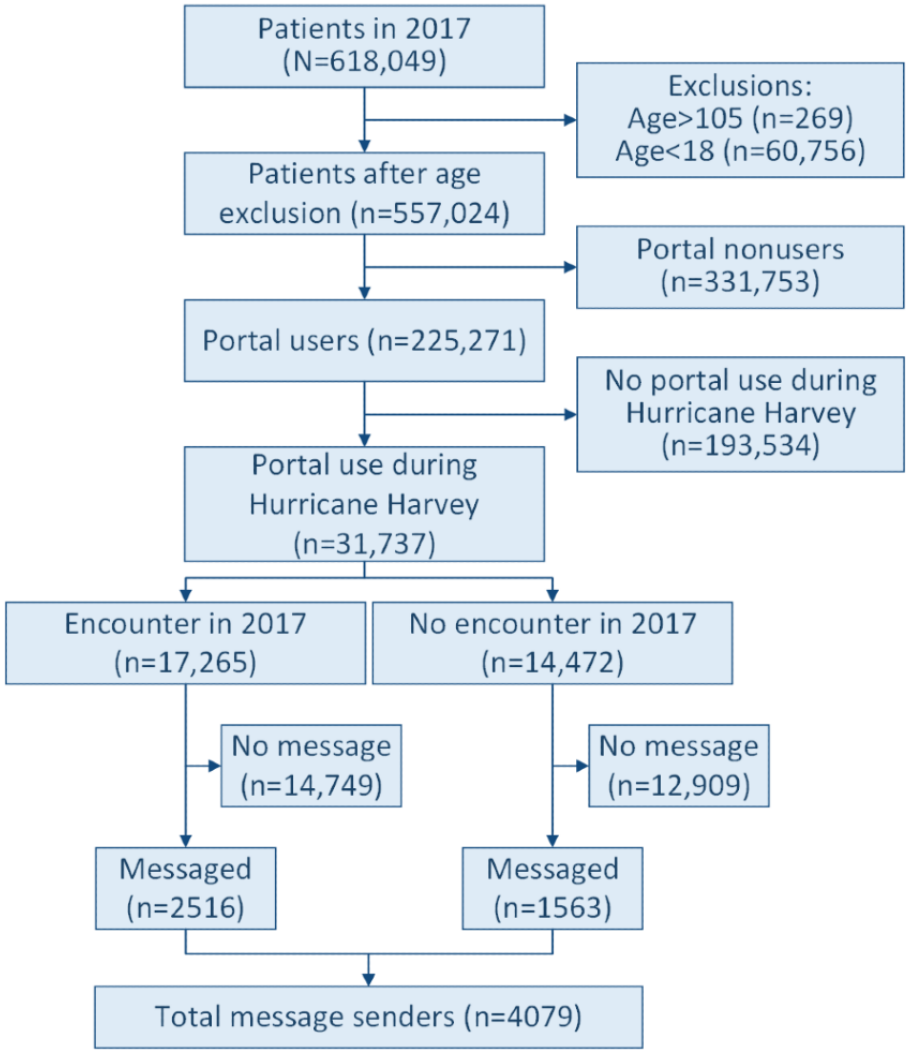
The coding of the remaining messages (4962/9316, 53.26%) was guided by these questions and performed using Quirkos software version 2.4.1 (Quirkos) [35]. Both initial coding rounds were combined into a Microsoft Excel (Microsoft Corporation) file, where the second-level theme assignment was carried out individually by JB and BSF. Weekly group meetings, which included TM, JB, BSF, and JMK, with other authors attending as needed, were held to discuss and condense second-level themes. The codes created by the 2 primary coders (JB and BSF) were compared during these meetings, and inconsistencies between them were resolved through discussion to reach a consensus.

Results

Overview

The study sample included 557,024 patients aged between 18 and 105 years. We categorized this population into portal users, portal nonusers, portal users during Harvey, and message senders during Harvey, in addition to categorizing portal users with and without encounters in 2017 (Figure 1). The percentage of patients with an activated portal account (portal users) and those with no account (portal nonusers) was 45.82% (255,271/557,024) and 59.55% (331,753/557,024), respectively. Among portal users in 2017, 12.43% (31,737/255,271) used the portal during Harvey, which is 5.69% (31,737/557,024) of the total Houston Methodist patient population. It is interesting to note that of these patients, 54.4% (17,265/31,737; encounter) had at least one appointment in 2017, and 24.63% (7816/31,737) of these patients had an appointment surrounding (4 weeks before or 4 weeks after) or during the week of Harvey. Of the 4079 patients who sent a portal message during Harvey, 2516 (61.68%) had an appointment in 2017 and 1317 (32.28%) had an appointment either 4 weeks before, the week of, or 4 weeks after Harvey.

Figure 1. Flow diagram of selection process for study population and total number of patients who messaged during Hurricane Harvey.



A total of 11.96% unique patients (66,647/557,024 of the total patients) sent a portal message at least once in 2017, totaling 364,617 messages, with an average of 1.62 messages per patient. Among these messages, approximately 2.55% (9316/364,617) were sent during Harvey, with an average of 2.3 messages per patient. Analysis of the number of messages sent per day showed that the most number of messages were sent on Thursday (the last day) during the week of Harvey (2981/9316, 31.99%), followed by Wednesday (the sixth day: 1981/9316, 21.26%) and Friday (the first day: 1726/9316, 18.53%). Overall, we found that although patient portal use declined during Harvey, portal messaging maintained near average patient-level use. In

2017, 13.8% of unique patients of the total activated users used the messaging function of the portal [32] compared with 12.85% (4079/31,737) of the patients who logged into the portal during Harvey and who also sent a message.

Quantitative Results

Portal Access and Portal Use During Harvey

Multimedia Appendix 1 shows the descriptive statistics of the total number of patients in 2017 (n=557,024) and of the subpopulations of interest: patients who used the portal during Harvey (n=31,737) and patients who messaged during Harvey (n=4079). Most of the patients who messaged and used the portal were aged 45 years or older, female, White, and non-Hispanic/Latinx. Of the total number of portal users during Harvey, 43.51% (13,808/31,737) had an outpatient encounter, as did more than half (2065/4079, 50.62%) of the patients who messaged during Harvey, which is consistent with the number of overall portal users in 2017 because patients with outpatient encounters composed the largest percentage of portal users in general. When comparing portal users and portal nonusers in 2017, we found that portal users were more likely to be aged below 65 years, female, White, and non-Hispanic/Latinx compared with portal nonusers. Age, race, and ethnicity were statistically significant different between the portal user groups (SMD >0.10), although sex was not. The patients who used the portal in 2017 had lower percentages of most of the known comorbidities except cancer, connective tissue disease, and mild liver disease compared with their counterparts; however, SMDs for all comorbidities were less than 0.10.

Portal Use During Hurricane Harvey

To examine the predictors of patient portal use during Hurricane Harvey, we used multivariate logistic regression (model 1; Table 1). We found that the patients who had used the portal before Harvey in 2017 were approximately 13.7 times more likely to use the portal during Harvey than those who had not used the portal (odds ratio [OR] 13.688, 95% CI 12.929-14.491; *P*<.001). Compared with the patients aged 18-44 years, those aged 65 years or older were 1.13 times as likely to use the portal during Harvey (OR 1.128, 95% CI 1.091-1.166; *P*<.001), but those aged 45-64 years were less likely to use the portal during Harvey (OR 0.949, 95% CI 0.921-0.978; *P*=.001). Male and Hispanic/Latinx patients were less likely to use the portal during Harvey than female and non-Hispanic/Latinx patients (OR 0.963, 95% CI 0.938-0.989, *P*=.005 and OR 0.860, 95% CI 0.823-0.898, *P*<.001, respectively). The results also showed that patients who had an appointment within 4 weeks before or after Harvey were more likely to use the portal during Harvey than those who did not have any appointment in all types of encounters (OR 2.052, 95% CI 1.973-2.134 for outpatient; OR 1.810, 95% CI 1.681-1.949 for surgery; OR 1.608, 95% CI 1.444-1.790 for observations; and OR 1.673, 95% CI 1.551-1.804 for inpatient, *P*<.001 and OR 1.085, 95% CI 1.001-1.175 for ED, *P*=.046). In addition, the positive associations between having an appointment during Harvey and using the portal during Harvey were found to be stronger for each type of encounter than with having an appointment within 4 weeks before or after Harvey (OR 5.897, 95% CI 5.094-6.826 for outpatient; OR 4.337, 95% CI 3.106-6.057 for surgery; OR 4.697, 95% CI 3.366-6.554 for observations; OR 2.917, 95% CI 2.340-3.635 for ED; and OR 2.743, 95% CI 2.307-3.261 for inpatient; *P*<.001).

Table 1. Results of logistic regression models fitted to the outcomes: used the patient portal during Harvey (Model 1) and messaged during Harvey (Model 2).

Independent variables	Model 1: used patient portal during Harvey		Model 2: messaged during Harvey	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value
Used portal before Harvey in 2017 (reference: no portal use in before Harvey in 2017)				
Yes	13.688 (12.929-14.491)	<.001	14.172 (11.879-16.907)	<.001
Age (years; reference: 18-44 years)				
45-64	0.949 (0.921-0.978)	.001	1.168 (1.081-1.262)	<.001
≥65	1.128 (1.091-1.166)	<.001	1.298 (1.192-1.413)	<.001
Sex (reference: female)				
Male	0.963 (0.938-0.989)	.005	0.899 (0.841-0.962)	.002
Race (reference: White)				
Black	0.977 (0.940-1.017)	.26	0.859 (0.775-0.952)	.004
Asian	0.987 (0.939-1.038)	.61	0.793 (0.690-0.912)	.001
Other	1.097 (1.031-1.168)	.004	1.081 (0.919-1.272)	.34
Unknown	0.990 (0.929-1.056)	.76	1.015 (0.859-1.198)	.86
Ethnicity (reference: non-Hispanic/Latinx)				
Hispanic/Latinx	0.860 (0.823-0.898)	<.001	0.854 (0.763-0.955)	.006
Unknown	0.978 (0.900-1.061)	.59	1.008 (0.814-1.248)	.94
Outpatient (reference: no encounter in 2017)				
Encounter in 2017 outside Harvey period[b]	0.898 (0.872-0.925)	<.001	1.037 (0.961-1.120)	.35
Encounter peri-Harvey[b]	2.052 (1.973-2.134)	<.001	2.465 (2.260-2.689)	<.001
Encounter during Harvey[b]	5.897 (5.094-6.826)	<.001	5.009 (3.944-6.362)	<.001
Surgery (reference: no encounter in 2017)				
Encounter in 2017 outside Harvey period[b]	0.953 (0.912-0.996)	.03	0.989 (0.890-1.100)	.85
Encounter peri-Harvey[b]	1.810 (1.681-1.949)	<.001	1.762 (1.514-2.050)	<.001
Encounter during Harvey[b]	4.337 (3.106-6.057)	<.001	0.583 (0.184-1.843)	.36
Observations (reference: no encounter in 2017)				
Encounter in 2017 outside Harvey period[b]	0.861 (0.810-0.915)	<.001	0.931 (0.807-1.075)	.33
Encounter peri-Harvey[b]	1.608 (1.444-1.790)	<.001	1.444 (1.156-1.802)	<.001
Encounter during Harvey[b]	4.697 (3.366-6.554)	<.001	1.201 (0.549-2.628)	.65
Emergency department (reference: no encounter in 2017)				
Encounter in 2017 outside Harvey period[b]	0.722 (0.690-0.755)	<.001	0.900 (0.809-1.002)	.06
Encounter peri-Harvey[b]	1.085 (1.001-1.175)	.046	1.407 (1.188-1.667)	<.001
Encounter during Harvey[b]	2.917 (2.340-3.635)	<.001	2.317 (1.505-3.565)	<.001
Inpatient (reference: no encounter in 2017)				
Encounter in 2017 outside Harvey period[b]	0.903 (0.861-0.947)	<.001	1.039 (0.930-1.161)	.50
Encounter peri-Harvey[b]	1.673 (1.551-1.804)	<.001	1.882 (1.619-2.186)	<.001
Encounter during Harvey[b]	2.743 (2.307-3.261)	<.001	2.088 (1.498-2.909)	<.001

[a]OR: odds ratio.

[b]Outside Harvey period defined as the periods January 1-July 27, 2017, and September 29-December 31, 2017; peri-Harvey period defined as the 4-week periods both immediately before Harvey (July 28-August 24, 2017) and immediately after Harvey (September 1-September 28, 2017); and during

Harvey period defined as August 25-August 31, 2017.

Message Sending During Hurricane Harvey

To explore the factors associated with patients sending a message during Harvey, we also used multivariate logistic regression analysis (model 2; Table 1). The results revealed that prior portal use in 2017 was positively associated with sending a message during Harvey (OR 14.172, 95% CI 11.879-16.907; *P*<.001). Patients aged 45-64 years or those aged 65 years or older were more likely to send a message through the portal during Harvey than those aged 18-44 years (OR 1.168, 95% CI 1.081-1.262 for those aged 45-64 years and OR 1.298, 95% CI 1.192-1.413 for those aged 65 years or older; *P*<.001). Male, Black, Asian, and Hispanic/Latinx patients were less likely to send a portal message during Harvey than their counterparts (OR 0.899, 95% CI 0.841-0.962, *P*=.002 for sex; OR 0.859, 95% CI 0.775-0.952, *P*=.004 for Black patients; OR 0.793, 95% CI 0.690-0.912, *P*=.001 for Asian patients; OR 0.854, 95% CI 0.763-0.955, *P*=.006 for Hispanic/Latinx patients). We found that patients with an appointment within 4 weeks before or 4 weeks after Harvey were more likely to send a portal message during Harvey than those without an encounter for each type of encounters (OR 2.465, 95% CI 2.260-2.689 for outpatient; OR 1.762, 95% CI 1.514-2.050 for surgery; OR 1.444, 95% CI 1.156-1.802 for observations; OR 1.407, 95% CI 1.188-1.667 for ED; and OR 1.882, 95% CI 1.619-2.186 for inpatient; *P*<.001). Furthermore, patients who had outpatient, ED, or inpatient encounters during Harvey showed higher ORs for sending a message during Harvey than those who had no encounter (OR 5.009, 95% CI 3.944-6.362 for outpatient; OR 2.317, 95% CI 1.505-3.565 for ED; and OR 2.088, 95% CI 1.498-2.909 for inpatient; *P*<.001).

Qualitative Results

Overview

A total of 9316 electronic messages (inclusive of sent and received messages) were paired with 4079 unique patients during the study period and evaluated to identify emergent themes. Despite electricity outages during Hurricane Harvey, portal users were still able to communicate with providers during this disruption of care, although portal use decreased overall during this period compared with the rest of the year. We categorized message content into the following themes that summarize patients' needs and providers' responses during Hurricane Harvey: access, prescription requests, medical advice, mental health, technical difficulties, and provider constraints.

Access

Messages related to access included no *physical access to health care* (eg, physician office closures requiring rescheduling or rerouting), no *physical access to pharmacy services* (eg, pharmacy closures due to flooding and power outages), *structural access issues* (eg, roadway conditions or flooding blocking transportation to health services), and patients' *inability to access established health services* (eg, evacuation to a nonflooded area or being stranded away from home). Sometimes, these varied types of access would overlap; for example, a displaced patient unable to access their established surgical team described a structural access issue despite having physical access that required rescheduling:

> *Due to the weather they called me yesterday to cancel my preop appointment, but they kept my surgery for tomorrow. Unfortunately, there is NO WAY I can get out of my house tomorrow, the water on the street is too high to allow us to drive to the hospital...I need to reschedule all my appointments.* [Portal user]

Several patients requested a web-based referral because the disaster required them to temporarily relocate to another city and cancel existing appointments; however, the serious nature of the condition required ongoing specialized care.

> *My wife and I fortunately relocated to [city name] just before Harvey as it turns out and now are only dealing with issues regarding our house in Houston which was flooded. Since I have a need for a cardiologist in [city], could you kindly refer me to one?* [Portal user]

In addition, patients sought direction from providers, exchanging messages on when and how to safely reach health facilities. Patients sometimes asked for suggestions on where to find accessible pharmacy locations, and providers routinely resent prescriptions and laboratory test orders to different locations to enable patient access:

> *Script is on the way! Hope the pharmacy is open! Let me know either way. We are trying to assess the area. All the roads out of [location] are blocked by water.* [Provider]

Patients and providers used the portal to discuss and confirm appointments while requesting or providing advice on navigating the roadways:

> *Do you know about [location] area?...I'll just try to keep my appointment for tomorrow...* [Portal user]
>
> *249 and 1960 are open and I'm here today if you want to come in this morning. Let me know.* [Provider]

Prescription Requests

Patients frequently messaged providers regarding prescription needs due to *their inability to acquire their medications*, fear of *running out during the disaster period* (eg, prescription refills), and *emerging medical needs* that resulted from the storm (eg, new prescription requests or replacement medications). Most patients asked to refill prescriptions because they were running out of medication. Some patients described situations in which their medications were destroyed or unreachable because of flooding and requested a prescription refill:

> *My prescriptions along with everything in my house have been destroyed. My whole house is under water and I haven't gotten to take any of my medicine...please put in another prescription to me for [pharmacy name]? I haven't even taken one single pill out of the medicine you gave me.* [Portal user]

Several messages came from patients who could not return because of Harvey but needed to acquire the medication out of state:

> *I am currently stranded outside of Boston, MA, and we're not sure when we'll make it back. I have two more days of medication. Is there any way you can send in an emergency prescription to [pharmacy name] at [address in MA]?* [Portal user]

Other patients messaged that they were struggling to get the medication because of weather-related pharmacy closures or lack of medication available at their local pharmacy. They asked providers to resend a prescription to an alternative available pharmacy, saying as follows:

> *The pharmacist assistant at [pharmacy name] on [pharmacy location road] informed me that the prescription that Dr.[X] called in for me for [beta blockers] is difficult to get...Can Dr. [X] prescribe something else for me?* [Portal user]

Some messages included context related to delay or disruption of care due to difficulties in acquiring medication from a pharmacy or a delayed prescription approval from a provider during the hurricane situation:

> *He took the last of his [immunomodulatory drug] on Aug. 22...However, the pharmacy notified us that because of the storm the medication could not be delivered...As of today, we have not received his prescription.* [Portal user]

Medical Advice

A varied heterogeneous family of portal message topics reflected patients' need for medical advice during the disaster. We found that the types of medical advice that patients were seeking were related to *chronic conditions*, *acute care*, *urgent needs*, and *Harvey-related injuries*.

Advice was sought during the disaster for existing *chronic health conditions*, including diabetes, hypertension, seizures, and migraines. The questions included medication adjustments, symptom assessment, and management until regular health care operations could resume. A patient messaged during the night with concern regarding elevated hypertension and associated physical symptoms. The patient described severe headaches, facial numbness, and dizziness and used the portal to seek medical advice after avoiding a potentially crowded medical facility:

> *I need your advise how [sic] to manage my [hypertension] at night...I was told Methodist WB is so packed and people were waiting to be seen I elected not to go, I was able to sleep for 3 hrs and this morning my bp was 146/90, Can I take [antihypertension drug] 50 mg in am and [Calcium channel blocker] 10 mg at night...* [Portal user]

Patients used the portal to seek medical advice for *acute illnesses* that occurred during the disaster period. These messages detailed symptoms of common ailments, including allergies, urinary tract infections, and upper respiratory infections. Providers were able to readily address many of these acute care needs through the portal:

> *I came home with a wicked sore throat. It hurts when I talk and swallow. I get strep all the time and I'm worried about that. I checked with a flashlight and it looks like there is white in the back of my throat. I am not 100% sure. I'm worried about driving in, and worried you won't have an appointment. If I can't get into see you what should I do?* [Portal user]

> *Hi [name], I am sending amoxicillin to your [pharmacy name] now.* [Provider]

There were instances in which patients reached out to their providers for *urgent guidance*, including patients who had recently undergone surgery in the weeks before the hurricane who messaged for postsurgery guidance. In addition, pregnant patients nearing anticipated delivery dates used the portal to reach their providers to describe new symptoms that could affect their pregnancy and to make alternative birthing plans if delivery were to occur during the disaster or if they were displaced because of evacuation. An urgent message request was sent by a pregnant patient to obtain guidance and to transfer her patient information:

> *I'm leaking some sort of fluid can a nurse call me [phone number]...Can you send over my records to the hospital I'm in now...* [Portal user]

Other patients sought advice regarding Hurricane *Harvey–related injuries*. The portal reflected specific questions about injuries resulting from disaster preparations to protect property (eg, strained muscles from moving furniture) as well as postdisaster recovery efforts, including exposure to unclean flood water. These types of messages included concerns related to infection and tetanus through minor injuries:

> *...are you opened [sic] today? I got a cut on my leg while helping rescuing on Monday. I spoke with Dr [name] and he said I need to be seen and need a tetanus sho[t]...* [Portal user]

Other patients sought medical advice after a substantial injury and the inability to receive routine medical care during the disaster:

> *Dr [name], during the flood I slipped in my house and believe I have broken at least one toe...My foot continues to swell and 3 toes are bruised...Would you be able to just send an order for an xray to San Jacinto outpatient bldg for xray? The...[m]edical clinic close to my house has not been open all wee[k]...* [Portal user]

Mental Health

Several patients expressed feelings of experiencing tragedy, anxiety, and depression in relation to the storm, flooding, and loss of property. Patients used words such as "devastated" and "surreal" to convey their feelings to health care providers. These messages were often sent as prescription requests for short-term medication options to aid in coping during the disaster. Portal users described physical symptoms associated with the mental

health aspects of storm-related anxiety, especially the inability to sleep:

> *We had flooding in our house, and it's been stressful. I'm wondering if you could prescribe me [controlled substance – sedative] or similar to help with anxiety/sleeplessness.* [Portal user]

Some patients noted how stress brought on by the circumstances of the hurricane resulted in physiological responses:

> *The stress of this situation has induced a few panic attacks. I realize I am due for an office visit but it is not possible during this flooding situation and caring for my elderly mother. I would appreciate a one month renewal on these medications until we get this situation squared away.* [Portal user]

Others conveyed descriptive accounts of their own circumstances during Hurricane Harvey that led to feelings of reduced mental health and well-being when making prescription requests:

> *...Unfortunately, [we] (our 6 year-old son)...were not so lucky. We took on serious amounts of water/sewage throughout our entire home and ended up having to make a treacherous emergency evacuation walking in water above our chest (with my son on my head) for 90 minutes through rushing currents to reach safety...The stress of worrying about my son and [name] (who slipped under the water and was sucked down [from] the undercurrent) is leaving me completely anxious and unable to calm down...* [Portal user]

Some users asked to be connected directly with a mental health professional through the portal during the disaster. This situation was addressed by providers sending referrals through the portal:

> *...this catastrophe...likely caused blackout[s] for people who suffered in past. Do you want to see therapist or psychiatrist, [so] I can place correct order[?]* [Provider]

Technical Difficulties

Patients' messages described communication difficulties that were often related to office closures, phone malfunctions, and other crisis-related technical difficulties. Some patients said that they wanted to call an office to confirm their appointment or check if the office was open, but the office phone was not working:

> *We have tried calling the [location] office, but the recording says that the line is not in service. We've also tried calling the [alternative location] office number but could not get through. Please advice [sic].* [Portal user]

Several messages reflected that patients' prescriptions were not delivered to the pharmacy because of a problem with the pharmacy's electronic system. One provider messaged to inform a patient of the circumstance:

> *We have noticed in our office that since the storm some electronic prescriptions are not going through, I am guessing due to damaged fax lines and EMR malfunctions. Please let us know if your prescription did not transmit and we can call it in.* [Provider]

Some patients mentioned that communication was delayed because internet access was limited on account of Hurricane Harvey. Other patients messaged to report difficulties in rescheduling an appointment or surgery because of an issue in the web-based scheduling system:

> *I tried to re-schedule my appt from...yesterday for Fri 09/01, at 2:00. Online scheduling showed that date/time available but now it doesn't show as scheduled...I have no idea whether I should come in Friday at 2:00 or schedule another time or?* [Portal user]

Provider Constraints

Occasionally, in their responses to patients, providers noted their own difficulties in fully addressing patient needs through the portal during the disaster. The providers mentioned technical complications, including loss of power and internet access, which delayed message responses; flooding that inhibited their ability to travel to medical facilities; technical issues related to accessing the EHR; office closure due to flooding; and inadequate ability to assess some medical conditions through a digital portal. Many complications in addressing patient messages were related to providers' home or office damage (eg, flooding and loss of power) caused by the storm. Further obstacles were noted due to inadequate access to patient records through the EHR or limited to no access to the internet and patient portal to readily respond to patients' requests:

> *Sorry I did not answer you sooner - we did not have any electricity or internet so I could not get on [the patient portal]. Please call the oncall number if you have any problems - they will contact me or another post coordinator to assist...* [Provider]

We found that while providers were navigating storm-related issues, other members of the health care team were able to alert patients regarding delayed provider responses:

> *Good Morning, I apologize for the inconvenience upon you at this time. Dr. [X] is not available like others he is unable to return to work at this time, but he is checking his messages throughout the day. When he responds I will definitely respond to this message.* [Provider]

Additional complications prevented providers from adequately addressing patient concerns through the portal. Certain conditions must be evaluated in office; thus, some providers noted that the digital portal was not sufficient and referred patients to urgent care centers with guidance as appropriate:

> *Dr. [X] is not able to access [electronic health record software] at this time, and the office is closed Monday and Tuesday. If your symptoms [s]ignificantly worsen or anything changes, go to urgent care...* [Provider]

For additional representative quotations in each of the themes discussed above, see Textbox 1.

Textbox 1. Representative quotes of qualitative portal message themes. All quotes are from patient portal users, unless otherwise noted.

Access

- "I am sorry but I will have to postpone my today's appointment due to Hurricane Harvey and associated rain and traffic mess on Houston roads. I will reschedule after next week once everything calms down."
- "Due to the storm, my appointment with the Orthopedic surgeon got cancelled...I don't know for how long. Can you refer me to another Doctor as I don't want to keep waiting...in case my foot is not healing properly?"

Prescription requests

- "I have been flooded out of my house and don't have any of my meds. Can you call in to [my pharmacy location] with a few days of each of my meds?"
- "Sunday I was rescued by boat from my home and a holed up now at [location]. – also lost both vehicles in the flood. Really stressed out. Can you prescribe something to calm me through the next few weeks?"

Medical advice

- "...based on the condition with the weather would it be possible for me to go to urgent care to get my stitches remove, it is becoming increasingly uncomfortable."
- "All week my eyes were fine while I was taking the steroids. Since yesterday they are starting to itch and the lids are red and swollen again. What would you suggest I do?..."

Mental health

- "Due to recent events, (Tropical Storm Harvey,...multiple family tragedies, moving,...etc), I am not feeling the effects from taking this medication. I was wondering if it was possible to at least temporarily increase the dosage or increase to twice daily?"
- "...my home was completely flooded I have lost all of my meds and breathing machine and inhalers I am having a hard time breathing and my anxiety is out the roof being in a shelter my stress is extremely high...please help me figure this out..."

Technical difficulties

- "I was away from my phone and did not answer in time. When I tried to call back, I get an error message...I am just wondering if my appointment is still schedule[d]..."
- "My [pharmacy] lost electricity and phones. They reopened today but still do not have phones. Please call in a script to [pharmacy] at [alternative location]."

Provider constraints

- "I did forward your message to Dr. [X], he is not able to get out his neighborhood. I hope to hear from him today on what labs he wants to order for you. I will let you know as soon as I put in the orders." [provider]
- "Please call our office so one of our medical assistants can send your order. At this time I'm not able to fax anything, we are just working from a laptop and I'm currently in Dallas due to the storm." [provider]

Discussion

Principal Findings

We found that prior use of the patient portal was highly predictive of both using the patient portal and sending a message during Hurricane Harvey. This study uniquely included encounter data as a proxy measure, which we expected to be predictive of portal use generally but also specifically during a disaster situation. More than half of the total number of portal users during Hurricane Harvey had an outpatient encounter within 4 weeks before or after the week of the hurricane. Patients with recent and upcoming interactions with the health system were also more likely to have used the portal and messaged providers during the main week of Hurricane Harvey, indicating the importance of including health care needs in portal use analyses.

Previous evidence during normal operations suggests that active portal users had higher comorbidity scores [36], and patients classified as high risk with comorbidities may benefit from using the patient portal messaging system [24,37-39], which extends to the importance of portal use during widespread natural disasters to address potentially urgent or serious medical care. Given that patients with health conditions who use secure electronic messaging may show an improvement in health outcomes [40,41], it could be a valuable extension of medical care to encourage patients with ongoing medical needs (eg, chronic comorbidities) to actively use the portal, particularly in emergency circumstances.

The significance of prior use lends support for portal training interventions or enrollment efforts to decrease racial, ethnic, and age disparities seen in the uptake of digital health tools [42,43]. A recent systematic review on barriers and facilitators to patient portal use noted the importance of prior use of the portal in combination with provider buy-in [44]. The Diffusion of Innovation Theory categorizes technology adoption based on time to adoption [45], which is influenced by the availability of, and familiarity with, technology [46] and relates to

technology literacy and comfort with sharing health information [47]. Our results also showed a continued disparity in the use of the portal by race and ethnicity, with a higher likelihood of accessing the patient portal during Hurricane Harvey for White, non-Hispanic/Latinx users, which indicates utility in targeting populations classified as high risk and vulnerable for portal enrollment and training because portal use has been linked to increased patient engagement [48-50].

The overall message frequency decreased during the week of the hurricane, likely because of power outages, flooding, and infrastructure instability, but our qualitative message analysis indicated that patient portal messages have an important role to play in rapidly communicating the urgent health needs of patients to providers in situations where hospital or physician access becomes extremely limited. The analysis of such a large number of inquiries and responses sent during a disaster situation serves as a natural needs assessment of critical problems surrounding patients' health and continuity of medical care during a disaster, including issues with care or medication access and technical and infrastructural issues caused by the storm. The portal uniquely allows providers the ability to address certain problems for their patients in the absence of routine operations during a disaster, which may potentially reduce the risk of treatment disruption for acute conditions (eg, postoperative care and disaster-related injuries) and management of chronic conditions (eg, diabetes, cancer, and hypertension), in addition to meeting or rerouting mental health needs as appropriate.

Although patient-provider communication was supported during the disaster through the portal, providers were unable to fully address all concerns in the messages sent by patients. In some cases, the medical advice sought was beyond a provider's ability to give through web-based messaging, necessitating a rerouting of care or a specialist referral. Providers were sometimes unable to comply with medication requests made during Harvey because of technical difficulties and rules governing controlled substances. It is also notable that not all communication during a disaster directly relates to the crisis but highlights the necessity of maintaining health care continuity for the many patients with ongoing care needs. Examining the successful uses and acknowledging the limitations of the patient portal as it is currently positioned provide the basis for extending the portal functionalities to create an intentionally designed, disaster-prepared patient portal.

Disaster-Prepared Portal Functionality

Key Logistical Suggestions

Proper disaster preparedness for health care is critical in areas where natural disasters, including hurricanes, blizzards, and wildfires, are frequent. Using the portal more readily in these scenarios could streamline communication but would likely require temporarily reducing restrictions to EHR access among health care providers within a given health care entity. Traditional workflows and contingency plans are typically inadequate during the rapidly evolving chaos of natural disasters. For example, cross-coverage teams for an entire facility or practice may not be able, or available, to communicate effectively with staff or patients. Facilitating access to patient information for broader cross-coverage and triage teams than would normally comprise an on-call or ride-out team is necessary to increase the pool of potential providers who have access to the EHR. Processes can be adopted to *precredential* qualified providers for access across practices or facilities in the event that this scenario occurs again. Doing so would allow a defined *disaster team* of providers, similar to on-call coverage, to respond to urgent needs through the patient portal, eliminating some of the delays we saw in communication during Harvey while reducing the burden of electronic communication systemwide for all clinicians during times of limited access. When operationalizing a disaster communication approach through the portal to include an expanded recipient list to enable rapid triaging of these messages in times of limited access, it should be noted that the implementation would likely not have a one-size-fits-all approach. On the basis of this study's findings, we offer key logistical suggestions below to realize an intentionally designed, disaster-prepared patient portal.

Triaging and Rerouting Needs During Crisis Situations

Common portal requests during our study period included a wide range of medical advice, and the patient portal could be used to triage and appropriately redirect some of these needs to available sources of care, specifically now with the increase in use of telemedicine visits. A common practice for portal communications is to assign a dedicated triage nurse to serve as the messaging gatekeeper for patient portal communications [51]; this same principle could be used in the intentional design of disaster-prepared patient portals for crisis-related requests. During a disaster, messages could be automatically triaged based on patients' responses to screening questions to direct or reroute patients to appropriate and available in-person visits or electronic consultations until normal care can resume. Using the patient portal for pressing care requests is not an intended use of the typical portal messaging system; however, as many of our messages reflected, this type of urgent request is likely to occur and increase in frequency in disaster situations.

Having an emergency-prepared portal that can triage patients based on need specifically in times of crisis is a means of preparing for times of disrupted health care access. For example, the use of patient portals to address mental health concerns has demonstrated positive impacts for patients [52,53], and providing guidance or redirection seems especially important for mental health–related needs during times of crisis. The patient portal could also have triaging questions specific to anxiety, stress, and mental health care (eg, "Would you like to speak to someone about your current level of stress?") and provide information on the available mental health services. All emergency medical incidents should be rerouted to readily available behavioral health patient portal–certified care providers [54-58].

Automating Medication Refills

Several messages in our study were related to prescription medication requests, which further highlights the necessity of care continuity during a disaster and raises the question and feasibility of being able to automate select pharmacy requests through the portal. An unknown duration of postdisaster recovery leaves many patients without adequate quantities of

medication for chronic and life-threatening conditions such as hypertension, diabetes, seizures, and cancer. Even in the presence of a disaster declaration policy that allowed pharmacists to prescribe a 30-day supply of medication without prescriber authorization [59], the volume of message content in our study pertaining to prescription refill requests suggests that either the requested medications fell outside the disaster declaration allowances (eg, schedule II medications, opioids, stimulants, and depressants) or that pharmacists, physicians, or patients were not aware of the policy. The portal is a means by which established patients can communicate their medication needs to their provider, and an intentionally designed disaster portal could enable a function to allow these patients to have refill requests be sent or redirected automatically to a new pharmacy or new location. Automating some of these requests using the functionality of the portal would provide a useful tool to patients in emergency situations where immediacy is heightened, thereby decreasing physician-required approvals during these disasters when providers too experience limited availability.

Prioritized Automated Rescheduling

Infrastructure deficiencies in telecommunication due to disaster damage may also hinder access to the standard pathways of medical communication. Rescheduling canceled appointments and procedures is a logistical challenge following days of closure because of a natural disaster. The patients' messages expressed concern regarding the delay of their scheduled appointments and the inability to reach providers through landline phones because of widespread infrastructure damage in the Houston area. The portal could be used to allow patients who had appointment cancelations to reschedule at their convenience once the health system has established providers' availability and return to work. This may be particularly useful for urgent (nonemergent) surgical procedures so that patients can make informed decisions on when they would be able to receive treatment versus pursuing alternative treatment locations. The feasibility and implementation of automating this rescheduling feature for a prioritized listing of patients based on unavoidable disaster delays would be a topic for an important future study to extend current portal functionalities. Although patients would likely benefit from a prioritized automated rescheduling patient portal functionality for disaster use, appropriate conditions must be in place to allow a smooth transition between normal and disaster operations.

Implications for Portal Use in Future Disasters

The 2020 hurricane season was uniquely challenging in that the highest number of storms on record—30 storms [60]—formed, in addition to the beginning of the COVID-19 global pandemic that occurred in the months before the start of the hurricane season. These large-scale disasters occurring simultaneously created unprecedented, long-term health care challenges. As cities and states across the United States reached peak COVID-19 cases and hospitalizations, health care systems relied on public health guidance to limit in-person patient appointments and services [61,62]. There has been a large increase in the adoption of the patient portal in terms of general use and depth of use [63], including substantial dependence on the use of telehealth platforms and portal messaging for care management of patients [61,64]. Medical information technology allowed health care providers a key method with which to continue accommodating patient services safely while reserving hospital capacity for the most critically ill. Although the time frames of these disasters differ, inferences for health management can be drawn from both disaster situations to inform how the patient portal can facilitate health care during ongoing and future disaster operations.

Notably, singular disaster events such as Hurricane Harvey resulted in a short-term decrease in the overall frequency of patient portal messages and portal features, whereas the COVID-19 pandemic has necessitated a long-term demand for tools and platforms supporting virtual care (eg, telehealth or virtual visits and messaging through portals or other technology) that have increased the adoption of digital health tools. Both disasters illustrate the inherent value of information technology tools in providing quality care conventions that are typically attributed to direct, in-person patient-provider interactions. In addition, electronic consultations offer the benefit of disease containment and increased access to care, specifically for patients living in rural areas, and can take place without increasing time or travel demands for patients, pivotal for natural disasters, while limiting human contact and potential virus spread during this pandemic period.

Limitations

This study includes several limitations. First, this study was conducted at a single health care system; therefore, the results have limited generalizability because portal use may differ by region. However, because the study site is located in a large, demographically diverse urban city, these findings may be applicable to other health systems to support the integration of preparedness and medicine during emergency events. Second, we noted that there were some messages where a proxy user (eg, an adult family member messaging on behalf of a patient) self-identified in the message; however, we were unable to verify proxy users through portal accounts because this variable was not a component of our data set. Third, the comorbidity data were based on discharge disposition, which accompanies all encounter types; however, not all patients had an encounter in 2017. Thus, it is possible that we have underrepresented the incidence of the disease. The comorbidity data were not included in the final multivariate models for this reason; instead, we used encounter data, realizing that although it is not a replacement for controlling for disease severity, there is likely some correlation. Fourth, we were unable to control for, or examine, differential portal use by payer type but suggest future researchers do so. Finally, this study's needs assessment relied on the analysis of retrospective data, which should be complemented by user-centered prospective methods of deriving functional requirements to ensure well-informed intentional design [65].

Conclusions

The potential uses for the patient portal, classified and categorized into themes, lend support for the broader integration of patient portals as part of routine care coordination and management. Identifying patients' health needs that were

communicated through electronic patient portal messaging during an acute, widely destructive natural disaster that resulted in severe erosion of infrastructure and medical care is the first step in expanding the use of the portal for disaster health management. Future research should use this natural needs assessment and engage providers, informaticists, and data architects to expand the capabilities of the patient portal in disaster situations when normal care operations are limited and patient needs are extreme. Future research and practice should also consider how to use the portal for disaster planning relevant to chronic care and pregnancy. We raised considerations for disaster-specific elements of the portal interface based on patients' met and unmet needs. Most importantly, enabling provider teams to respond to patient needs during disasters would alleviate delays when provider access is inhibited because of the circumstances of the disaster, which was demonstrated in some of our provider messages (eg, internet access, power outages, and flooding). Our study conveyed the need for immediate medical advice during a disaster period when infrastructure is severely debilitated and typical emergency care is nonoperational or not feasible because of safety concerns. The patient portal could serve as an intentional bridge during times of disaster when the typical workflow structure of health care disintegrates.

Authors' Contributions

TM and RAP conceptualized the research aim; TM outlined the research design and supervised the project; JN, JMK, and SLJ contributed to the study design; JN acquired and cleaned the requisite data; JB and BSF conducted the qualitative coding; JB, AL, and TM conducted the quantitative analysis; JB, BSF, RAP, SLJ, and TM interpreted the study results; and JB, BSF, JMK, and TM drafted the manuscript. All authors reviewed, revised, and approved the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Descriptive statistics of the total 2017 population, the patient subgroups of interest, and a comparison between portal users and portal nonusers.
[DOCX File , 21 KB - jmir_v23i9e31264_app1.docx]

References

1. Abramson D, Garfield R. On the edge: children and families displaced by hurricanes Katrina and Rita face a looming medical and mental health crisis. Columbia University Libraries, New York, NY. 2006. URL: https://academiccommons.columbia.edu/doi/10.7916/D88S4Z4B [accessed 2021-08-16]
2. Weather, climate, and catastrophe insight: 2017 annual report. Aon Benfield. 2017. URL: https://reliefweb.int/sites/reliefweb.int/files/resources/20180124-ab-if-annual-report-weather-climate-2017.pdf [accessed 2021-08-16]
3. Karapiperis D. Natural catastrophes, insurance and alternate risk transfer. The Center for Insurance Policy and Research CIPR Newsletter. 2017. URL: https://www.naic.org/cipr_newsletter_archive/vol23.htm#story3 [accessed 2021-08-16]
4. Blake E, Zelinsky D. Hurricane Harvey. National Hurricane Center Tropical Cyclone Report. 2018. URL: https://www.nhc.noaa.gov/data/tcr/AL092017_Harvey.pdf [accessed 2021-08-16]
5. Amadeo K. Hurricane Harvey facts, damage and costs: what made Harvey so devastating. The Balance. 2019. URL: https://www.thebalance.com/hurricane-harvey-facts-damage-costs-4150087 [accessed 2020-09-24]
6. Phillips R, Schwartz R, McKeon W, Boom M. Lessons in Leadership: how the world's largest medical center braced for Hurricane Harvey. NEJM Catalyst. 2017. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0386 [accessed 2021-08-16]
7. Kousky C. Impacts of natural disasters on children. Future Child 2016;26(1):73-92. [doi: 10.1353/foc.2016.0004]
8. Alpert JM, Dyer KE, Lafata JE. Patient-centered communication in digital medical encounters. Patient Educ Couns 2017 May 02:1852-1858. [doi: 10.1016/j.pec.2017.04.019] [Medline: 28522229]
9. Irizarry T, DeVito DA, Curran CR. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
10. Peremislov D. Patient use of the electronic communication portal in management of type 2 diabetes. Comput Inform Nurs 2017 Mar 21:473-482. [doi: 10.1097/CIN.0000000000000348] [Medline: 28323648]
11. What is a patient portal? HealthIT.gov. URL: https://www.healthit.gov/faq/what-patient-portal [accessed 2021-03-17]
12. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
13. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
14. Wright A, Feblowitz J, Samal L, McCoy AB, Sittig DF. The Medicare Electronic Health Record Incentive Program: provider performance on core and menu measures. Health Serv Res 2014 Feb;49(1 Pt 2):325-346 [FREE Full text] [doi: 10.1111/1475-6773.12134] [Medline: 24359554]

15. Heath S. Patient access to health data, portal access surges to 92%. PatientEngagementHIT. 2016. URL: https://patientengagementhit.com/news/patient-access-to-health-data-portal-access-surges-to-92 [accessed 2021-03-17]
16. Heath S. Patient portal adoption tops 90%, but strong patient use is needed. PatientEngagementHIT. 2018. URL: https://patientengagementhit.com/news/patient-portal-adoption-tops-90-but-strong-patient-use-is-needed [accessed 2021-03-17]
17. Wolff JL, Darer JD, Berger A, Clarke D, Green JA, Stametz RA, et al. Inviting patients and care partners to read doctors' notes: OpenNotes and shared access to electronic medical records. J Am Med Inform Assoc 2016 Aug 07:166-172. [doi: 10.1093/jamia/ocw108] [Medline: 27497795]
18. Haun JN, Lind JD, Shimada SL, Martin TL, Gosline RM, Antinori N, et al. Evaluating user experiences of the secure messaging tool on the Veterans Affairs' patient portal system. J Med Internet Res 2014;16(3):e75 [FREE Full text] [doi: 10.2196/jmir.2976] [Medline: 24610454]
19. Mazor KM, Beard RL, Alexander GL, Arora NK, Firneno C, Gaglio B, et al. Patients' and family members' views on patient-centered communication during cancer care. Psychooncology 2013 Nov;22(11):2487-2495 [FREE Full text] [doi: 10.1002/pon.3317] [Medline: 23780672]
20. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc 2013 May 1;20(3):519-525 [FREE Full text] [doi: 10.1136/amiajnl-2012-001253] [Medline: 23242764]
21. Zhong X, Liang M, Sanchez R, Yu M, Budd PR, Sprague JL, et al. On the effect of electronic patient portal on primary care utilization and appointment adherence. BMC Med Inform Decis Mak 2018 Dec 16;18(1):84 [FREE Full text] [doi: 10.1186/s12911-018-0669-8] [Medline: 30326876]
22. Alturkistani A, Greenfield G, Greaves F, Aliabadi S, Jenkins RH, Costelloe C. Patient portal functionalities and uptake: systematic review protocol. JMIR Res Protoc 2020 Jul 31;9(7):e14975 [FREE Full text] [doi: 10.2196/14975] [Medline: 32734928]
23. Steitz B, Cronin RM, Davis SE, Yan E, Jackson GP. Long-term patterns of patient portal use for pediatric patients at an academic medical center. Appl Clin Inform 2017 Aug 02;8(3):779-793 [FREE Full text] [doi: 10.4338/ACI-2017-01-RA-0005] [Medline: 28765865]
24. Griffin A, Skinner A, Thornhill J, Weinberger M. Patient Portals: Who uses them? What features do they use? And do they reduce hospital readmissions? Appl Clin Inform 2016;7(2):489-501. [doi: 10.4338/ACI-2016-01-RA-0003] [Medline: 27437056]
25. Portz JD, Brungardt A, Shanbhag P, Staton EW, Bose-Brill S, Lin C, et al. Advance care planning among users of a patient portal during the COVID-19 pandemic: retrospective observational study. J Med Internet Res 2020 Aug 11;22(8):e21385 [FREE Full text] [doi: 10.2196/21385] [Medline: 32716900]
26. Patel PD, Cobb J, Wright D, Turer R, Jordan T, Humphrey A, et al. Rapid development of telehealth capabilities within pediatric patient portal infrastructure for COVID-19 care: barriers, solutions, results. J Am Med Inform Assoc 2020 Apr 17:1116-1120 [FREE Full text] [doi: 10.1093/jamia/ocaa065] [Medline: 32302395]
27. Bhatia AK, Waldman L, Hunt M, Leader A, Palidora J, Sabonjian M, et al. Creation of a telehealth task force to improve successful use of telehealth to maintain patient access during the COVID-19 pandemic. J Clin Oncol 2020 Oct 10;38(29_suppl):264. [doi: 10.1200/jco.2020.38.29_suppl.264]
28. Robinson JR, Valentine A, Carney C, Fabbri D, Jackson GP. Complexity of medical decision-making in care provided by surgeons through patient portals. J Surg Res 2017 Jun 15;214:93-101 [FREE Full text] [doi: 10.1016/j.jss.2017.02.077] [Medline: 28624066]
29. North F, Crane SJ, Stroebel RJ, Cha SS, Edell ES, Tulledge-Scheitel SM. Patient-generated secure messages and eVisits on a patient portal: are patients at risk? J Am Med Inform Assoc 2013;20(6):1143-1149 [FREE Full text] [doi: 10.1136/amiajnl-2012-001208] [Medline: 23703826]
30. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [doi: 10.1016/0021-9681(87)90171-8] [Medline: 3558716]
31. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med 2009 Nov 10;28(25):3083-3107 [FREE Full text] [doi: 10.1002/sim.3697] [Medline: 19757444]
32. Menser T, Kolman J, Sasangohar F, Phillips R, Jones S, Tian Y, et al. Patient portal usage during Hurricane Harvey - abstract. AcademyHealth: Annual Research Meeting. 2018. URL: https://academyhealth.confex.com/academyhealth/2018arm/meetingapp.cgi/Paper/26029 [accessed 2021-06-30]
33. Menser T, Baek J, Nicolas J, Jones S, Kash B. Use of patient portal during Hurricane Harvey in 2017: a qualitative study. American Public Health Association Annual Meeting. 2019. URL: https://academyhealth.confex.com/academyhealth/2018arm/meetingapp.cgi/Paper/26029 [accessed 2021-08-16]
34. Hewitt-Taylor J. Use of constant comparative analysis in qualitative research. Nurs Stand 2001;15(42):39-42. [doi: 10.7748/ns2001.07.15.42.39.c3052] [Medline: 12212430]
35. Quirkos. URL: https://www.quirkos.com [accessed 2021-04-01]

36. Lafata JE, Miller CA, Shires DA, Dyer K, Ratliff SM, Schreiber M. Patients' adoption of and feature access within electronic patient portals. Am J Manag Care 2018 Nov 01;24(11):352-357 [FREE Full text] [Medline: 30452203]
37. Han H, Gleason KT, Sun C, Miller HN, Kang SJ, Chow S, et al. Using patient portals to improve patient outcomes: systematic review. JMIR Hum Factors 2019 Dec 19;6(4):e15038 [FREE Full text] [doi: 10.2196/15038] [Medline: 31855187]
38. Antonio MG, Petrovskaya O, Lau F. The state of evidence in patient portals: umbrella review. J Med Internet Res 2020 Nov 11;22(11):e23851 [FREE Full text] [doi: 10.2196/23851] [Medline: 33174851]
39. Coquet J, Blayney DW, Brooks JD, Hernandez-Boussard T. Association between patient-initiated emails and overall 2-year survival in cancer patients undergoing chemotherapy: Evidence from the real-world setting. Cancer Med 2020 Nov;9(22):8552-8561 [FREE Full text] [doi: 10.1002/cam4.3483] [Medline: 32986931]
40. Zhou YY, Kanter MH, Wang JJ, Garrido T. Improved quality at Kaiser Permanente through e-mail between physicians and patients. Health Aff (Millwood) 2010 Jul;29(7):1370-1375 [FREE Full text] [doi: 10.1377/hlthaff.2010.0048] [Medline: 20606190]
41. Avdagovska M, Menon D, Stafinski T. Capturing the impact of patient portals based on the quadruple aim and benefits evaluation frameworks: scoping review. J Med Internet Res 2020 Dec 08;22(12):e24568 [FREE Full text] [doi: 10.2196/24568] [Medline: 33289677]
42. Grossman LV, Creber RM, Benda NC, Wright D, Vawdrey DK, Ancker JS. Interventions to increase patient portal use in vulnerable populations: a systematic review. J Am Med Inform Assoc 2019 Aug 01;26(8-9):855-870. [doi: 10.1093/jamia/ocz023] [Medline: 30958532]
43. Ramsey A, Lanzo E, Huston-Paterson H, Tomaszewski K, Trent M. Increasing patient portal usage: preliminary outcomes from the MyChart genius project. J Adolesc Health 2018 Jan;62(1):29-35 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.08.029] [Medline: 29169768]
44. Zhao JY, Song B, Anand E, Schwartz D, Panesar M, Jackson GP, et al. Barriers, facilitators, and solutions to optimal patient portal and personal health record use: a systematic review of the literature. AMIA Annu Symp Proc 2017;2017:1913-1922 [FREE Full text] [Medline: 29854263]
45. Daniel L. Examining the usage barriers for patient portal systems with patients ages over 45. Dissertations & Theses: Capella University. 2020. URL: https://www.proquest.com/openview/2a403ce694983bd7b156ce1c4ee217e8/1?pq-origsite=gscholar&cbl=18750&diss=y [accessed 2021-06-30]
46. Grossman LV, Creber RM, Ancker JS, Ryan B, Polubriaginof F, Qian M, et al. Technology access, technical assistance, and disparities in inpatient portal use. Appl Clin Inform 2019 Jan;10(1):40-50 [FREE Full text] [doi: 10.1055/s-0038-1676971] [Medline: 30650448]
47. Dalal AK, Bates DW, Collins S. Opportunities and challenges for improving the patient experience in the acute and postacute care setting using patient portals: the patient's perspective. J Hosp Med 2017 Dec;12(12):1012-1016. [doi: 10.12788/jhm.2860] [Medline: 29073310]
48. Price-Haywood EG, Luo Q. Primary care practice reengineering and associations with patient portal use, service utilization, and disease control among patients with hypertension and/or diabetes. Ochsner J 2017;17(1):103-111 [FREE Full text] [Medline: 28331456]
49. Creber R, Prey J, Ryan B, Alarcon I, Qian M, Bakken S, et al. Engaging hospitalized patients in clinical care: study protocol for a pragmatic randomized controlled trial. Contemp Clin Trials 2016 Mar;47:165-171 [FREE Full text] [doi: 10.1016/j.cct.2016.01.005] [Medline: 26795675]
50. Sarkar U, Lyles CR, Parker MM, Allen J, Nguyen R, Moffet HH, et al. Use of the refill function through an online patient portal is associated with improved adherence to statins in an integrated health system. Med Care 2014 Mar;52(3):194-201 [FREE Full text] [doi: 10.1097/MLR.0000000000000069] [Medline: 24374412]
51. Patient portal increases communication between patients and providers. HealthIT.gov. 2018. URL: https://www.healthit.gov/case-study/patient-portal-increases-communication-between-patients-and-providers [accessed 2021-07-01]
52. Turvey CL, Lindsay JA, Chasco EE, Klein DM, Fuhrmeister LA, Dindo LN. Current practices in electronic capture of patient-reported outcomes for measurement-based care and the use of patient portals to support behavioral health. Psychiatr Clin North Am 2019 Dec;42(4):635-647 [FREE Full text] [doi: 10.1016/j.psc.2019.08.006] [Medline: 31672213]
53. Sasseville M, LeBlanc A, Boucher M, Dugas M, Mbemba G, Tchuente J, et al. Digital health interventions for the management of mental health in people with chronic diseases: a rapid review. BMJ Open 2021 Apr 05;11(4):e044437 [FREE Full text] [doi: 10.1136/bmjopen-2020-044437] [Medline: 33820786]
54. Espinel Z, Galea S, Kossin JP, Caban-Aleman C, Shultz JM. Climate-driven Atlantic hurricanes pose rising threats for psychopathology. Lancet Psychiatry 2019 Sep;6(9):721-723. [doi: 10.1016/S2215-0366(19)30277-9] [Medline: 31327708]
55. Taioli E, Tuminello S, Lieberman-Cribbin W, Bevilacqua K, Schneider S, Guzman M, et al. Mental health challenges and experiences in displaced populations following Hurricane Sandy and Hurricane Harvey: the need for more comprehensive interventions in temporary shelters. J Epidemiol Community Health 2018 Oct;72(10):867-870. [doi: 10.1136/jech-2018-210626] [Medline: 29706599]
56. Schwartz RM, Tuminello S, Kerath SM, Rios J, Lieberman-Cribbin W, Taioli E. Preliminary assessment of hurricane harvey exposures and mental health impact. Int J Environ Res Public Health 2018 May 13;15(5):974 [FREE Full text] [doi: 10.3390/ijerph15050974] [Medline: 29757262]

57. Fitzpatrick KM. Post-traumatic stress symptomatology and displacement among Hurricane Harvey survivors. Soc Sci Med 2021 Feb;270:113634. [doi: 10.1016/j.socscimed.2020.113634] [Medline: 33433371]
58. Behavioral health data exchange/ primary care and behavioral health integration. HealthIT.gov. 2019. URL: https://www.healthit.gov/topic/health-it-health-care-settings/behavioral-health-data-exchange-primary-care-and-behavioral [accessed 2021-08-16]
59. Emergency Dispensing of Prescription Medications. Texas Administrative Code. 2007. URL: https://www.pharmacy.texas.gov/files_pdf/2018_flood_notification.pdf [accessed 2021-08-16]
60. Record-breaking Atlantic hurricane season draws to an end. National Oceanic and Atmospheric Administration. 2020. URL: https://www.noaa.gov/media-release/record-breaking-atlantic-hurricane-season-draws-to-end [accessed 2021-03-17]
61. Monaco A, Palmer K, Faber N, Kohler I, Silva M, Vatland A, et al. Digital health tools for managing noncommunicable diseases during and after the COVID-19 pandemic: perspectives of patients and caregivers. J Med Internet Res 2021 Jan 29;23(1):e25652 [FREE Full text] [doi: 10.2196/25652] [Medline: 33464206]
62. Koonin LM, Hoots B, Tsang CA, Leroy Z, Farris K, Jolly T, et al. Trends in the use of telehealth during the emergence of the COVID-19 pandemic - United States, January-March 2020. MMWR Morb Mortal Wkly Rep 2020 Oct 30;69(43):1595-1599 [FREE Full text] [doi: 10.15585/mmwr.mm6943a3] [Medline: 33119561]
63. Heath S. Was COVID-19 healthcare's use case for the patient portal? PatientEngagementHIT. 2020. URL: https://patientengagementhit.com/features/was-covid-19-healthcares-use-case-for-the-patient-portal [accessed 2021-07-01]
64. Reddy A, Arthur J, Dalal S, Hui D, Subbiah I, Wu J, et al. Rapid transition to virtual care during the COVID-19 epidemic: experience of a supportive care clinic at a tertiary care cancer center. J Palliat Med 2021 Feb 02:0737. [doi: 10.1089/jpm.2020.0737] [Medline: 33535019]
65. Sadeghi M, Thomassie R, Sasangohar F. Objective assessment of functional information requirements for patient portals. Proc Hum Factors Ergon Soc Annu Meet 2017 Oct 20;61(1):1788-1792. [doi: 10.1177/1541931213601928]

Abbreviations

ED: emergency department
EHR: electronic health record
OR: odds ratio
SMD: standardized mean difference

Edited by G Eysenbach; submitted 09.07.21; peer-reviewed by E Staton; comments to author 30.07.21; revised version received 02.08.21; accepted 02.08.21; published 01.09.21.

Please cite as:
Baek J, Simon-Friedt B, Lopez A, Kolman JM, Nicolas J, Jones SL, Phillips RA, Menser T
Assessing Patient Needs During Natural Disasters: Mixed Methods Analysis of Portal Messages Sent During Hurricane Harvey
J Med Internet Res 2021;23(9):e31264
URL: https://www.jmir.org/2021/9/e31264
doi:10.2196/31264
PMID:34468328

Review

The Impact of Digital Patient Portals on Health Outcomes, System Efficiency, and Patient Attitudes: Updated Systematic Literature Review

Elettra Carini[1*], MD; Leonardo Villani[1*], MD; Angelo Maria Pezzullo[1], MD, MSc; Andrea Gentili[1], MD; Andrea Barbara[2,3], MD; Walter Ricciardi[1], MD, MPH; Stefania Boccia[1,4], MSc, PhD

[1]Section of Hygiene, University Department of Life Sciences and Public Health, Università Cattolica del Sacro Cuore, Rome, Italy

[2]Hygiene and Public Health Service, ASL Roma 1, Rome, Italy

[3]Department of Public Health and Infectious Diseases, Sapienza University of Rome, Rome, Italy

[4]Department of Women, Children and Public Health Sciences - Public Health Area, Fondazione Policlinico Universitario A Gemelli IRCCS, Rome, Italy

[*]these authors contributed equally

Corresponding Author:
Angelo Maria Pezzullo, MD, MSc
Section of Hygiene
University Department of Life Sciences and Public Health
Università Cattolica del Sacro Cuore
L go F Vito 1
Rome, 00168
Italy
Phone: 39 0630154396
Email: angelomaria.pezzullo01@icatt.it

Abstract

Background: Patient portals are becoming increasingly popular worldwide even though their impact on individual health and health system efficiency is still unclear.

Objective: The aim of this systematic review was to summarize evidence on the impact of patient portals on health outcomes and health care efficiency, and to examine user characteristics, attitudes, and satisfaction.

Methods: We searched the PubMed and Web of Science databases for articles published from January 1, 2013, to October 31, 2019. Eligible studies were primary studies reporting on the impact of patient portal adoption in relation to health outcomes, health care efficiency, and patient attitudes and satisfaction. We excluded studies where portals were not accessible for patients and pilot studies, with the exception of articles evaluating patient attitudes.

Results: Overall, 3456 records were screened, and 47 articles were included. Among them, 11 studies addressed health outcomes reporting positive results, such as better monitoring of health status, improved patient-doctor interaction, and improved quality of care. Fifteen studies evaluated the impact of digital patient portals on the utilization of health services with mixed results. Patient characteristics were described in 32 studies, and it was reported that the utilization rate usually increases with age and female gender. Finally, 30 studies described attitudes and defined the main barriers (concerns about privacy and data security, and lack of time) and facilitators (access to clinical data and laboratory results) to the use of a portal.

Conclusions: Evidence regarding health outcomes is generally favorable, and patient portals have the potential to enhance the doctor-patient relationship, improve health status awareness, and increase adherence to therapy. It is still unclear whether the use of patient portals improves health service utilization and efficiency.

(J Med Internet Res 2021;23(9):e26189) doi:10.2196/26189

KEYWORDS

digital health; patient portal; health outcomes; patient satisfaction; patient attitudes; health service research; health care efficiency

Introduction

In recent years, electronic tools that allow patients to interact with health care professionals have considerably increased with consequences on the awareness of citizens about their own health [1]. The adoption of these technologies might represent an important measure to improve the quality and efficiency of health care services and is a key feature for the construction of a more equitable, effective, and safe health care system [2]. Indeed, the rapid growth and diffusion of digital health, including health information sources, such as electronic medical records (EMRs), has made online access to information by patients and health care professionals a crucial component of health care delivery [3].

In this context, patient portals are thought to allow patients secure access to health-related information and to communicate and share information with providers [4]. Besides guaranteeing protected access to EMRs, more advanced patient portals allow secure message exchange between health professionals and citizens, consultation of educational material adapted to patients' own characteristics, appointment scheduling, automatic renewal of medical prescriptions for chronic diseases, and facilitation of payments. Despite their potential benefits, several studies have proved underuse or inappropriate use of patient portals and their limited impact [5]. Furthermore, the majority of studies available on this topic have focused on users' characteristics and satisfaction, and few studies have considered the consequences on health outcomes [6-8]. Patient portals are relatively new technologies with continuous updates. Several types are released every year, and this may explain the lack of research in this area [5].

A systematic literature review in 2013, which addressed the effect of patient portals on patient clinical care, reported that evidence was limited to evaluate whether patient portals had a positive, negative, or neutral impact on users' health [4]. Some of the most effective examples refer to patients with chronic diseases, such as diabetes, hypertension, and depression, specifically when the portal is included in a shared health care pathway [9-13]. The effect of patient portals on health care utilization and efficiency, instead, is unclear due to the scarcity of studies examining the impact of patient portals on key indicators, such as inpatient hospitalizations, emergency department (ED) and outpatient visits, length of stay, and telephone contacts [14]. The aim of this systematic review was to update the study performed in 2013, by summarizing evidence on the impact of digital patient portals on patients' health outcomes, health care efficiency, and patients' attitudes and satisfaction.

Methods

Search Strategy

A search of relevant articles was performed in the PubMed and Web of Science databases using the query reported in Multimedia Appendix 1. The resulting records were entered in a dedicated work sheet to be subsequently screened according to the inclusion/exclusion criteria. Following the removal of duplicates, the selection was made by reading titles and abstracts, and then the full texts.

Inclusion/Exclusion Criteria

Eligible studies were primary studies reporting on the impact of patient portal adoption in relation with health outcomes, health care efficiency, and patients' attitudes and satisfaction. Articles included were published from January 1, 2013, to October 31, 2019, and written in English, Italian, Spanish, or French. We excluded studies describing portals that were not accessible for patients, as well as pilot studies, with the exception of studies evaluating patients' attitudes.

Selection Process and Data Extraction

Two authors screened the articles, and each reference retrieved was screened by two researchers independently, with any disagreement finally discussed and resolved by a third researcher, if necessary. The following information was extracted from the studies: first author name, publication year, study country, study design, study population, study setting, study duration and time, health information technology, study objective, main findings according to health outcomes, health care efficiency/utilization, patient characteristics, and patient attitudes and satisfaction. The systematic literature review was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2009 checklist [15].

Results

Characteristics of the Included Studies

The database search, after duplicate removal, identified a total of 3456 records. According to the inclusion/exclusion criteria, the screening resulted in the inclusion of 47 full-text articles (Figure 1).

The study designs were grouped into six categories according to the characteristics of the articles. Overall, 17 were descriptive quantitative studies [8,16-31], two were descriptive mixed-methods studies [32,33], 14 were observational hypothesis testing studies [20,34-46], seven were descriptive qualitative studies [47-53], five were interventional studies, other than randomized controlled trials (RCTs) [54-58], and three were RCTs [59-61].

With regard to country, 33 (70%) studies were based in the United States [17,18,20,21,31,32,34,43-46,49,60], three in Canada [35,36,50], three in the Netherlands [19,26,57], two in Finland [55,56], two in the United Kingdom [16,48], one in Australia [22], one in France [59], one in Israel [47], and one in Sweden [24].

Various patient portals have been described in the studies. Multimedia Appendix 2 provides details on the functionalities of the portals, and Multimedia Appendix 3 provides qualitative descriptions of the portals.

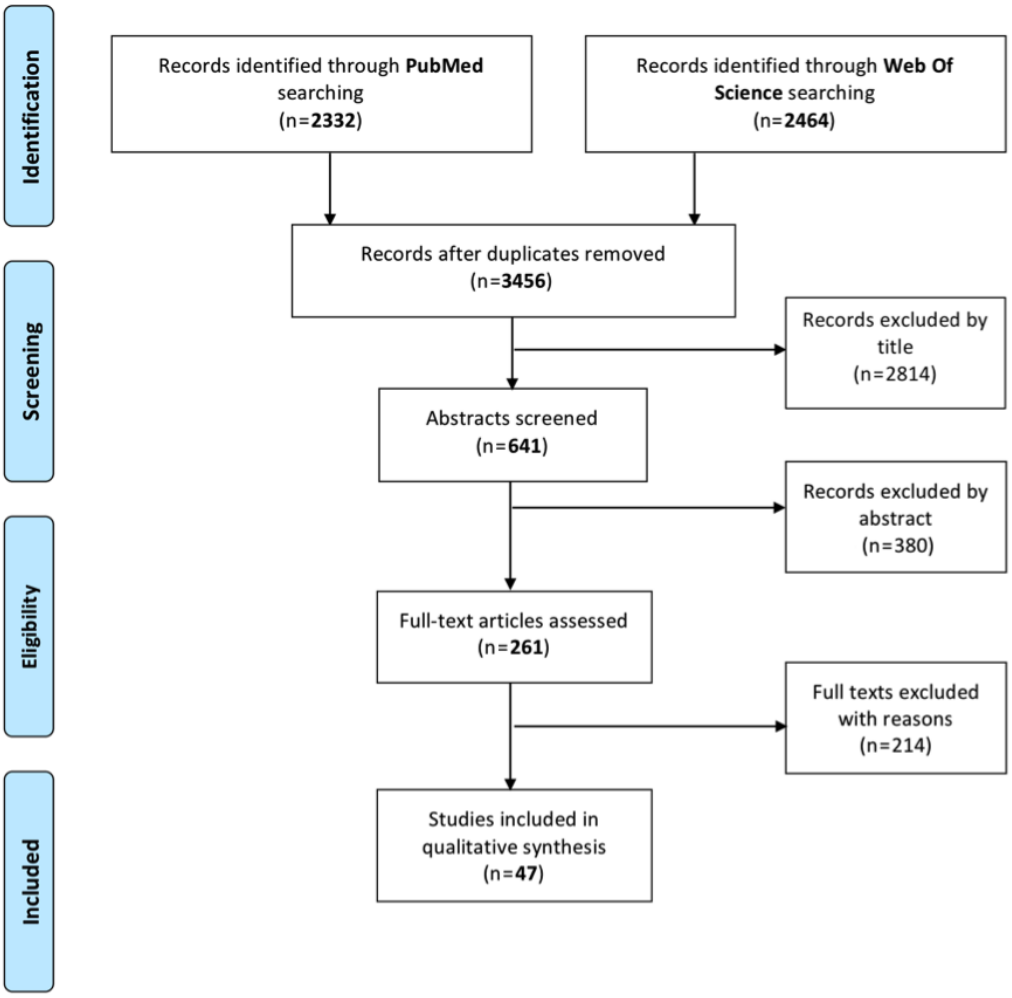
Most of the portals were not addressed to a defined population subgroup, and only some of them were specific to a clinical specialty/condition, such as endocrinology-diabetes [8,21,26,34,36,41], primary care [33,37,38,40,60], mental health

[31,35,54], multiple chronic conditions [40,55,56], pulmonology and asthma [32,44,60], rheumatology [50,57,59], cardiology [37,44], internal medicine [40,61], nephrology [30,46], pregnancy [34], cancer [53], and gastroenterology [48].

The population included in the studies was heterogeneous in terms of sample size (from 24 [52] to 2,171,325 patients [31]) and groups of included patients (eg, pediatric [32], older [31], oncology [52], and diabetic patients [60]).

The results were summarized in four categories, albeit the same study could belong to more than one category. In particular, 11 studies analyzed health outcomes and adherence, intended as a change in individual or population health, attributable to health-related interventions. Adherence is the degree to which a patient follows medical advice, especially drug compliance. Overall, 15 studies focused on health care efficiency/utilization (utilization of health care services), 32 studies referred to patient characteristics, and 30 studies analyzed attitudes and satisfaction.

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



Health Outcomes and Adherence

Eleven articles presented results on patient outcomes (Table 1), such as prevention, diabetes, blood pressure control, asthma, mental health, and medication adherence.

In particular, a significant association was found between patients' preventive health behaviors and portal use [45]. Considering diabetes [34,36] and blood pressure control [37], portal users were significantly more likely to control their HbA_{1c} levels successfully compared to nonusers.

An improved clinical condition over time was observed in the management of asthma in children, where the group using the patient portal reported better flare control compared to the control group [60].

Concerning mental health, patients using the portal had a positive impact related to the clinical condition in many domains [35], such as the reduction of drinking days [54]. However, concerning the health status, no marked short-term impact was described, and differences in patient-reported physical and mental health changes were minor [55].

Finally, higher medication adherence was described in portal users compared with nonusers, especially among pediatric patients with asthma and patients with rheumatic disorders [32,57].

Table 1. Summary of the findings on health outcomes and adherence.

First author, year	Country	Health information technology	Study design	Sample size	Main findings
Ancker, 2019 [34]	United States	Blood glucose flow sheet (EpicCare and Weill Cornell Connect portal)	Observational, hypothesis testing	53 patients	Pregnant: average BMI dropped while average blood pressure increased significantly more in the 9 months among uploaders than nonuploaders. Chronic disease patients: after 9 months, uploaders had significantly larger reductions in HbA_{1c} and BMI than nonuploaders. One subset of uploaders had low well-controlled HbA_{1c} values before and during PGHD[a] upload. Another uploader subset began to upload when their HbA_{1c} levels were elevated and experienced a decrease in HbA_{1c} levels followed by a plateau.
Fiks, 2015 [60]	United States	MyAsthma (clinical interface in MyChart)	RCT[b]	60 families of children	No significant differences in baseline control, quality of life, or parent activation between the two study arms (P>.2 for all comparisons). Frequency of asthma flares improved in the intervention group over time by 2.0 points on a 25-point scale (P=.02). Families in the intervention group had a marginally significant improvement in symptoms during periods without flares. A nonstatistically significant improvement in quality of life in terms of daytime symptoms and functional limitations was observed in the intervention group. There were no significant changes in parent activation.
Fiks, 2016 [32]	United States	MyAsthma	Descriptive, mixed methods	237 families	Portal users with uncontrolled asthma had significantly more medication changes after using the portal relative to the year earlier (increase of 14%).
Huang, 2019 [45]	United States	MyPennMedicine (branded version of Epic MyChart)	Observational, hypothesis testing	10,000 patients	Patients' preventive health behaviors were significantly associated with portal use. The proportions of annual flu vaccination, blood pressure checks, and lipid level screening were substantially higher in portal users compared with nonusers (OR[c]=1.58, 1.13, and 1.50, respectively; P<.001). The average composite prevention score was significantly higher among portal users compared with nonusers (mean difference=0.22; P<.001). The proportion of colorectal cancer screening between users and nonusers was statistically significant (P<.001, OR very close to 1). No clinically meaningful difference between patient portal users and nonusers in chronic health outcomes.
Jhamb, 2015 [46]	United States	Free patient portal tethered to an ambulatory EHR[d]	Observational, hypothesis testing	1098 patients	In the fully adjusted model (controlling for hyperlipidemia, nephrolithiasis, history of kidney transplant, CCI[e], proteinuria, eGFR[f], number of nephrology and outpatient visits, and university affiliated PCP[g]), the association was not significant (OR 1.11, 95% CI 0.99-1.24).
Kipping, 2016 [35]	Canada	Ontario Shores HealthCheck Patient Portal	Observational, hypothesis testing	91 patients	The overall Mental Health Recovery Measure score increased from 70.4 (SD 23.6) at baseline to 81.7 (SD 25.1) at follow-up (P=.01). Of the eight domains, seven increased from baseline to follow-up (overcoming stuckness, self-empowerment, basic functioning, overall well-being, new potentials, spirituality, and advocacy/enrichment; all P<.05. No change for learning and self-redefinition).
Lau, 2014 [36]	Canada	BCDiabetes.ca	Observational, hypothesis testing	1957 patients	Overall, 28 of 50 users had a follow-up HbA_{1c} ≤7%, whereas 22 of 50 did not (56% success rate). Only 16 of 50 nonusers achieved a follow-up HbA_{1c} ≤7%, while 34 of 50 did not (32% success rate). Users were significantly more likely to control their HbA_{1c} levels successfully than nonusers (McNemar test, P=.03). The HbA_{1c} level at the last follow-up was significantly lower for users compared to nonusers (P=.02).

First author, year	Country	Health information technology	Study design	Sample size	Main findings
Manard, 2016 [37]	United States	Online patient portal	Observational, hypothesis testing	1571 patients	After adjusting for age, users were more likely to achieve BP[h] control (HR[i] 1.24, 95% CI 1.06-1.45). After adjustment for sociodemographics, portal use was no longer associated with BP control (HR 0.98, 95% CI 0.83-1.16).
Quanbeck, 2018 [54]	United States	Seva	Interventional, other than RCT	268 patients	Significant reductions in the numbers of risky drinking days, which declined by 44% ([0.7-1.25]/1.25) from baseline to 6 months, and illicit drug-use days, which declined by 34% ([2.14-3.22]/3.22). Two of the three abstinence outcomes showed significant improvements (any illicit drug use and/or any drink or drug). Significant effects were found for two of the three quality of life scores (overall quality of life and mental health).
Riippa, 2015 [55]	Finland	Patient portal by The Finnish Medical Society, Duodecim	Interventional, other than RCT	876 patients	Minor differences in patient-reported physical and mental health changes that changed the sign from the matched (physical health mean=1.2, 95% CI −3.3 to 5.7; mental health mean=0.8, 95% CI −3.6 to 5.2) to the unmatched sample (physical health mean=−0.4, 95% CI −4.7 to 3.9; mental health mean=−0.4, 95% CI −4.8 to 4.0). Patient activation improved more in the intervention group, but it was not statistically significant. There was no marked short-term impact on health status based on the SF-36v2 measure.
Van der Vaart, 2014 [57]	Netherlands	Medisch Spectrum Twente	Interventional, other than RCT	360 patients	Overall, 56% of the respondents had a score of 7 (out of 8) on medication adherence.

[a]PGHD: patient-generated health data.

[b]RCT: randomized controlled trial.

[c]OR: odds ratio.

[d]EHR: electronic health record.

[e]CCI: Charlson Comorbidity Index.

[f]eGFR: estimated glomerular filtration rate.

[g]PCP: primary care practice.

[h]BP: blood pressure.

[i]HR: hazard ratio.

Efficiency/Utilization

Fifteen articles described the relationship between portal use and health care service efficiency and utilization (Table 2). The use of a digital portal had an effect on the utilization of health care services in terms of the number of clinical visits, especially for asthmatic patients [32,60], while no statistically significant changes in the number of primary care visits was reported in association with the use of secure messaging [38].

Table 2. Summary of the findings on health care efficiency.

First author, year	Country	Health information technology	Study design	Sample size	Main findings
Ancker, 2019 [34]	United States	Blood glucose flow sheet (EpicCare and Weill Cornell Connect portal)	Observational, hypothesis testing	53 patients	Uploaders had more clinical visits and portal logins before initial data upload.
Bidmead, 2016 [48]	England (United Kingdom)	Patients Know Best (PKB)	Descriptive, qualitative	56 patients	The portal enabled clinicians to manage stable patients, facilitating clinical and cost-effective use of specialist nurses, and improved two-way communication and more optimal use of outpatient appointments and consultant time. It also facilitated a single rationalized pathway for stable patients, enabling access to information and proactive support.
Fiks, 2015 [60]	United States	MyAsthma (clinical interface in MyChart)	RCT[a]	60 families of children	The intervention group had a marginally significant reduction in the proportion of parents missing at least 1 day of work (reduction of 47%, *P*=.07). Families in the intervention group reported fewer ED[b] visits and hospitalizations for asthma over 6 months than the control group (3 vs 9 and 0 vs 2, respectively). Only two intervention families reported at least one ED visit (vs six control families), and no intervention families reported hospitalizations. Children in the intervention group had fewer visits with asthma specialists or primary care. Results were similar on stratifying by asthma severity.
Fiks, 2016 [32]	United States	MyAsthma	Descriptive, mixed methods	237 families	Portal users with uncontrolled asthma had significantly more primary care asthma visits after using the portal than the year earlier (increases of 16%).
Foster, 2019 [43]	United States	Epic MyChart	Observational, hypothesis testing	208,635 tests	ED visits: 80.56% (n=20,430) of patients had a single ED visit with laboratory testing, 16.04% (n=4069) had two or three ED visits, 3.16% (n=802) had four to 10 ED visits, and only 0.24% (n=60) had more than 10 ED visits. Activation rates were lower for those with only a single ED visit (7312/20,430, 35.79%) compared with either those with two to three ED visits (1770/4069, 43.50%; *P*<.001) or four or more ED visits (368/862, 42.7%; *P*<.001).
Griffin, 2016 [44]	United States	My UNC Chart	Observational, hypothesis testing	2975 patients	The odds of being readmitted within 30 days for active users was 66% higher than that for nonusers, holding all other variables constant in the model. There was no significant difference in 30-day readmission between nonusers and light users.
Jahn, 2018 [49]	United States	My HealtheVet	Descriptive, qualitative	29 participants	Secure messaging tasks were inefficient as related to clinical document sharing (it took almost 5 minutes for providers to only attach and send a clinical document).
Kipping, 2016 [35]	Canada	Ontario Shores HealthCheck Patient Portal	Observational, hypothesis testing	91 patients	Fewer missed appointments and a reduced number of requests for information in the year following portal implementation. The odds of a portal user attending an appointment were 67% (CI 56%-79%) greater than for nonusers over the follow-up period. Compared with 2014, in 2015, there was an 86% and 57% decrease in requests for information among users and nonusers, respectively (61% overall).
North, 2014 [38]	United States	Mayo Clinic Health System	Observational, hypothesis testing	2357 primary care patients	Primary care patients who sent at least one secure message or e-visit had a mean of 2.43 (SD 2.3) annual face-to-face visits before the first message and 2.47 (SD 2.8) after, with a nonsignificant difference (*P*=.45). After adjustment for a first message surge in visits, no significant visit frequency differences were observed (mean, 2.35 annual visits per patient both before and after the first message; *P*=.93). Subgroup analysis showed no significant change in visit frequency for patients with higher message utilization or for those who had used the messaging feature longer.

First author, year	Country	Health information technology	Study design	Sample size	Main findings
Plate, 2019 [39]	United States	MyChart; Epic Systems Corporation	Observational, hypothesis testing	6426 patients	Active MyChart status was not associated with 90-day ED return (*P*=.78) or readmission (*P*=.51) based on univariable analysis. Similarly, during multivariable analysis controlling for age, gender, BMI, and ASA[c] category, active MyChart utilization was not significantly associated with 90-day ED visits (OR[d] 1.019, 95% CI 0.843-1.231; *P*=.85) or readmissions (OR 0.966, 95% CI 0.747-1.249; *P*=.79). Patients who sent secure messages within 90 days from surgery (2200 patients, 48% of active users) were not less likely to present to the ED (*P*=.63) or be readmitted (*P*=.59) within 90 days. For patients who sent two or more messages (1354 patients), provider or staff response rate <75% was significantly associated with 90-day readmission (*P*=.004) with greater 90-day ED visits that neared statistical significance (*P*=.07).
Quanbeck, 2018 [54]	United States	Seva	Interventional, other than RCT	268 patients	Significant reduction in hospitalizations and a trend toward fewer ER[e] visits. Increase in HIV screening rates. Change in the rates of HIV risk behaviors (eg, condom use) and receiving other addiction treatments appeared to be nonsignificant.
Riippa, 2015 [55]	Finland	Patient portal by The Finnish Medical Society, Duodecim	Interventional, other than RCT	876 patients	The effect on the cost of care was ambiguous; costs decreased by an average of €91 in the unadjusted model, but increased by €48 in the adjusted model. Due to the controversial result, the unadjusted analysis showed an 89% probability of cost-effectiveness with no willingness to pay for increased patient activation, whereas in the adjusted sample, the probability of the portal being more cost-effective than care as usual exceeded 50% at a willingness to pay €700 per clinically significant increase in the patient activation score. For doctor visits, portal access (n=80): 3.8 (SD 3.3) and control (n=57): 3.0 (SD 3.1) (*t*=1.4; *P*=.18). For nurse visits, portal access (n=80): 3.5 (SD 2.6) and control (n=57): 4.1 (SD 2.5) (*t*=−1.3; *P*=.18).
Tsai, 2019 [28]	United States	Epic's personal health record system	Descriptive, quantitative	109,200 patients	Active users had more outpatient and inpatient visits and fewer ER visits. Patients without a portal account had on average fewer outpatient visits per month (0.31 vs 0.89, *P*<.001) and fewer inpatient visits per month (0.007 vs 0.059, *P*<.001), but had more ER visits per month than patients who were active with the portal (0.047 vs 0.014, *P*<.001). The difference between no-show appointments was not significant.
Wallace, 2016 [58]	United States	MyChart by Epic health record system	Interventional, other than RCT	36,549 patients	The number of visits for 12 months was strongly associated with an increased likelihood of MyChart activation and with more frequent MyChart logins.

First author, year	Country	Health information technology	Study design	Sample size	Main findings
Zhong, 2018 [42]	United States	MyUFHealth (also known as MyChart by Epic)	Observational, hypothesis testing	15,659 nonusers and 5494 users	At the time of adoption, the quarterly PCP[f] office visit RR[g] of users to nonusers was 1.33 (95% CI 1.27-1.39; *P*<.001). The RRs were between 0.94 and 0.99 up to four quarters after portal adoption (*P*=.75, .10, .13, and .09, respectively), and it was significantly less than 1 at the seventh (RR 0.82, 95% CI 0.73-0.91; *P*<.001) and eighth (RR 0.80, 95% CI 0.70-0.90; *P*<.001) quarters post adoption. The no-show rate proxies in the user group were significantly lower than in the nonuser group. RRs were between 0.60 and 0.83 for eight out of 11 quarters, and for the remaining three quarters, differences were not significant (*P*=.65, .29, and .44, respectively). Differences in cancellation rate proxies were not significant (*P*>.05). Overall, appointment adherence improved after portal adoption.

[a]RCT: randomized controlled trial.

[b]ED: emergency department.

[c]ASA: American Society of Anesthesiology.

[d]OR: odds ratio.

[e]ER: emergency room.

[f]PCP: primary care physician.

[g]RR: rate ratio.

Concerning emergency room (ER) visits, a decreasing trend was described [54,60], and active portal users showed more outpatient and inpatient visits and fewer ER visits per month, compared with patients without an account [28]. The number of visits over 6 months for asthmatic patients was lower among users than among nonusers (3 vs 9) [60]. User status was not significantly associated with ED access in the study by Plate et al, and the use of secure messages showed a nonsignificant result [39].

A reduction in hospitalizations was described for asthma [60] and mental health conditions [54]. Different findings were observed in patients with acute myocardial infarction, congestive heart failure, or pneumonia. In these cases, the odds of 30-day readmission for active users was 66% higher than that for nonusers, while no significant difference was described between nonusers and light users [44]. Considering patients who underwent total hip or knee arthroplasty, there was a significant association between 90-day readmission and sending two or more secure messages [39].

Concerning clinicians' attitudes, the portal seemed to enable a new way of managing stable patients, facilitating clinical and cost-effective use of specialist nurses (improved two-way communication, and more optimal use of outpatient appointments and consultant time). The portal also facilitated a single rationalized pathway for stable patients, enabling access to information and proactive support [48].

Portal use by patients reduced missing appointments [35] and showed an improvement in appointment adherence after portal adoption [42]. However, no significant association between user/nonuser status and no-show appointments was found by one study [28]. The difference in cancellation rate proxies was also not significant between the user and nonuser groups [42].

Finally, concerning other aspects, only one study took into consideration the associated costs with ambiguous results. In this study, costs decreased in the unadjusted model but increased after adjusting for relevant variables [55]. An increase in HIV screening rates was described, but changes in the rates of HIV risk behaviors (eg, condom use) and modification of addiction treatments appeared nonsignificant in mental health patients [54]. Portal use had a positive effect on days of work lost due to asthma patients' issues [60]. Moreover, information services were positively affected by portal use, as it led to a reduced number of requests [35].

Patient Characteristics

Concerning patients' characteristics, 32 articles presented relevant results that were mainly related to demographic information, such as age, gender, education, and household status (Table 3).

Table 3. Summary of the findings on patient characteristics.

First author, year	Country	Health information technology	Study design	Sample size	Main findings
Abel, 2018 [31]	United States	My HealtheVet (MHV) and Clinical Video Telehealth (CVT)	Descriptive, quantitative	2,171,325 patients	African American and Latino patients were significantly less likely to engage in use compared with White patients. Low-income patients with free care were significantly less likely to be users. Patients with schizophrenia or schizoaffective disorder were significantly less likely to be users than those with other mental health diagnoses (OR[a] 0.50, CI 0.47-0.53 and OR 0.75, CI 0.69-0.80, respectively). Although rural patients had 17% lower odds of My HealtheVet adoption compared with urban patients (OR 0.83, 95% CI 0.80-0.87), they were more likely to engage in Clinical Video Telehealth and dual use (OR 2.45, 95% CI 1.95-3.09 for Clinical Video Telehealth and OR 2.11, 95% CI 1.81-2.47 for dual use).
Ancker, 2019 [34]	United States	Blood glucose flow sheet (EpicCare and Weill Cornell Connect portal)	Observational, hypothesis testing	53 patients	Pregnant patients: Uploaders were similar to the comparison group in terms of race, ethnicity, age, and socioeconomic status. Uploaders had more clinical encounters and portal logins before initial data upload, earlier establishment of patient portal accounts, and worse baseline blood pressure. 30 chronic disease patients: Uploaders were more likely to be Asian-American and younger, but the groups did not have other significant demographic differences.
Chan, 2018 [18]	United States	MyChart (EpicCare, Verona, WI)	Descriptive, quantitative	17,699 patients	Positive associations between portal use and being aged 50-74 years, White, privately insured, English-speaking, and living outside San Francisco.
Fiks, 2015 [60]	United States	MyAsthma (Clinical interface in MyChart)	RCT[b]	60 families of children	Parents of children with moderate or severe asthma used the portal more frequently (75% were frequent users vs 47% were parents whose child had mild persistent asthma).
Fiks, 2016 [32]	United States	MyAsthma	Descriptive, mixed methods	237 families	Users were more likely to have children aged 6-9 years (P=.009), be White (P<.001), be privately insured (P<.001), have mild persistent or moderate or severe persistent asthma (P=.002), be on an asthma controller medication (P<.001), and be receiving a greater number of asthma medications at baseline on average than nonusers (P<.001). Those with persistent asthma were twice as likely to use the portal versus those with intermittent asthma (2.37% vs 1.25% at hospital practices where these data were available; P<.001). Sustained portal users were more likely than one-time users to be Hispanic (P=.02), have private insurance (P=.02), and be from the Northeast (P=.001). Sustained use parents had higher educational levels (P=.002). Positively associated with portal adoption in multivariable logistic regression: receipt of a controller medication at baseline (OR 2.0, 95% CI 1.5-2.7), private insurance (OR 2.0, 95% CI 1.3-3.1), lower child age (OR 1.4, 95% CI 1.1-1.9), and greater asthma severity (OR 1.9, 95% CI 1.2-3.0 for mild and OR 1.9, 95% CI 1.0-3.5 for moderate or severe persistent vs intermittent).

First author, year	Country	Health information technology	Study design	Sample size	Main findings
Foster, 2019 [43]	United States	Epic MyChart	Observational, hypothesis testing	208,635 tests	Females (5546/13,149, 42.18%) were significantly more likely to activate the portal than males (3897/12,212, 31.91%; *P*<.001). Activation rates were highest for Asian (262/451, 58.1%) and White individuals (8155/20,637, 39.52%) and lower for African American/Black (491/2254, 21.78%; *P*<.001 compared with White), Hispanic/Latino (333/1257, 26.49%; *P*<.001 compared with White), and other individuals (241/762, 31.6%; *P*<.001 compared with White). The activation rate for patients aged 18-70 years was 41.61% (7593/18,246). The overall pattern of radiologic image viewing with respect to age and gender showed similar trends to those described for laboratory testing.
Gordon, 2016 [20]	United States	Kaiser Permanente Northern California patient portal	Observational, hypothesis testing/descriptive, quantitative	231,082 patients/4980 patients	Older seniors (aged 70-74 and 75-79 years) were significantly less likely than those aged 65-69 years to have registered, and to have used the patient portal to send a secure message, view laboratory test results online, or order prescription refills at least once by the end of the year. Slightly over 70% had been diagnosed with a chronic cardiovascular condition, and 90% reported taking at least one prescription medication for a chronic condition.
Gossec, 2017 [59]	France	Sanoia	RCT	320 patients	In multivariate analyses, the only factor related to connecting more than twice to the platform was being a member of a patient association (OR 1.44, 95% CI 1.17-1.77; *P*<.001). In the groups with high and low numbers of connections, the percentages of patient association members were 24.7% and 6.5%, respectively.
Griffin, 2016 [44]	United States	My UNC Chart	Observational, hypothesis testing	2975 patients	Active users had a higher proportion of Caucasian patients, higher Charlson Comorbidity scores, and a higher proportion of patients admitted to an academic medical center than light users.
Huang, 2019 [45]	United States	myPennMedicine (branded version of Epic MyChart)	Observational, hypothesis testing	10,000 patients	Users were more likely to be younger (63.46 years [users] vs 66.08 years [nonusers]; *P*<.001) and have higher income (US $74,172 [users] vs US $62,940 [nonusers]; *P*<.001) than nonusers. The percentage of White race was substantially higher among users (72.77% [4317/5932] [users] vs 52.58% [2139/4068] [nonusers]; *P*<.001). For users, the percentage of payments by commercial insurance was higher (60.99% [3618/5932] [users] vs 40.12% [1632/4068] [nonusers]; *P*<.001) and the percentage of payments by Medicare or Medicaid was lower (Medicare: 34.91% [2071/5932] [users] vs 48.72% [1982/4068] [nonusers]; *P*<.001; and Medicaid: 3.49% [207/5932] [users] vs 10.08% [410/4068] [nonusers]; *P*<.001). The difference in sex between users and nonusers was not statistically significant. No significant difference was found in any provider-level characteristic between the two groups.
Jhamb, 2015 [46]	United States	Free patient portal tethered to an ambulatory EHR[c]	Observational, hypothesis testing	1098 patients	Users were younger and more likely to be non-Black, be married, have private insurance, and have higher neighborhood median household income. Users were less likely to have diabetes, coronary artery disease, or congestive heart failure, but were more likely to have had a kidney transplant. Older age, Black race, unmarried status, Medicaid or Medicare insurance (vs private), and lower neighborhood median household income were associated with not using the portal.

First author, year	Country	Health information technology	Study design	Sample size	Main findings
Kipping, 2016 [35]	Canada	Ontario Shores HealthCheck Patient Portal	Observational, hypothesis testing	91 patients	A similar proportion of patients (1756/3158, 55.6%) and portal users (266/432, 61.6%) were female. Age distribution was relatively similar. The majority of users were between 25 and 34 years.
Krist, 2014 [33]	United States	AllscriptsTouchworks EHR	Descriptive, mixed methods	112,893 patients	Older patients were more likely to create a PHR[d] account as they had chronic conditions.
Laranjo, 2017 [22]	Australia	Portuguese National patient portal	Descriptive, quantitative	109,619 participants	Geographic analysis revealed higher proportions of PHR adoption in urban centers when compared with rural noncoastal districts.
Lau, 2014 [36]	Canada	BCDiabetes.ca	Observational, hypothesis testing	1957 patients	Users tended to be younger (mean difference of 4.28 years; *P*=.06), have lower baseline HbA_{1c} (mean difference of 0.89%; *P*<.01), and have higher baseline weight (mean difference of 7.53 kg; *P*=.06) than nonusers. There was no difference in gender or total follow-up time. Follow-up HbA_{1c} levels tended to be lower in users than nonusers (mean difference of 0.75%; *P*<.01), and users were significantly more likely to have HbA_{1c} of 7% at their last follow-up visit (*P*=.01). No significant differences in LDL[e] and SBP[f] were observed between users and nonusers at initial visits and follow-up visits.
Manard, 2016 [37]	United States	Online patient portal	Observational, hypothesis testing	1571 patients	Users were significantly younger (*P*<.001), more often White (*P*<.001), and more often married (*P*<.001) than nonusers. Users were significantly from upper-middle to the highest socioeconomic status compared with nonusers (*P*<.001). Portal use was more common among FM[g] patients than GIM[h] patients (*P*<.001), and users were more often high health care utilizers (*P*<.02). Portal use was less common among current smokers (*P*<.001). Users were more likely to have depression (*P*<.01) and lower comorbidity scores (*P*<.001).
Mishra, 2019 [23]	United States	OpenNotes within the HealtheLife patient portal	Descriptive, quantitative	1487 patients	Overall, 90% (n=784) were above 30 years, with 8% between 18 and 29 years; 40% were above 60 years; and 50% were between 30 and 59 years. One participant stated accessing the portal for a minor and another as a family surrogate. Moreover, 92% (n=797) had a college degree or greater, 24% (n=205) had a graduate degree, and 1% (n=12) had less than 12th grade education. Individuals in the 18-29 and >60 years groups were more likely to find the notes helpful. Greater note comprehension was correlated with greater education. Noncollege participants were more likely to access notes “many times” than college participants (*P*=.02).
North, 2014 [38]	United States	Mayo Clinic Health System	Observational, hypothesis testing	2357 primary care patients	The majority of patients sending messages were female, were White, lived locally, and were employed by the Mayo Clinic.
Plate, 2019 [39]	United States	MyChart; Epic Systems Corporation	Observational, hypothesis testing	6426 patients	Overall, 4623 people registered on MyChart logged into the patient portal at least once within 1 year from surgery, and 1803 (28%) patients were not registered users. Active users were significantly more likely to be young, have a healthy ASA[i] score (ASA 1 or 2), be Caucasian, be married, be employed, be privately insured, and be discharged to home. Patients not using MyChart had a higher ASA score (ASA 3 or 4) and were more likely to be African American, unmarried, and unemployed. Patients without MyChart were more likely to have Medicare or Medicaid insurance and be discharged to a skilled nursing facility.

First author, year	Country	Health information technology	Study design	Sample size	Main findings
Portz, 2019 [52]	United States	My Health Manager (Kaiser Permanente Colorado patient portal)	Qualitative descriptive study	24 patients	The mean age was 78 years. Patients were primarily White (12, 80%) and women (12, 80%). Education: high school graduate, 1 (7%); some college graduate, 7 (47%); and college graduate, 7 (47%). Income: <US $30,000, 2 (13%); US $30,000-49,999, 7 (47%); US $50,000-74,999, 2 (13%); >US $75,000, 2 (13%); chose not to answer, 2 (13%).
Powell, 2018 [40]	United States	FollowMyHealth portal	Observational, hypothesis testing	500 patients	No significant relationship between the number of logins and any of the demographic variables; however, when those with zero logins were removed from the model, age, distance separating the patient from his or her provider, and having a diagnosis of heart failure were all significant predictors of portal use (P<.05).
Price-Haywood, 2017 [25]	United States	MyOchsner patient portals (Epic System), wearable technology, smartphone mobile apps	Descriptive, quantitative	247 patients	Portal users had higher levels of education, lower rates of inadequate health literacy, and higher rates of using the internet and having an interest in websites or smartphone apps for tracking health. The odds of portal use increased with total eHEALS scores (health literacy scale) and decreased among Black patients.
Riippa, 2014 [56]	Finland	No specific portal	Interventional, other than RCT	876 patients	Patients with a severe diagnosis during the intervention showed the greatest positive change in patient activation (mean change 5.4, SD 8.4). Patients diagnosed 1-2 years ago (mean change 2.3, SD 15.7) and patients with no severe diagnoses (mean change 1.6, SD 13.1) showed a positive change in patient activation.
Ronda, 2014 [26]	Netherlands	Digitaal Logboek	Descriptive, quantitative	1390 patients	Multivariable analysis showed that increasing age and smoking were associated with not using the portal. A higher educational level, treatment by an internist, using insulin, polypharmacy, better diabetes knowledge, and more hyperglycemic episodes were less likely to be associated with not using the portal.
Smith, 2015 [27]	United States	EpicCare	Descriptive, quantitative	534 patients	Significant predictors of registering were as follows: gender (male 65.3% vs female 55.1%), race (White 71.7% vs African American 27.7% vs “other” races 41.7%), education (more educated people were more likely to register), number of chronic conditions (70.9% with zero conditions, 63.2% with one condition, and 50.0% with two or more conditions), health literacy (adequate 72.7% vs marginal 46.4% vs limited health literacy 21.7%).
Sun, 2019 [41]	United States	Epic’s personal health record system	Observational, hypothesis testing	38,399 patients	Almost one-third of patients (n=12,615; 32.9%, 95% CI 32.38%-33.32%) had used the portal for a mean of 2.5 (SD 1.9) years prior to the study period. Portal use was higher on weekdays (P<.001). An increase in portal use was observed in response to email reminders. A nonlinear relationship between age and portal use was observed and depended on several other predictors (P<.05). Patients living in more rural areas with low income were at lower odds to use the portal (P=.02), and this finding also applied to non‐Whites with low income (P<.001). More chronic conditions and a higher initial HbA_{1c} value were associated with portal use (P=.01).
Tsai, 2019 [28]	United States	Epic’s personal health record system	Descriptive, quantitative	109,200 patients	Active portal users were on average older (49.45 vs 46.22 years) and frequently female (62.59% vs 54.91%). Both the differences in mean age (P=.008) and gender (P=.04) were significant. There was a bimodal peak in terms of active users, with active users more likely to be in their 30s and 60s. The difference among age groups was significant (P<.001). Differences in racial composition, insurance, and language were not significant.

First author, year	Country	Health information technology	Study design	Sample size	Main findings
Van der Vaart, 2014 [57]	Netherlands	Medisch Spectrum Twente	Interventional, other than RCT	360 patients	Univariate analyses showed that age, marital status, education level, employment, health literacy, and internet-related characteristics were significantly related to portal use. Nonusers were more often older, single, lower educated, and unemployed. Respondents with higher health literacy were more inclined to login on the portal, and respondents who used the internet more often had more years of experience and perceived their own skills as better.
Wade-Vuturo, 2013 [8]	United States	MyHealthAtVanderbilt patient portal	Descriptive, quantitative	54 patients	Participant age, gender, race, income, and education level were not associated with using SM[j] to send a message to a provider for any reason or using SM to schedule an appointment.
Wallace, 2016 [58]	United States	MyChart by Epic health record system	Interventional, other than RCT	36,549 patients	Men, non-White patients, and Hispanic patients were significantly less likely to login once, 2 to 23 times, or 24 times than women, White patients, or non-Hispanic patients. Patients with public insurance were less likely to login than those with private insurance across all MyChart usage categories. Patients with income levels 100% of the FPL[k] were more likely to login one time than those below the FPL level.
Wedd, 2019 [30]	United States	Unspecified patient portal	Descriptive, quantitative	710 patients	Black patients were less likely to use the portal vs White patients among both kidney (Black 57% vs White 74%) and liver (Black 28% vs White 55%) transplant recipients. In adjusted multivariable analyses, kidney transplant recipients were more likely to use the portal if they had higher education. Among liver recipients, patients who were White and had higher education were more likely to use the portal.
Zhong, 2018 [42]	United States	MyUFHealth (also known as MyChart by Epic)	Observational, hypothesis testing	15,659 nonusers and 5494 users	The user group comprised 53.1% patients with more than four chronic problems (vs 40.2% of the matched nonuser group), and had more patients bearing 10 or more chronic problems (18.2% vs 12.2%). Individuals enrolled in the patient portal were mostly middle aged (31-64 years) and female. Married patients were more likely to adopt the portal. Medicare and Medicaid patients, and Black or African American patients were less likely to be adopters. Portal adoption was also associated with the baseline number of active medical problems (P<.05).

[a]OR: odds ratio.

[b]RCT: randomized controlled trial.

[c]EHR: electronic health record.

[d]PHR: personal health record.

[e]LDL: low-density lipoprotein.

[f]SBP: systolic blood pressure.

[g]FM: family medicine.

[h]GIM: general internal medicine.

[i]ASA: American Society of Anesthesiology.

[j]SM: secure messaging.

[k]FPL: federal poverty level.

In terms of age, results were not homogeneous. Higher use in older patients (aged >50 years) was described in four studies [18,20,21,33]. In a middle age and elderly cohort, the mean age of users was significantly lower [45]. On the other hand, younger patients were also the major users in disease/specialty-specific cohorts [32,35,46].

A total of 19 studies mentioned the origin and ethnicity of users [8,18,25,27,30-32,34,37-39,42-46,52,58]. White patients were usually the most likely to use the portals described in the different studies [18,27,30-32,37,38,43,45,52]. Two studies found no significant association between ethnicity and portal use [8,28].

Most of the articles found a positive association between female gender and portal use [28,35,38,42,43,58]. Only one study, conversely, found male gender to be a predictor of registering [27]. No statistically significant association between sex and user/nonuser status was found in two articles [36,45].

In a cohort of patients with mental disorders, having schizophrenia or schizoaffective disorders was negatively associated with portal use [31]. Instead, depression was positively associated [37]. Moderate or severe asthma was more linked to portal use [32,60]. Moreover, having a diagnosis of diabetes, hypertension, heart failure, or cardiovascular disease was a significant predictor of portal use [20,40], with one exception [46]. On the other hand, there were contrasting results concerning the association between the number of comorbidities and portal use [27,37,41,42,44].

People living in rural areas were less likely to use patient portals than urban citizens in three studies [22,31,41], while higher education levels were often related to broader use of portals [23,25-27,30,32,52,57]. Only one study showed no significant association [8]. Higher income was also generally associated with portal use [31,37,45,46,58]. Only one study found no association between income and the use of secure messages [8].

Studies conducted in the United States showed that having a private insurance was positively associated with portal use [18,32,39,42,45,46,58], with only one study reporting the absence of this association [28].

Other patient characteristics positively associated with portal use were being a member of a patient association [59] and being admitted to an academic medical center [44].

Attitudes and Satisfaction

Patient attitudes were evaluated in terms of perceived barriers and facilitators toward portal use. The overall satisfaction was also assessed, and it refers to the extent to which the patient is content about health care. Thirty articles addressed these topics (Multimedia Appendix 4).

Nine articles clearly addressed the barriers to portal use. Some of the main issues were related to perceived or preconceived security concerns [8,48], limited knowledge [26,51], satisfaction with current care [51], paying for the service [54], disinterest in managing one's own disease [26,57], personal/time constraints and not thinking about accessing the portal [57,61], doubts about the reliability of the patient portal to facilitate a timely and productive message exchange with providers [8], and prior negative experiences with secure messaging [8]. Other barriers were related to population characteristics, such as being a clinician older than 55 years or younger than 35 years and being male [33], and variation in provider availability for online appointment scheduling and response times to medical messages [25]. Limited computer and internet access [20,25,26,51,57], knowledge of technology [25,26,57], security concerns [25,48], and data integration [48] were also negative predictors of portal use.

On the other hand, nine articles specified the elements that facilitated portal use, including improved communication with specialists [22,48,60]; availability of information that led to an increased awareness of the health status [19,22,32,51,60] and tracking of disease control [32]; time-saving, convenient, and easy to use elements [19]; accuracy, timeliness, usefulness, and convenience of the functionalities included in the portal [22,50]; availability of surrogates (ie, daughter or family member) to act as intermediaries [51]; active involvement of the practice/staff in the promotion of the portal (ie, team approach strategy to engage staff in notifying patients) [33]; and active training of the patient for portal use [61]. Concerning satisfaction, patients were generally satisfied with the portals [8,17,21,29,49,50,59-61].

Discussion

Due to the considerable amount of literature published on the topic of patient portals, the aim of this review was to provide evidence and to gather information systematically. Similar to the review published in 2013 by Goldzweig et al [4], the outcomes were grouped into four aspects, namely, health outcomes and adherence, health care efficiency, patient characteristics, and attitudes and satisfaction, and showed nonunique results in terms of benefits brought by patient portals concerning patient experience and health. The functionalities available in the portals described by Goldzweig et al and our review were likewise the same.

Positive results were described relating to the enhancement of preventive behaviors [45], changes in chronic condition control with higher control of diabetes parameters [34,36], and asthma flares [60]. However, conflicting results were described concerning blood pressure control [34,37], mental health conditions [35,54,55], and medication adherence [32,57]. A possible explanation of these results is that these particular studies involved patients (often with a low sample size) who utilized portals for a short period of time, preventing them from having a possible consistent outcome. Similarly, concerning health outcomes, Goldzweig et al found generally positive results, which documented improvement in patients' disease control and maintenance [4], even though many of the included studies in our review did not find a significant difference between portal users and nonusers. Adherence to therapy was always improved [4] similar to that in this review.

The number of clinical and ED visits [28,32,39,54,60] and hospital readmissions [39,44,54,60] did not always decline in patients using portals. The number of missed appointments decreased [35,42], but this finding was not always significant [28,42]. The nondeclining trends of hospitalizations and outpatient visits as a consequence of portal use are concordant in the two studies, although different findings were described. Indeed, different from our review, in which we found more often a decrease in the utilization of in-person services, the review by Goldzweig et al reported that most of the studies found an increased number of outpatient visits and hospitalizations. As patient portals are normally designed to reduce inappropriate health care utilization, this might be partially explained by the fact that provider and patient adaptations to the patient portals have evolved over time.

The studies differed also in terms of the population included, with conflicting results in portal adoption and age, gender,

ethnicity, kind of disease, and number of comorbidities. Generally, having a higher income [31,37,45,46,58] and having a private insurance were associated with increased portal use [18,32,39,42,45,46,58].

The main barriers faced in using the portals were (1) user-related issues, such as time constraints [57,61], disinterest in managing the disease [26,57], and limited digital knowledge [26,51]; (2) clinician-related issues, such as age [33] and attitude toward the portal [33]; and (3) technology-related issues, such as limited internet access [25,26,51,57] or ability to use technology [25,26,57] and security concerns [25,48]. Privacy and security concerns, and ability to use technological appliances and systems were the most important barriers to utilization described in the studies retrieved by Goldzweig et al [4]. Indeed, other recent reviews on patient attitudes highlighted that privacy and security problems are the main barriers to the use of patient portals. These barriers are as evident in elderly patients (in whom there is also a limitation of portal use related to age) as in younger patients [62,63]. Moreover, technical problems due to patient capacity and difficulties in using the portal also represent important barriers. Thus, correct and adequate information on safety issues and education on the technical use of the portal represent the best facilitators. Furthermore, engaging patients and making them realize that the portal represents a useful tool to support the management of their pathology (especially for chronic diseases) without replacing the doctor-patient relationship are important to encourage the use of digital portals [13,62,64]. In addition, technical improvements in the usability of portals could increase patient enrollment.

Security problems are complex issues that must be considered in any part of medical care. The use of information technologies in health care that can be accessed by multiple types of users (physician, patient, caregiver, and hospital administrative staff) represents the basis of the discussion about computer security [65]. Indeed, uncertainties about security of clinical data might hinder adoption of systems by both hospitals and patients [48]. Patients expressed concern about their privacy and the privacy of their family members, and asked for further information about confidentiality, as vulnerable data might be accessed by external providers, such as insurance providers, who are the main actors of health care access in many countries [47].

The facilitators retrieved were the prompt availability of health information that caused an increase in the awareness of the health status [19,22,32,51,60], improved communication with health care professionals [22,48,60], and the accuracy, timeliness, usefulness, and convenience of the functionalities included in the portal [22,50]. In this latter argument, some functionalities of the portal were found to be more useful than others, including laboratory tests and imaging [23,24,29,43,52], medical notes [23,26], messaging with providers [52], medication refill [52], and current medication list [29]. Generally and as observed by Goldzweig et al [4], patients declared being satisfied with the use of digital tools [8,17,21,29,49,50,59-61].

Despite the considerable number of studies included, the high heterogeneity in terms of outcomes and described portal functionalities did not allow us to perform meta-analyses and to draw generalizable and strong conclusions concerning the utility of the unique features of the portals.

Technological and digital innovations in health care could contribute to achieving the health system goals of equity, efficiency, accessibility, quality, and sustainability, if they are purposefully designed and cost-effectively implemented. When designing a new patient portal or a new functionality, developers and providers should always consider to which health care need they are trying to respond and if other nondigital interventions may be more effective or as effective at a lower cost.

Moreover, the adoption of a new technology is a complex process, depending on the content and the context in which it is introduced. As an example, our review demonstrated that it is feasible to achieve better medication adherence in chronic disease patients through portal use, and highlighted the main facilitators (eg, prompt availability of reliable information and accessibility of communication with disease specialists) and barriers (eg, security and usability concerns, and limited digital knowledge) to portal use. Keeping in mind these contextual factors could ease the difficult task of identifying the best digital tool for a specific population.

Before designing or implementing a new tool, it can be useful to analyze the ideal conditions needed for the adaptation, transfer, absorption, up-scaling, and enhancement of digital technologies. By ideal conditions, we basically mean a situation where the new technology has demonstrated effectiveness in trials or pilots, the provider is committed to guarantee continuous improvement in user accessibility and usability, and the main barriers in the target population are given due consideration. In the absence of these conditions, satisfactory results may be difficult to reach or may take many years to be observed.

The benefits of digitalization cannot be taken for granted and the use of technology does not always lead to an improvement in patient care and health system performance; thus, there is a need for evidence, which is, to date, scarce. The identification of a set of main features with proven efficacy for a patient portal is a useful starting point for the development and implementation of patient-oriented portals. Further studies should be conducted in different aspects of digitalization in health care. None of the studies retrieved analyzed the cost-effectiveness of portal use. Similarly, none of the studies compared the portals to each other, which could be interesting to point out the best practices and features.

Even though a patient portal is not a new concept, its real utilization and implementation are still far from optimal, and it seem to be still considered a “future technology.” It is important to adapt the portal functions to the needs and capacities of patients, in order to facilitate the use of this technology and improve its dissemination. In particular, overcoming ethnic and literacy barriers to portal use represents a fundamental goal to create more equitable, effective, and safe health care systems.

Authors' Contributions

All authors contributed to the study conception and design. Material preparation and data collection were performed by EC, LV, AMP, AG, and AB. EC and LV screened the articles, with any disagreement finally discussed and resolved with AMP. The first draft of the manuscript was written by EC, LV, AMP, AG, and AB. WR and SB commented on the latest version of the manuscript. WR and SB supervised the study. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search string.
[DOCX File , 12 KB - jmir_v23i9e26189_app1.docx]

Multimedia Appendix 2
Functions and details of patient portals.
[DOCX File , 25 KB - jmir_v23i9e26189_app2.docx]

Multimedia Appendix 3
Qualitative descriptions of the portals and/or features studied in the included articles.
[DOCX File , 30 KB - jmir_v23i9e26189_app3.docx]

Multimedia Appendix 4
Summary of the findings on patient attitudes and satisfaction.
[DOCX File , 27 KB - jmir_v23i9e26189_app4.docx]

References

1. Pagliari C, Detmer D, Singleton P. Potential of electronic personal health records. BMJ 2007 Aug 18;335(7615):330-333 [FREE Full text] [doi: 10.1136/bmj.39279.482963.AD] [Medline: 17703042]
2. Global Strategy on Digital Health 2020-2024. World Health Organization. 2019. URL: https://extranet.who.int/dataform/upload/surveys/183439/files/Draft%20Global%20Strategy%20on%20Digital%20Health.pdf [accessed 2020-11-07]
3. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient Portals Facilitating Engagement With Inpatient Electronic Medical Records: A Systematic Review. J Med Internet Res 2019 Apr 11;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
4. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
5. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. How outcomes are achieved through patient portals: a realist review. J Am Med Inform Assoc 2014 Jul;21(4):751-757 [FREE Full text] [doi: 10.1136/amiajnl-2013-002501] [Medline: 24503882]
6. Wong J, Steitz B, Rosenbloom S. Characterizing the impact of health literacy, computer ability, patient demographics, and portal usage on patient satisfaction with a patient portal. JAMIA Open 2019 Dec;2(4):456-464 [FREE Full text] [doi: 10.1093/jamiaopen/ooz058] [Medline: 32025642]
7. D'Costa SN, Kuhn IL, Fritz Z. A systematic review of patient access to medical records in the acute setting: practicalities, perspectives and ethical consequences. BMC Med Ethics 2020 Mar 02;21(1):18 [FREE Full text] [doi: 10.1186/s12910-020-0459-6] [Medline: 32122332]
8. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc 2013 May 01;20(3):519-525 [FREE Full text] [doi: 10.1136/amiajnl-2012-001253] [Medline: 23242764]
9. Green BB, Cook AJ, Ralston JD, Fishman PA, Catz SL, Carlson J, et al. Effectiveness of home blood pressure monitoring, Web communication, and pharmacist care on hypertension control: a randomized controlled trial. JAMA 2008 Jun 25;299(24):2857-2867 [FREE Full text] [doi: 10.1001/jama.299.24.2857] [Medline: 18577730]
10. Simon GE, Ralston JD, Savarino J, Pabiniak C, Wentzel C, Operskalski BH. Randomized trial of depression follow-up care by online messaging. J Gen Intern Med 2011 Jul;26(7):698-704 [FREE Full text] [doi: 10.1007/s11606-011-1679-8] [Medline: 21384219]
11. Tang PC, Overhage JM, Chan AS, Brown NL, Aghighi B, Entwistle MP, et al. Online disease management of diabetes: engaging and motivating patients online with enhanced resources-diabetes (EMPOWER-D), a randomized controlled trial.

J Am Med Inform Assoc 2013 May 01;20(3):526-534 [FREE Full text] [doi: 10.1136/amiajnl-2012-001263] [Medline: 23171659]
12. Cho AH, Arar NH, Edelman DE, Hartwell PH, Oddone EZ, Yancy WS. Do diabetic veterans use the Internet? Self-reported usage, skills, and interest in using My HealtheVet Web portal. Telemed J E Health 2010 Jun;16(5):595-602. [doi: 10.1089/tmj.2009.0164] [Medline: 20575727]
13. Amante DJ, Hogan TP, Pagoto SL, English TM. A systematic review of electronic portal usage among patients with diabetes. Diabetes Technol Ther 2014 Nov;16(11):784-793. [doi: 10.1089/dia.2014.0078] [Medline: 24999599]
14. Han H, Gleason K, Sun C, Miller H, Kang S, Chow S, et al. Using Patient Portals to Improve Patient Outcomes: Systematic Review. JMIR Hum Factors 2019 Dec 19;6(4):e15038 [FREE Full text] [doi: 10.2196/15038] [Medline: 31855187]
15. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
16. Abd-Alrazaq A, Bewick BM, Farragher T, Gardner P. Factors Affecting Patients' Use of Electronic Personal Health Records in England: Cross-Sectional Study. J Med Internet Res 2019 Jul 31;21(7):e12373 [FREE Full text] [doi: 10.2196/12373] [Medline: 31368442]
17. Byczkowski TL, Munafo JK, Britto MT. Family perceptions of the usability and value of chronic disease web-based patient portals. Health Informatics J 2014 Jun;20(2):151-162 [FREE Full text] [doi: 10.1177/1460458213489054] [Medline: 24056751]
18. Chan B, Lyles C, Kaplan C, Lam R, Karliner L. A Comparison of Electronic Patient-Portal Use Among Patients with Resident and Attending Primary Care Providers. J Gen Intern Med 2018 Dec;33(12):2085-2091 [FREE Full text] [doi: 10.1007/s11606-018-4637-x] [Medline: 30187376]
19. de Jong JM, Ogink PA, van Bunningen CG, Driessen RJ, Engelen LJ, Heeren B, et al. A Cloud-Based Virtual Outpatient Clinic for Patient-Centered Care: Proof-of-Concept Study. J Med Internet Res 2018 Sep 24;20(9):e10135 [FREE Full text] [doi: 10.2196/10135] [Medline: 30249584]
20. Gordon NP, Hornbrook MC. Differences in Access to and Preferences for Using Patient Portals and Other eHealth Technologies Based on Race, Ethnicity, and Age: A Database and Survey Study of Seniors in a Large Health Plan. J Med Internet Res 2016 Mar 04;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]
21. Jackson SL, DesRoches CM, Frosch DL, Peacock S, Oster NV, Elmore JG. Will use of patient portals help to educate and communicate with patients with diabetes? Patient Educ Couns 2018 May;101(5):956-959. [doi: 10.1016/j.pec.2017.11.004] [Medline: 29153758]
22. Laranjo L, Rodolfo I, Pereira AM, de Sá AB. Characteristics of Innovators Adopting a National Personal Health Record in Portugal: Cross-Sectional Study. JMIR Med Inform 2017 Oct 11;5(4):e37 [FREE Full text] [doi: 10.2196/medinform.7887] [Medline: 29021125]
23. Mishra VK, Hoyt RE, Wolver SE, Yoshihashi A, Banas C. Qualitative and Quantitative Analysis of Patients' Perceptions of the Patient Portal Experience with OpenNotes. Appl Clin Inform 2019 Jan;10(1):10-18 [FREE Full text] [doi: 10.1055/s-0038-1676588] [Medline: 30602196]
24. Moll J, Rexhepi H, Cajander Å, Grünloh C, Huvila I, Hägglund M, et al. Patients' Experiences of Accessing Their Electronic Health Records: National Patient Survey in Sweden. J Med Internet Res 2018 Nov 01;20(11):e278 [FREE Full text] [doi: 10.2196/jmir.9492] [Medline: 30389647]
25. Price-Haywood EG, Harden-Barrios J, Ulep R, Luo Q. eHealth Literacy: Patient Engagement in Identifying Strategies to Encourage Use of Patient Portals Among Older Adults. Popul Health Manag 2017 Dec;20(6):486-494. [doi: 10.1089/pop.2016.0164] [Medline: 28384076]
26. Ronda MCM, Dijkhorst-Oei L, Rutten GEHM. Reasons and barriers for using a patient portal: survey among patients with diabetes mellitus. J Med Internet Res 2014 Nov 25;16(11):e263 [FREE Full text] [doi: 10.2196/jmir.3457] [Medline: 25424228]
27. Smith SG, O'Conor R, Aitken W, Curtis LM, Wolf MS, Goel MS. Disparities in registration and use of an online patient portal among older adults: findings from the LitCog cohort. J Am Med Inform Assoc 2015 Jul 25;22(4):888-895 [FREE Full text] [doi: 10.1093/jamia/ocv025] [Medline: 25914099]
28. Tsai R, Bell EJ, Woo H, Baldwin K, Pfeffer MA. How Patients Use a Patient Portal: An Institutional Case Study of Demographics and Usage Patterns. Appl Clin Inform 2019 Jan;10(1):96-102 [FREE Full text] [doi: 10.1055/s-0038-1677528] [Medline: 30727003]
29. Turvey C, Klein D, Fix G, Hogan TP, Woods S, Simon SR, et al. Blue Button use by patients to access and share health record information using the Department of Veterans Affairs' online patient portal. J Am Med Inform Assoc 2014 Jul 01;21(4):657-663 [FREE Full text] [doi: 10.1136/amiajnl-2014-002723] [Medline: 24740865]
30. Wedd J, Basu M, Curtis LM, Smith K, Lo DJ, Serper M, et al. Racial, Ethnic, and Socioeconomic Disparities in Web-Based Patient Portal Usage Among Kidney and Liver Transplant Recipients: Cross-Sectional Study. J Med Internet Res 2019 Apr 22;21(4):e11864 [FREE Full text] [doi: 10.2196/11864] [Medline: 31008707]

31. Abel EA, Shimada SL, Wang K, Ramsey C, Skanderson M, Erdos J, et al. Dual Use of a Patient Portal and Clinical Video Telehealth by Veterans with Mental Health Diagnoses: Retrospective, Cross-Sectional Analysis. J Med Internet Res 2018 Nov 07;20(11):e11350 [FREE Full text] [doi: 10.2196/11350] [Medline: 30404771]
32. Fiks AG, DuRivage N, Mayne SL, Finch S, Ross ME, Giacomini K, et al. Adoption of a Portal for the Primary Care Management of Pediatric Asthma: A Mixed-Methods Implementation Study. J Med Internet Res 2016 Jun 29;18(6):e172 [FREE Full text] [doi: 10.2196/jmir.5610] [Medline: 27357835]
33. Krist AH, Woolf SH, Bello GA, Sabo RT, Longo DR, Kashiri P, et al. Engaging primary care patients to use a patient-centered personal health record. Ann Fam Med 2014;12(5):418-426 [FREE Full text] [doi: 10.1370/afm.1691] [Medline: 25354405]
34. Ancker J, Mauer E, Kalish R, Vest J, Gossey J. Early Adopters of Patient-Generated Health Data Upload in an Electronic Patient Portal. Appl Clin Inform 2019 Mar 10;10(2):254-260 [FREE Full text] [doi: 10.1055/s-0039-1683987] [Medline: 30970383]
35. Kipping S, Stuckey MI, Hernandez A, Nguyen T, Riahi S. A Web-Based Patient Portal for Mental Health Care: Benefits Evaluation. J Med Internet Res 2016 Nov 16;18(11):e294 [FREE Full text] [doi: 10.2196/jmir.6483] [Medline: 27852556]
36. Lau M, Campbell H, Tang T, Thompson DJS, Elliott T. Impact of patient use of an online patient portal on diabetes outcomes. Can J Diabetes 2014 Feb;38(1):17-21. [doi: 10.1016/j.jcjd.2013.10.005] [Medline: 24485208]
37. Manard W, Scherrer JF, Salas J, Schneider FD. Patient Portal Use and Blood Pressure Control in Newly Diagnosed Hypertension. J Am Board Fam Med 2016;29(4):452-459 [FREE Full text] [doi: 10.3122/jabfm.2016.04.160008] [Medline: 27390376]
38. North F, Crane SJ, Chaudhry R, Ebbert JO, Ytterberg K, Tulledge-Scheitel SM, et al. Impact of patient portal secure messages and electronic visits on adult primary care office visits. Telemed J E Health 2014 Mar;20(3):192-198 [FREE Full text] [doi: 10.1089/tmj.2013.0097] [Medline: 24350803]
39. Plate JF, Ryan SP, Bergen MA, Hong CS, Attarian DE, Seyler TM. Utilization of an Electronic Patient Portal Following Total Joint Arthroplasty Does Not Decrease Readmissions. J Arthroplasty 2019 Feb;34(2):211-214. [doi: 10.1016/j.arth.2018.11.002] [Medline: 30497899]
40. Powell KR, Deroche C. Predictors and patterns of portal use in patients with multiple chronic conditions. Chronic Illn 2020 Dec;16(4):275-283. [doi: 10.1177/1742395318803663] [Medline: 30284917]
41. Sun R, Burke LE, Saul MI, Korytkowski MT, Li D, Sereika SM. Use of a Patient Portal for Engaging Patients with Type 2 Diabetes: Patterns and Prediction. Diabetes Technol Ther 2019 Oct;21(10):546-556. [doi: 10.1089/dia.2019.0074] [Medline: 31335206]
42. Zhong X, Liang M, Sanchez R, Yu M, Budd PR, Sprague JL, et al. On the effect of electronic patient portal on primary care utilization and appointment adherence. BMC Med Inform Decis Mak 2018 Oct 16;18(1):84-12 [FREE Full text] [doi: 10.1186/s12911-018-0669-8] [Medline: 30326876]
43. Foster B, Krasowski MD. The Use of an Electronic Health Record Patient Portal to Access Diagnostic Test Results by Emergency Patients at an Academic Medical Center: Retrospective Study. J Med Internet Res 2019 Jun 28;21(6):e13791 [FREE Full text] [doi: 10.2196/13791] [Medline: 31254335]
44. Griffin A, Skinner A, Thornhill J, Weinberger M. Patient Portals: Who uses them? What features do they use? And do they reduce hospital readmissions? Appl Clin Inform 2016;7(2):489-501 [FREE Full text] [doi: 10.4338/ACI-2016-01-RA-0003] [Medline: 27437056]
45. Huang J, Chen Y, Landis JR, Mahoney KB. Difference Between Users and Nonusers of a Patient Portal in Health Behaviors and Outcomes: Retrospective Cohort Study. J Med Internet Res 2019 Oct 07;21(10):e13146 [FREE Full text] [doi: 10.2196/13146] [Medline: 31593546]
46. Jhamb M, Cavanaugh KL, Bian A, Chen G, Ikizler TA, Unruh ML, et al. Disparities in Electronic Health Record Patient Portal Use in Nephrology Clinics. Clin J Am Soc Nephrol 2015 Nov 06;10(11):2013-2022 [FREE Full text] [doi: 10.2215/CJN.01640215] [Medline: 26493242]
47. Bajracharya A, Crotty B, Kowoloff H, Safran C, Slack WV. Patient experience with family history tool: analysis of patients' experience sharing their family health history through patient-computer dialogue in a patient portal. J Am Med Inform Assoc 2019 Jul 01;26(7):603-609 [FREE Full text] [doi: 10.1093/jamia/ocz008] [Medline: 30946464]
48. Bidmead E, Marshall A. A case study of stakeholder perceptions of patient held records: the Patients Know Best (PKB) solution. Digit Health 2016 Sep 21;2:2055207616668431-2055207616668415 [FREE Full text] [doi: 10.1177/2055207616668431] [Medline: 29942567]
49. Jahn MA, Porter BW, Patel H, Zillich AJ, Simon SR, Russ AL. Usability Assessment of Secure Messaging for Clinical Document Sharing between Health Care Providers and Patients. Appl Clin Inform 2018 Apr;9(2):467-477 [FREE Full text] [doi: 10.1055/s-0038-1660521] [Medline: 29949815]
50. King G, Maxwell J, Karmali A, Hagens S, Pinto M, Williams L, et al. Connecting Families to Their Health Record and Care Team: The Use, Utility, and Impact of a Client/Family Health Portal at a Children's Rehabilitation Hospital. J Med Internet Res 2017 Apr 06;19(4):e97 [FREE Full text] [doi: 10.2196/jmir.6811] [Medline: 28385680]
51. Mishuris RG, Stewart M, Fix GM, Marcello T, McInnes DK, Hogan TP, et al. Barriers to patient portal access among veterans receiving home-based primary care: a qualitative study. Health Expect 2015 Dec 12;18(6):2296-2305 [FREE Full text] [doi: 10.1111/hex.12199] [Medline: 24816246]

52. Portz JD, Bayliss EA, Bull S, Boxer RS, Bekelman DB, Gleason K, et al. Using the Technology Acceptance Model to Explore User Experience, Intent to Use, and Use Behavior of a Patient Portal Among Older Adults With Multiple Chronic Conditions: Descriptive Qualitative Study. J Med Internet Res 2019 Apr 08;21(4):e11604 [FREE Full text] [doi: 10.2196/11604] [Medline: 30958272]
53. Schultz CL, Alderfer MA. Are on-line patient portals meeting test result preferences of caregivers of children with cancer? A qualitative exploration. Pediatr Blood Cancer 2018 Nov 14;65(11):e27306. [doi: 10.1002/pbc.27306] [Medline: 30007016]
54. Quanbeck A, Gustafson DH, Marsch LA, Chih M, Kornfield R, McTavish F, et al. Implementing a Mobile Health System to Integrate the Treatment of Addiction Into Primary Care: A Hybrid Implementation-Effectiveness Study. J Med Internet Res 2018 Jan 30;20(1):e37 [FREE Full text] [doi: 10.2196/jmir.8928] [Medline: 29382624]
55. Riippa I, Linna M, Rönkkö I. A Patient Portal With Electronic Messaging: Controlled Before-and-After Study. J Med Internet Res 2015 Nov 09;17(11):e250 [FREE Full text] [doi: 10.2196/jmir.4487] [Medline: 26553595]
56. Riippa I, Linna M, Rönkkö I. The effect of a patient portal with electronic messaging on patient activation among chronically ill patients: controlled before-and-after study. J Med Internet Res 2014 Nov 19;16(11):e257 [FREE Full text] [doi: 10.2196/jmir.3462] [Medline: 25413368]
57. van der Vaart R, Drossaert CHC, Taal E, Drossaers-Bakker KW, Vonkeman HE, van de Laar MAFJ. Impact of patient-accessible electronic medical records in rheumatology: use, satisfaction and effects on empowerment among patients. BMC Musculoskelet Disord 2014 Mar 26;15:102 [FREE Full text] [doi: 10.1186/1471-2474-15-102] [Medline: 24673997]
58. Wallace LS, Angier H, Huguet N, Gaudino JA, Krist A, Dearing M, et al. Patterns of Electronic Portal Use among Vulnerable Patients in a Nationwide Practice-based Research Network: From the OCHIN Practice-based Research Network (PBRN). J Am Board Fam Med 2016 Oct;29(5):592-603 [FREE Full text] [doi: 10.3122/jabfm.2016.05.160046] [Medline: 27613792]
59. Gossec L, Cantagrel A, Soubrier M, Berthelot J, Joubert J, Combe B, et al. An e-health interactive self-assessment website (Sanoia) in rheumatoid arthritis. A 12-month randomized controlled trial in 320 patients. Joint Bone Spine 2018 Dec;85(6):709-714. [doi: 10.1016/j.jbspin.2017.11.015] [Medline: 29246532]
60. Fiks AG, Mayne SL, Karavite DJ, Suh A, O'Hara R, Localio AR, et al. Parent-reported outcomes of a shared decision-making portal in asthma: a practice-based RCT. Pediatrics 2015 Apr;135(4):e965-e973 [FREE Full text] [doi: 10.1542/peds.2014-3167] [Medline: 25755233]
61. Stein JN, Klein JW, Payne TH, Jackson SL, Peacock S, Oster NV, et al. Communicating with Vulnerable Patient Populations: A Randomized Intervention to Teach Inpatients to Use the Electronic Patient Portal. Appl Clin Inform 2018 Oct;9(4):875-883 [FREE Full text] [doi: 10.1055/s-0038-1676333] [Medline: 30541152]
62. Sakaguchi-Tang DK, Bosold AL, Choi YK, Turner AM. Patient Portal Use and Experience Among Older Adults: Systematic Review. JMIR Med Inform 2017 Oct 16;5(4):e38 [FREE Full text] [doi: 10.2196/medinform.8092] [Medline: 29038093]
63. Zhao JY, Song B, Anand E, Schwartz D, Panesar M, Jackson GP, et al. Barriers, Facilitators, and Solutions to Optimal Patient Portal and Personal Health Record Use: A Systematic Review of the Literature. AMIA Annu Symp Proc 2017;2017:1913-1922 [FREE Full text] [Medline: 29854263]
64. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015 Feb 20;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
65. Giansanti D. Cybersecurity and the : The Challenge of This Millennium. Healthcare (Basel) 2021 Jan 11;9(1):62 [FREE Full text] [doi: 10.3390/healthcare9010062] [Medline: 33440612]

Abbreviations

ED: emergency department
EMR: electronic medical record
ER: emergency room
RCT: randomized controlled trial

Edited by R Kukafka; submitted 03.12.20; peer-reviewed by S Pillon, N Gordon, I Wilson, F Kaliyadan; comments to author 10.01.21; revised version received 24.01.21; accepted 24.05.21; published 08.09.21.

Please cite as:
Carini E, Villani L, Pezzullo AM, Gentili A, Barbara A, Ricciardi W, Boccia S
The Impact of Digital Patient Portals on Health Outcomes, System Efficiency, and Patient Attitudes: Updated Systematic Literature Review
J Med Internet Res 2021;23(9):e26189
URL: https://www.jmir.org/2021/9/e26189
doi:10.2196/26189
PMID:34494966

Review

Patient Portals to Support Palliative and End-of-Life Care: Scoping Review

M Pilar Ingle[1*], BA, MSW; Cristina Valdovinos[2*], MPH; Kelsey L Ford[3*], DrPH; Shou Zhou[3*], PhD; Sheana Bull[3*], PhD; Starlynne Gornail[3*], MS; Xuhong Zhang[3*], PhD; Susan Moore[3], PhD; Jennifer Portz[4*], MSW, PhD

[1]Graduate School of Social Work, University of Denver, Denver, CO, United States

[2]David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CO, United States

[3]Colorado School of Public Health, University of Colorado, Aurora, CO, United States

[4]Department of General Internal Medicine, University of Colorado, Aurora, CO, United States

[*]these authors contributed equally

Corresponding Author:
M Pilar Ingle, BA, MSW
Graduate School of Social Work
University of Denver
2148 S High St
Denver, CO, 80210
United States
Phone: 1 720 515 8595
Email: Pilar.ingle@du.edu

Abstract

Background: Although patient portals are widely used for health promotion, little is known about the use of palliative care and end-of-life (PCEOL) portal tools available for patients and caregivers.

Objective: This study aims to identify and assess the user perspectives of PCEOL portal tools available to patients and caregivers described and evaluated in the literature.

Methods: We performed a scoping review of the academic literature directed by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) extension for Scoping Review and searched three databases. Sources were included if they reported the development or testing of a feature, resource, tool, or intervention; focused on at least one PCEOL domain defined by the National Coalition for Hospice and Palliative Care; targeted adults with serious illness or caregivers; and were offered via a patient portal tethered to an electronic medical record. We independently screened the titles and abstracts (n=796) for eligibility. Full-text (84/796, 10.6%) sources were reviewed. We abstracted descriptions of the portal tool name, content, targeted population, and reported user acceptability for each tool from included sources (n=19).

Results: In total, 19 articles describing 12 tools were included, addressing the following PCEOL domains: ethical or legal (n=5), physical (n=5), and psychological or psychiatric (n=2). No tools for bereavement or hospice care were identified. Studies have reported high acceptability of tools among users; however, few sources commented on usability among older adults.

Conclusions: PCEOL patient portal tools are understudied. As medical care increasingly moves toward virtual platforms, future research should investigate the usability and acceptability of PCEOL patient portal resources and evaluate their impact on health outcomes.

(J Med Internet Res 2021;23(9):e28797) doi:10.2196/28797

KEYWORDS

patient portal; electronic health record; digital health; palliative care; end-of-life care

Introduction

Patient portals are secure websites that provide access to personal health information and health care services that often include web-based tools for medical visits, health records, and medications [1]. In addition to providing patients with access to their health records, health care organizations offer other digital health resources and functions through patient portals, including messaging with providers, general medical information, and prescription refills [2]. The growing adoption

of portals shows that approximately 90% of health care organizations provide a portal system [3]. Through access to health records, patient portals promote self-management of health and disease and help improve patient-clinician communication [4-7].

Patients with chronic or serious illness may especially benefit from access to patient portals; in fact, having a chronic illness is a predictor of portal enrollment and use [8]. Older adults—80% of whom have at least one chronic illness [9]—are increasingly becoming the focus of chronic condition management and population health initiatives that involve the adoption and use of digital health technology, including patient portals. Although there are concerns about the low use of patient portals among this population [10-13], evidence also suggests a growing trend of internet and technology use among older adults [14,15]. This emphasizes the need to develop and tailor patient portals for the use of older adults. The rise of patient portal use also presents the opportunity to maximize patient empowerment, education, and patient-clinician communication among patients facing serious illnesses.

In light of the potential benefits of digital health technologies for older adults and individuals with chronic or serious illness, interest is also growing in digital health interventions targeted toward palliative care and end-of-life (PCEOL) patients, including the use of patient portals [16]. Palliative care focuses on pain and symptom management and improving quality of life for patients with serious illness and their families, and the Institute of Medicine has named access to these services as essential [17,18]. Despite the potential benefits of palliative care [19], issues with access to this specialized care persist because of barriers, including lack of availability and public awareness, particularly in rural areas [20,21]. Patient portals can be used to deliver advance care planning (ACP) education and provide access to materials for documenting advance directives, Medical Durable Power of Attorney (MDPOA), and health care proxies. Studies implementing ACP efforts through patient portals demonstrate patient acceptability as well as successful increases in ACP-related documentation [22-24]. Although this evidence is encouraging for the use of patient portals to promote PCEOL outcomes, there is also a need to understand the utility of patient portals for PCEOL patient education and resources and whether patient portal use is associated with other important PCEOL health outcomes, such as symptom management, hospice use (high quality palliative care near the end of life, typically in the last 6 months) [25], and documentation of care preferences.

As a first step in this line of research, we aim to conduct a scoping review of PCEOL patient portal research. The aims of this study are to (1) identify PCEOL patient portal tools available for patients and caregivers that are described and evaluated in the literature and (2) document patient and caregiver perspectives regarding these tools. By describing currently available tools and their respective user perspectives, this review intends to inform future research examining the association between tool use and PCEOL outcomes while guiding future PCEOL patient portal tool development.

Methods

Scoping Review

As little is known about the use of patient portals for the delivery of palliative and end-of-life care, a scoping review of the academic literature was conducted to identify sources describing currently available patient portal PCEOL tools. Our methods were aligned with methodological framework for scoping reviews by Arksey and O'Malley [26], which includes the following stages: identifying the research questions; identifying relevant studies; study selection; charting the data; and collating, summarizing, and reporting the results. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) extension for Scoping Review checklist guided our reporting (Multimedia Appendix 1) [27]. Our scoping review goals and procedures were registered with the Open Science Framework on March 26, 2020 (DOI 10.17605/OSF.IO/N34JZ).

Identifying Relevant Studies

To identify PCEOL patient portal tools described in the academic literature, we searched three databases (Ovid MEDLINE, CINAHL, and Web of Science) for peer-reviewed sources, including both qualitative and quantitative studies. The search terms included *patient portal* in combination with PCEOL terms such as *palliative care*, *hospice*, *end-of-life*, *terminal illness*, *cancer*, *advance directives*, *symptom management*, *ACP*, *grief*, and *caregiver support*. Our specific search strategy with exclusion terms is presented in Multimedia Appendix 2. The PCEOL search terms have previously been used to identify digital health solutions directed toward palliative care [28,29]. No time limits restricted the search, and only sources in English were included. We used Covidence, a web-based systematic review management program to remove duplicates, track citations, and screen references [30]. Our search was conducted in March 2020.

Sources were included in the scoping review if they (1) described a developed feature, resource, tool, or intervention that focused on at least one domain of PCEOL as defined by the National Coalition for Hospice and Palliative Care [31]; (2) targeted adults with serious illness (a health condition with a high risk of mortality impacting daily functioning or quality of life) [32] or their family and caregivers; and (3) were offered via a patient portal. The domains of palliative care include the following:

- Physical aspects of care including pain and symptom management, as well as quality of life assessment
- Psychological aspects of care, including care about anxiety, depression, stress, cognitive impairment, psychosocial support, coping skills, and support around patient or family grief and bereavement
- Social aspects of care, including patient and family education and caregiver support
- Spiritual and cultural aspects of care, including spiritual, existential, or religious support or attention to language, ritual, or dietary needs
- Care of the imminently dying, including end-of-life care and education

- Ethical and legal aspects of care, including ACP and documentation (appointment of MDPOA or health care proxy, and completion of advance directives)

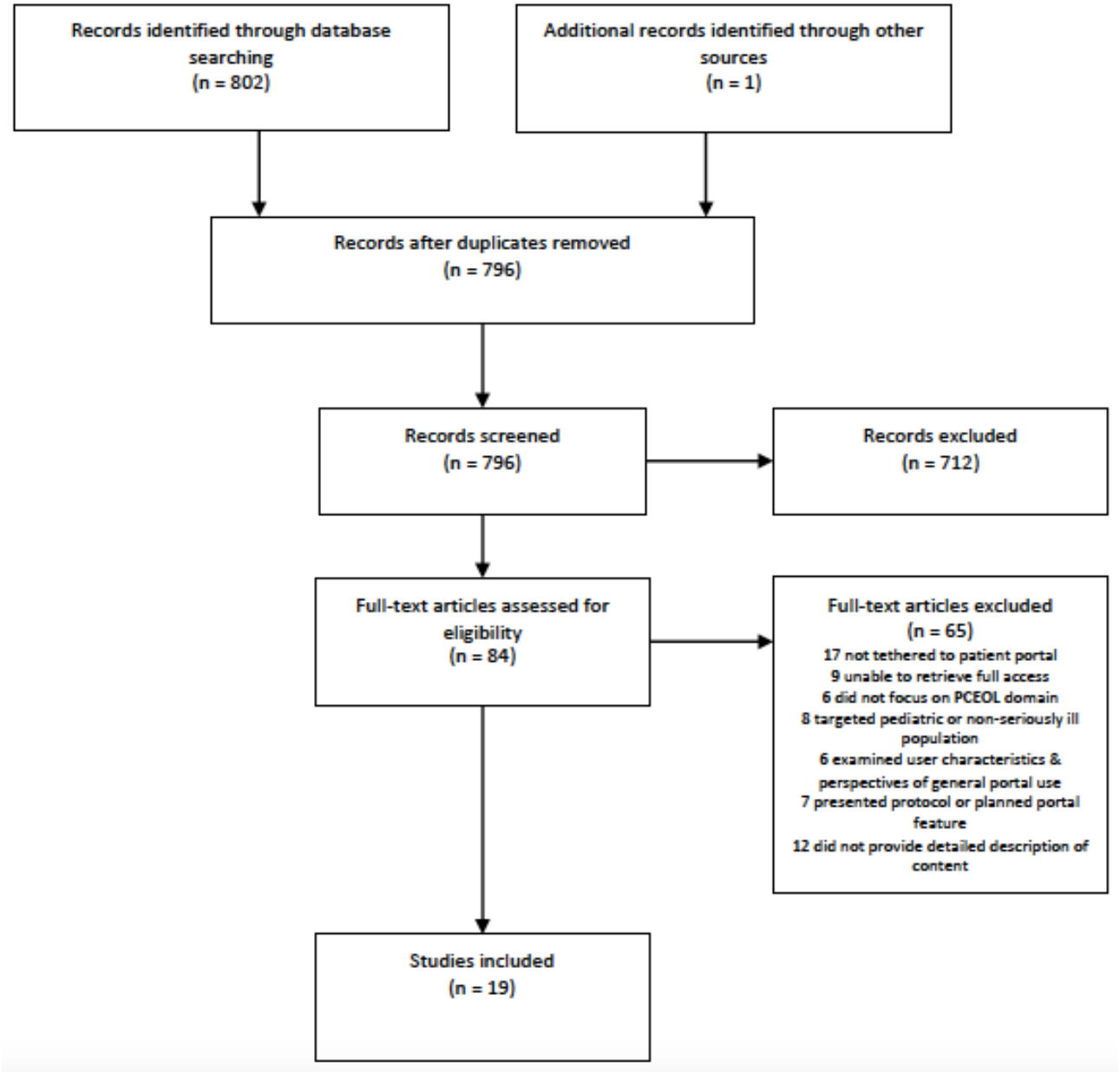
To determine whether sources met the inclusion criteria, 2 coders (MPI and CV) independently reviewed the academic sources by title and abstract.

Full-Text Review

Full texts of sources included by title and abstract were downloaded and further assessed for eligibility by 2 coders (MPI and CV). We determined whether the portal tool was targeted toward patients, caregivers, or both by reviewing the full text of the included sources for explicit mention of the intended users of the portal tool. The full text was also reviewed for a thorough description of the tool's purpose and available features, apps, and content. Sources that presented a protocol or planned portal feature, did not provide a detailed description of the resources or content (eg, abstract and review articles), did not focus on a PCEOL domain, targeted a pediatric or non–seriously ill population, examined the user characteristics and perspectives of general portal use, were not published in English, described a tool that was not tethered to a patient portal (eg, free-standing website or mobile app), or for which the full text was not retrievable, were excluded from the review. Pediatric populations were excluded because of their different care needs, access to technology, and technological skills.

The coders regularly reviewed the sources and discussed questions and possible disputes. Both coders agreed on the features included in the review, with a third reviewer (JDP) reconciling disputes. A summary of the search and screening processes is shown in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. PCEOL: palliative care and end-of-life.



Charting the Data

The full text of each of the included sources was reviewed, and the elements and features of the described PCEOL patient portal tool were abstracted by 2 reviewers (MPI and CV) into a Microsoft Excel spreadsheet. To ensure consistency, the reviewers double-extracted five articles before dividing the remaining articles. The abstracted data included the following components: portal name, health system user, developer (research team or organization), target illness and/or palliative care element, intended audience (patient, patient's caregiver, or both), and a summary of the features. When described, patient and caregiver perspectives about the tool, including measurements and qualitative reports of acceptability or usability and user satisfaction, were abstracted.

Collating, Summarizing, and Reporting the Results

We used an iterative and mixed inductive-deductive approach to summarize the charted information and results from the included papers to address our aims [26]. The charted information from each paper was reviewed extensively to identify similarities and differences in patient portal features, PCEOL domains, study population or targeted audience, and acceptability and usability findings (when available). The first, second, and senior authors (MPI, CV, and JDP) reviewed and agreed upon the thematic groupings of the results.

Results

Study Selection

The initial database search identified 796 unique sources that we screened by title and abstract (Figure 1). At this stage, we excluded 89.4% (712/796) of sources and reviewed 10.6% (84/796) of full-text sources. We excluded an additional 65 sources after full-text review because they presented a protocol or planned portal feature (7/65, 11%), examined user characteristics and perspectives of general portal use (6/65, 9%), described a tool that was not tethered to a patient portal (17/65, 26%), did not focus on a specific PCEOL domain (6/65, 9%), or was targeted at a pediatric or non–seriously ill population (8/65, 12%). At this stage, we also excluded sources that did not provide detailed descriptions of the tool's intended audience, content, and/or features (12/65, 18%) and sources for which the full text was not retrieved through the 3 separate university libraries we had access to (9/65, 13%). Using these methods, we identified 18 sources eligible for inclusion. During the full-text review of a selected article [33], an additional article that met the inclusion criteria [34] was identified in the text, bringing the total number of included sources to 19.

Identification of PCEOL Patient Portal Tools

We abstracted data from 19 sources published between the years 2013 and 2020 that describe 12 unique PCEOL patient portal tools (Table 1). The identified tools were developed by researchers and clinicians with the specific intent of enhancing PCEOL care. The identified PCEOL patient portal tools address the following domains: ethical and legal (n=5), physical (n=5), and psychological aspects of care (n=2). No tools address spiritual or bereavement support or hospice care education. Tools addressing the physical and psychological aspects of palliative care focus on quality of life and psychosocial support tools and are designed for cancer (n=6) and chronic kidney disease (n=1) patients. Patients are the intended audience for most PCEOL patient portal tools (n=9). Most tools were developed in an academic setting (n=11) and in the United States (n=9).

Table 1. Summary of included articles.

Study	Portal name	PCEOL[a] domain	Sample size and characteristics	Study design	Developer	Summary of features	Summary of portal use	Promotion of use
Bajracharya et al [24]	PatientSite	Ethical or legal (ACP[b])	200 patients; 63% women; 82% White; mean age 55 years (SD 15.16)	Evaluation of portal implementation	Beth Israel Deaconess Medical Center, Boston, MA, United States	A guided HCP[c] interview delivered via patient portal that provides education on the HCP role, collects information necessary for the completion of an HCP form, allows patients to print the HCP form and submit their HCP form electronically. The HCP form may be incorporated into patient's personal health record.	Of the participants that did not have an HCP listed before partaking in the interview, 78% submitted HCP information for clinician review. Of 200 patients, 139 submitted updated or new information about HCP overall.	N/A[d]
Tieu et al [35]	Mayo Clinic Patient Online Services	Ethical or legal (ACP)	2526 patients; 51% female; mean age 72 years (SD 5.8)	Randomized controlled intervention	Mayo Clinic, Rochester, MN, United States	Tool delivers a personalized, electronic message to patients with a link to ACP education and a state-specific advance directive form.	Only 5.5% of the intervention group and 2% of the control group completed and returned an AD[e].	Patients randomized to intervention group received message encouraging them to complete AD; reminder was sent out 8 weeks after if patients had not completed.
Bose-Brill et al [34]	MyChart (operated through Epic)	Ethical or legal (ACP)	50 patients	Randomized controlled pilot intervention	The Ohio State University, Columbus, OH, United States	Tool delivers a secure message to patients consisting of an introduction to ACP and four ACP-related questions. The responses are automatically stored in the patient's medical record and integrated into an ACP previsit planning algorithm focused on enhancing patient-provider communication surrounding ACP preferences.	In the intervention group, 10 of 23 participants completed ACP documentation compared with 1 of 25 participants in the control group.	N/A
Bose-Brill et al [33]	MyChart (operated through Epic)	Ethical or legal (ACP)	419 patients; 65% female; median age 61 years; range 50-93	Pragmatic trial	The Ohio State University, Columbus, OH, United States	Tool delivers a secure message to patients consisting of an introduction to ACP and four ACP-related questions. The responses are automatically stored in the patient's medical record and integrated into an ACP previsit planning algorithm focused on enhancing patient-provider communication surrounding ACP preferences.	ACP documentation in EHR[f] increased by 27% in intervention group, compared with 0.7% in control group. 78% of the intervention group read the initial message sent to them and 19% responded to at least one ACP-related question.	N/A

Study	Portal name	PCEOL[a] domain	Sample size and characteristics	Study design	Developer	Summary of features	Summary of portal use	Promotion of use
Lum et al [22]	MyHealth Connection (operated through Epic)	Ethical or legal (ACP)	2184 patients; 69% female; mean age 45 years; range 17-98	Quality improvement intervention	University of Colorado Hospital, Aurora, CO, United States	Tool provides patient-centered website for ACP education, secure messaging with an ACP support team, the ability to complete and sign an MD-POA[g] form and to view the completed advance directive in the patient's personal health record.	Over 15 months, 2814 patients used ACP tool; 89% completed MDPOA form, 2% called or sent web-based messages; 8% viewed MDPOA form without completing.	N/A
Jordan et al [23]	MyHealth Connection (operated through Epic)	Ethical or legal (ACP)	46 patients; 63% female; mean age 49 years	Exploratory qualitative study	University of Colorado Hospital, Aurora, CO, United States	Tool provides patient-centered website for ACP education, secure messaging with an ACP support team, the ability to complete and sign an MD-POA form and to view the completed advance directive in the patient's personal health record.	Not collected	N/A
Brungardt et al [36]	MyHealth Connection (operated through Epic)	Ethical or legal (ACP)	105 older adult patients	Practice-based pilot initiative	University of Colorado Hospital, Aurora, CO, United States	Tool provides patient-centered website for ACP education, secure messaging with an ACP support team, the ability to complete and sign an MD-POA form and to view the completed advance directive in the patient's personal health record.	At 1 year, 63 patients read ACP message and 17 had taken at least one ACP action step.	N/A
Portz et al [37]	My Health Manager (operated through Epic)	Ethical or legal (ACP)	24 older adult patients with multiple chronic conditions; 71% female; 79% White; mean age 78 years (SD 5.4)	Qualitative case study	Kaiser Permanente, Denver, CO, United States	Tool provides ACP education, ACP resources, and links to external website to complete advance directives.	Not collected	N/A
Dalal et al [38]	Patient-Centered Toolkit	Physical (QOL[h])	119 admitted patients and 120 caregivers; 43% female; mean age 56 years	Evaluation of portal implementation	Brigham and Women's Hospital, Boston, MA, United States	Tool allows patients with serious illness to navigate their plan of care during hospitalization. Patient can establish a single recovery goal and rate priorities, review medication and test results, securely message their care team, access educational content, and view discharge checklists, tailored safety tips, and reminders.	66% users inputted daily goal; 41% inputted overall goal (to be cured, to live longer, to be comfortable, other); 32% communicated care preferences; 64% provided real-time feedback for team. Goals, results, team members, medications, and messages were the most frequently visited pages. Problem education was the least frequently visited page.	N/A

Study	Portal name	PCEOL[a] domain	Sample size and characteristics	Study design	Developer	Summary of features	Summary of portal use	Promotion of use
Dalal et al [39]	Patient-Centered Toolkit	Physical (QOL)	55 patients preintervention, 40% female, 89% White; mean age 59 years (SD 12.8); 46 patients postintervention, 46% female, 83% White, mean age 58 years (SD 13.5)	Prospective pre- or postintervention study	Brigham and Women's Hospital, Boston, MA, United States	Tool allows patients with serious illness to navigate their plan of care during hospitalization. Patient can establish a single recovery goal and rate priorities, review medication and test results, securely message their care team, access educational content, and view discharge checklists, tailored safety tips, and reminders.	Not collected	Participants were encouraged to enter recovery goals to the portal.
Kuijpers et al [40]	MyAVL or MijnAVL	Physical (QOL)	92 female patients with breast cancer; mean age 49 years (SD 11.4)	Pre- or posttest intervention	Netherlands Cancer Institute, Amsterdam, Netherlands	PRO[i] collection and symptom management for patients with breast or non–small cell lung cancer. Patients can complete quality of life and physical activity questionnaires and receive personalized physical activity advice based on the questionnaire responses.	Overview of appointments and EMR[j] were accessed most frequently; mean number of log-ins for on-treatment is 10.9 with mean of 11.3 minutes. Mean number of log-ins for off-treatment is 5.6 with mean of 15.2 minutes.	N/A
Groen et al [41]	MyAVL (MijnAVL)	Physical (QOL)	37 patients with lung cancer; 47% women; 100% White; mean age 59 years (SD 8.4); range 40-76 years	Feasibility intervention	Netherlands Cancer Institute, Amsterdam, Netherlands	PRO collection and symptom management for patients with breast or non–small cell lung cancer. Patients can complete quality of life and physical activity questionnaires and receive personalized physical activity advice based on the questionnaire responses.	Mean number of log-ins over 4-month study period—11.2 with mean duration of 12.9 minutes. Overview of appointments, access to EMR, and questionnaires were the most accessed.	N/A
Wagner et al [42]	MyChart (operated through Epic)	Physical (QOL)	636 female patients with cancer; 78% White; age mean 55 years (SD 12.8); range 21-90 years	Clinical quality improvement initiative	PROMIS[k], Northwestern University (Chicago, IL) and US Department of Health and Human Services (Washington, DC), United States	PRO collection and symptom management for patients with cancer. Delivers electronic PRO assessments and other assessments to identify psychosocial concerns, informational and nutritional needs. The results are immediately populated to the patient's EHR. This integration allows automated triage to multidisciplinary care team.	80% of patients read the initial MyChart message asking for e-PRO[l] assessment completion; about 33% completed the entire assessment. 90% of the patients that completed the assessment did so at home.	e-PRO assessments were sent before scheduled outpatient appointments through MyChart. Those who did not compete the assessment at home were provided with an iPad to complete the assessment during appointment check-in.

Study	Portal name	PCEOL[a] domain	Sample size and characteristics	Study design	Developer	Summary of features	Summary of portal use	Promotion of use
Garcia et al [43]	MyChart (operated through Epic)	Physical (QOL)	3521 patients with cancer; 68% female; 76% White; mean age 57 years (SD 13.39)	Clinical quality improvement initiative	PROMIS, Northwestern University (Chicago, IL) and US Department of Health and Human Services (Washington, DC), United States	PRO collection and symptom management for patients with cancer. Delivers electronic PRO assessments and other assessments to identify psychosocial concerns, informational and nutritional needs. The results are immediately populated to the patient's EHR. This integration allows automated triage to multidisciplinary care team.	Not collected	e-PRO assessments were sent before scheduled outpatient appointments through MyChart. Those who did not compete the assessment at home were provided with an iPad to complete the assessment during appointment check-in.
Brant et al [44]	Carevive Care Planning System	Physical (QOL)	121 female patients with gynecologic or breast cancer; 83% White; age mean 56 years (SD 10.94); range 28-81	Mixed methods pilot study	Carevive, United States	PRO collection and symptoms management for patients with breast or gynecologic cancer. Collects PRO and integrates responses and clinical data to create a tailored care plan that provides clinical decision support and self-management advice to patient and caregivers.	Not collected	N/A
Hazara et al [45]	Renal PatientView	Physical (QOL)	190 patients with chronic kidney disease; 34% female; median age 53 years; range 19-86	Quality improvement initiative	Renal Patient Exchange Group, United Kingdom	PRO collection and symptom management for patients with chronic kidney disease. Patients can document and monitor symptoms and other health indicators and access educational resources.	Between 2009-2013, tool had 421 registered users and 54% were active.	N/A
Hudson et al [46]	Renal PatientView	Physical (QOL)	10 patients with chronic kidney disease; 20% female; age mean 59 years	Ethnographic qualitative study	Renal Patient Exchange Group, United Kingdom	PRO collection and symptom management for patients with chronic kidney disease. Patients can document and monitor symptoms and other health indicators and access educational resources.	Not collected	N/A

Study	Portal name	PCEOL[a] domain	Sample size and characteristics	Study design	Developer	Summary of features	Summary of portal use	Promotion of use
Pai et al [47]	Provider	Psychological or psychiatric	22 male patients with prostate cancer; 95% White; mean age 64 years (51-76)	Nonexperimental posttest only	British Columbia Cancer Agency and the School of Health Information Science at the University of Victoria, Victoria, BC, Canada	Psychosocial support for patients with prostate cancer. Allows patients to view a summary of prostate cancer diagnosis, cancer treatment received, and monitor health indicators. Patients can also access a treatment decision support tool, a questionnaire on distress level, education on prostate cancer and its treatments, and a tailored clinical trial and research screening tool.	Mean registered logins over 6 months was 3.4. Medical records, appointment viewer, and PSA[m] monitoring tool were accessed most frequently by users.	N/A
Nahm et al [48]	CaS-PET[n]	Psychological or psychiatric	30 patients with cancer; 77% female; 33% White; mean age 56 years (SD 13.6)	One group pre- or posttest pilot study	University of Maryland Medical Center, Baltimore, MD, United States	Psychosocial support for patients with cancer. Provides 12-week content with six modules including transition to survivorship, nutrition, exercise, cancer and relationships, fear and mental health, and stress management using mindfulness. There are goal setting activities, discussion boards, and virtual libraries at the end of each module.	Not collected	Biweekly follow-up messages were sent out after initial message.

[a]PCEOL: palliative care and end-of-life.

[b]ACP: advance care planning.

[c]HCP: health care proxy.

[d]N/A: not applicable.

[e]AD: advance directive.

[f]EHR: electronic health record.

[g]MDPOA: Medical Durable Power of Attorney.

[h]QOL: quality of life.

[i]PRO: patient-reported outcome.

[j]EMR: electronic medical record.

[k]PROMIS: Patient-Reported Outcomes Measurement Information System.

[l]e-PRO: electronic patient-reported outcome.

[m]PSA: prostate-specific antigen

[n]CaS-PET: Cancer Survivorship Patient Engagement Toolkit.

Features of PCEOL Patient Portal Tools

Ethical or Legal

All five tools addressing ethical and legal aspects of care center on ACP, including ACP education and the opportunity to complete documentation of advance directive, or of an appointed MDPOA or health care proxy. The ACP tools in each of these patient portals provide education and a means to facilitate documentation of advance care preferences. We identified two tools (PatientSite and MyChart) that use a guided, patient-centered questionnaire to enable the completion of an advance directive [24,33,34]. Although most of the tools allow users to complete and submit an advance directive that directly integrates into the electronic health record [22,23,35,36], we identified one tool based on My Health Manager (a Kaiser Permanente portal built through Epic software) that links users to an external website to complete the advance directive [37].

Physical

The identified PCEOL patient portal tools addressing physical aspects of care focus on improving the quality of life of patients with serious illness. Four of the five tools identified allow users to complete patient-reported outcome assessments and other questionnaires electronically [42-46]. The results of the assessments were automatically integrated into the electronic

medical record and then used to generate an action. One of the tools incorporated the Patient-Reported Outcomes Measurement Information System (PROMIS), which employs computer-adaptive tests to measure patient-reported symptoms and health-related quality of life across various medical conditions [42,43]. Wagner et al [42] administered an assessment that included a subset of PROMIS computer-adapted tests measuring pain interference, fatigue, physical function, depression, and anxiety and a checklist to identify psychosocial, informational, and nutritional needs [43] in patients with gynecological cancers. For each outcome, a score above a specific threshold triggers an alert to the appropriate member of the multidisciplinary care team for further evaluation. This assessment was also conducted in a larger cohort of adult medical oncology outpatients [42].

Similarly, MyAVL (MijnAVL), a patient portal for patients with breast and non–small cell lung cancer in the Netherlands, allows users to complete quality of life and physical activity questionnaires. Users are provided with tailored physical activity advice based on their responses. The commercially developed Carevive Care Planning System allows for the collection of patient-reported outcomes in patients with breast and gynecological cancers [44]. Targeted at both patients and caregivers, the users' responses were used to create a care plan that included clinical decision support and patient education and self-management resources. Renal PatientView, developed in the United Kingdom, was the only tool identified in patients with chronic kidney disease [45,46]. Renal PatientView allows users to monitor symptoms and other health indicators (eg, weight, blood pressure, and blood glucose) electronically and access educational content specific to their diagnosis.

Only one of the tools addressing quality of life, the Patient-Centered Tool Kit (PCTK), was designed for use during acute hospitalization [38,39]. The PCTK has been studied in patients admitted to medical intensive care and oncology units and their caregivers. Integrated with the patient's electronic health record, the tool allows users to establish a single recovery goal that can be updated throughout hospitalization. Users can also access educational content, predischarge checklists, and navigate their plan of care.

Psychological

Two of the PCEOL patient portal tools addressed the psychological aspects of care, primarily providing resources for psychosocial support. The British Columbia Cancer Agency developed the *Provider* tool for patients with prostate cancer. This tool allows users to complete a questionnaire on distress levels and access treatment decision support resources [47]. The tool also provides users with education on prostate cancer, its treatment, and a personalized clinical trial screening tool. The Cancer Survivorship Patient Toolkit (CaS-PET) was developed to provide survivorship care plans for patients with cancer [48]. This tool uses a biweekly follow-up contact via secure messaging. Each message asks about the patient's general condition and support needs and includes links to well beyond cancer, a web-based resource with six modules that include transition to survivorship, nutrition, exercise, cancer and relationships, fear and mental health, and stress management using mindfulness. The modules incorporate goal setting, discussion boards, and virtual libraries.

Information on Use or Promotion of Use

Portal use measures were described for nearly all (n=10) patient portal tools in at least one study. Table 1 illustrates these findings in detail. Portal use was most often measured by the mean number of log-ins and duration in minutes [40,41,47], number of registered users [45], features accessed [38,40,41,47], or completed activities such as completing an AD or other ACP-related documentation [22,24,33-36] or inputting a daily goal [38].

Overall, an overview of appointments and medical records were the most frequently accessed features for portals [38,40,41,47]. The Patient-Centered Toolkit assessed that education on patients' *problem* or condition was the least visited feature; however, this could be explained by the study's in-patient setting where the patients and caregivers had more frequent access to providers [38,39].

Studies for four of the portals specifically described methods of promoting portal use, either by reminders via electronic messages [35,42,48] or in-person [39]. The impact of these promotions of use is mixed; two of the studies did not collect data on portal use to correspond to their promotion efforts. After sending electronic messages prompting patients to complete their e-PRO assessment before their scheduled appointment, Wagner et al [42] found that 80% of recipients read the message, while 33% of them completed the electronic patient-reported outcomes (e-PRO) assessment.

User Perspectives

Usability, acceptability, and user satisfaction were formally evaluated for nine of the identified PCEOL patient portal tools (Table 2). Only two studies have specifically assessed the user perspectives of older adults [37,49]. In general, usability and acceptability rankings were high, and users were satisfied with the tools. Users valued tools that allowed the electronic collection of patient-reported outcomes and symptoms. Aside from their perceived convenience, these tools allow users to gain a sense of control over their health, improve symptom management, and access novel resources [40,41,44,47,48].

Users viewed these tools as a means of improving communication. In particular, the integration of tools with electronic medical records was valued. In ACP tools, users appreciated that the advance directive forms were easily retrievable and viewable for their providers and family [23,24,37]. The tools addressing quality of life and psychosocial support were also perceived to improve communication with the care team [39,44,47,48]. There was evidence of increased care plan concordance between patients [39], and users noted that the tools better prepared them for their clinical visits [44,47,48]. Users of the CaS-PET felt that the tool helped them determine what questions to ask during their medical appointments.

Table 2. Portal tool usability and satisfaction.

Portal tool	Usability sample size (overall sample size)	Data collection	Mean usability or satisfaction scores or qualitative comments	Comments for improvement
PatientSite [24]	• 74 (200)	• Qualitative feedback	• Patient participants had positive feedback on usability, tool prompted patients to take the time to think about their HCP[a], gave them an opportunity to improve their HCP information, and helped patients tackle this difficult topic	• Patients made minor suggestions to improve the capability to edit form directly, to improve the ease of printing the form, and to add ability to appoint an additional HCP.
MyHealth Connection (Epic) [22,23]	• 11 (2814) • 46 (46); 63% female; mean age 49 years	• SUS[b] • Qualitative feedback	• 89 (a total of 70 or greater is typically considered acceptable for usability) • Patients were generally satisfied with the ease of use and were likely to recommend using the tool for ACP[c] documentation to others.	• N/A[d]
My Health Manager (Epic) [37]	• 24 (24); 71% female; 79% White; mean age 78 years (SD 5.4)	• Qualitative feedback (focus groups)	• Most participants reported interest in having AD[e] documentation features available in the electronic medical record	• N/A
PCTK[f] [38]	• 18 (239); 10 patients; 50% female; 80% White; 70% >51 years • 8 caregivers; 75% female; 87% White; 87% >51 years	• SUS	• 74 (a total of 70 or greater is typically considered acceptable for usability)	• Feedback included suggestion for improving technical features and displays to enhance clinical communication.
MyAVL (MijnAVL) [40,41]	• 28 (37) • 92 (92); all female; mean age 49 years (SD 11.4)	• WUS[g]; UTAUT[h] • WUS; UTAUT	• 3.9 (maximum score of 5, indicated the highest level of satisfaction); 93% reported tool easy to use, 69% reported tool valuable addition to health care experience • 3.8; 75% reported tool easy to use	• N/A • Focus groups expressed overall satisfaction with portal features, however expressed desire for educational content to be more tailored to their specific condition
Carevive Care Planning System [44]	• 94 (121)	• SUS	• 83 (a total of 70 or greater is typically considered acceptable for usability)	• N/A
Renal PatientView [45,46]	• 190 (190) [45]; 34% female; median age 53 years; range 19-86 years • 10 (10); 20% female; mean age 59 years	• Investigator-developed questionnaire • Qualitative feedback	• 45% of inactive users cited computer or password issues as primary issue for nonuse; 37% of inactive users said portal did not anything to their relationship with clinicians • Patients found the portal valuable to help prepare themselves and family for changes in care and had better understanding of how their symptoms, blood results, and physiological changes were connected; able to better involve family in care	• N/A • N/A

Portal tool	Usability sample size (overall sample size)	Data collection	Mean usability or satisfaction scores or qualitative comments	Comments for improvement
Provider [47]	• 22 (22); all male; 95% White; mean age 64 years; range 51-76 years	• Questionnaire not specified	• 88% of respondents rated overall satisfaction as excellent or very good; 88% of respondents would continue to use the tool	• N/A
CaS-PET[i] [48]	• 30 (30); 77% female; 33% White; mean age 56 years (SD 13.6)	• Health Website Usability Questionnaire subscale; open-ended questions	• Most patients (n=22) found the portal helpful and reported having helpful information or that it helped them stay healthy	• N/A

[a]HCP: health care proxy.

[b]SUS: Systematic Usability Scale.

[c]ACP: advance care planning.

[d]N/A: not applicable.

[e]AD: advance directive.

[f]PCTK: Patient-Centered Tool Kit.

[g]WUSQ: Website User Satisfaction.

[h]UTAUT: Unified Theory of Acceptance and Use of Technology.

[i]CaS-PET: Cancer Survivorship Patient Engagement Toolkit.

Discussion

Principal Findings

Patient portals may offer accessible avenues for keeping individuals with serious illness and their caregivers connected to PCEOL resources. Although the use of patient portals among patients with serious life-limiting illness is of growing interest, previous studies suggest that patients with serious illness and their caregivers are potentially extensive users of patient portal tools that are commonly available, particularly to access care team information, review laboratory results, and obtain medical records [50,51]. Interest in using patient portals among this population may be high and potentially an excellent venue for providing PCEOL-specific information and resources. This scoping review of 19 articles described available PCEOL features via patient portals tethered to electronic health records. Overall, the review highlighted some of the ways patient portals are offering PCEOL resources, as well as opportunities for future work to leverage patient portals to increase access and awareness of PCEOL care.

The results identified 12 unique patient portal resources supporting the following PCEOL domains: ethical or legal (in the case of this review, exclusively ACP education; n=5), physical (including symptom monitoring, goals of care, and patient-reported outcomes; n=5), and psychological support (n=2). Most resources were targeted toward patient audiences (n=9), with the remainder being targeted toward caregivers, or both patients and caregivers. Only two resources focused specifically on the user characteristics of older adults [37,49], which indicates a clear opportunity to improve our understanding of use and usability in this high-need patient population.

Evidence demonstrates that patients living with serious illnesses and their family members and caregivers can benefit from palliative care [52]. However, there are nearly 1 million patients who are admitted to the hospital annually who could benefit from palliative care and yet do not receive this specialized care [53]. In addition, people living in rural communities have less access to palliative care services than their urban counterparts [54]. Although we recognize that palliative care is highly personalized, patient portals offer the opportunity to leverage technology to promote access and education regarding PCEOL issues. Incorporating PCEOL resources into patient portal features may serve to introduce or supplement specialty palliative care services or bridge the gap where patients have no access to palliative care.

Furthermore, the role of digital health solutions, including patient portal resources, is more important than ever in the wake of the COVID-19 pandemic. Health systems across the United States, and the globe, are forced to incorporate digital care as part of their regular repertoire of available patient care as safety precautions and social distancing have become the norm [55,56]. Individuals who are appropriate for palliative care are highly vulnerable and must be especially considered when addressing how to bring quality virtual health care to individuals during a pandemic. These patients have an increased risk of contracting the virus and experiencing related complications (eg, hospitalization or death) because of their age and health status. They have also been uniquely impacted by the pandemic in that they still require continuous disease management and may experience disproportionate exacerbations of psychosocial needs [57]. Finally, people are dying or experiencing bereavement in isolation at higher rates in the wake of the pandemic [58,59], emphasizing the need for creative palliative care solutions, including better access to PCEOL information and resources for individuals with serious illness and their loved ones.

Nearly all the portals included in this review focused on quality of life, symptom management, or ACP education or completion of advance directives. ACP may be better leveraged through patient portals than other resource types given its more limited

and specific goal than some of the more complex and nuanced needs and components inherent in PCEOL care. Similarly, PROMIS tools and other electronic mechanisms for patient-reported outcomes are well suited for patient portals.

Although the legal and physical aspects are essential pieces of quality and holistic PCEOL services, equally important are the psychological and spiritual aspects, as well as end-of-life education around hospice. Only two of the portals reviewed focused on the psychological needs of patients with serious illness [47,49]. None of the portals included in the review offered features addressing grief and bereavement or information about hospice or comfort care. Despite the evidence supporting the benefits of hospice care [60-62], lack of information and understanding remains one of the most common barriers to hospice use [63-65]. Furthermore, bereaved individuals and families often benefit from educational resources related to the grief process [66]. Future research and development on portal tools for patients with serious illness and their family members should emphasize psychological and spiritual PCEOL domains and consider tools and educational resources regarding grief, bereavement, and hospice.

It is important to emphasize the special consideration of older adult patients with serious illness and patient portal tool development, availability, and adoption among PCEOL populations. In total, 80% of older adults in the United States have one or more chronic conditions, many of which can lead to serious illness [9]. As the older adult population continues to grow, and subsequently the number of older adults experiencing serious illness increases, the focus of patient portals' development and research must continue to include the perspectives and needs of patients aged ≥65 years. Despite the currently low adoption rate of portals by older patients and caregivers, research demonstrates that there is an interest in such tools among this group [37,67], an interest that might well be addressed by improving the design and availability of solutions tailored to and especially suited to the needs of older patients [68]. Future research on PCEOL patient portal development, use, and acceptability should explicitly incorporate older adult needs and feedback.

Limitations

To the best of our knowledge, this scoping review is the first to identify and describe patient portals that offer PCEOL features. However, this is not without limitations. First, the review only included results from academic literature and may be missing commercially available portals that were developed by industry without concurrent publication in the peer-reviewed literature. In addition, we did not evaluate the full text of articles in which we were unable to retrieve access, potentially overlooking additional PCEOL patient portal resources. Although some of the articles described the efficacy on clinically meaningful outcomes, we did not address this information as it was beyond the scope of our scoping view. Future research should focus on the efficacy of such tools. Finally, this review had a small sample size of articles (n=19) covering 12 different patient portals that are predominantly US based, potentially limiting the results to a Westernized PCEOL perspective.

Conclusions

The scoping review highlighted an important gap in PCEOL resources offered via patient portals linked to electronic health records, as well as opportunities for leveraging patient portals as a means of offering education and support for patients with serious illness. The results suggest that there are ongoing efforts to offer various PCEOL supports through patient portals, particularly with a focus on ACP. However, our review resulted in a small number of portals that met our criteria, of which included only a few of the expansive elements of PCEOL care. Future research and development of patient portals would benefit from offering comprehensive PCEOL features to increase access and education for patients experiencing serious illness.

Acknowledgments

This research was funded by the Medical Student Training in Aging Research Program, the National Institute on Aging (T35AG026736), the John A. Hartford Foundation, and the Lillian R. Gleitsman Foundation. JP was supported by the National Institute on Aging (NIA K76AG059934).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Review (PRISMA-ScR) checklist.
[PDF File (Adobe PDF File), 188 KB - jmir_v23i9e28797_app1.pdf]

Multimedia Appendix 2
Scoping review search terms.
[PDF File (Adobe PDF File), 52 KB - jmir_v23i9e28797_app2.pdf]

References

1. What is a patient portal? HealthIT. 2017 Sep 29. URL: https://www.healthit.gov/faq/what-patient-portal [accessed 2020-10-15]

2. Bourgeois F, Mandl K, Shaw D, Flemming D, Nigrin D. Mychildren's: integration of a personally controlled health record with a tethered patient portal for a pediatric and adolescent population. AMIA Annu Symp Proc 2009 Nov 14;2009:65-69 [FREE Full text] [Medline: 20351824]
3. Are portals a means to patient activation? Medical Group Management Association. 2018. URL: https://www.mgma.com/data/data-stories/are-portals-a-means-to-patient-activation [accessed 2020-03-24]
4. Byczkowski T, Munafo J, Britto M. Family perceptions of the usability and value of chronic disease web-based patient portals. Health Informatics J 2014 Jun;20(2):151-162 [FREE Full text] [doi: 10.1177/1460458213489054] [Medline: 24056751]
5. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Dec 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
6. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015 Mar 20;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
7. Otte-Trojel T, de Bont A, Rundall T, van de Klundert J. How outcomes are achieved through patient portals: a realist review. J Am Med Inform Assoc 2014;21(4):751-757 [FREE Full text] [doi: 10.1136/amiajnl-2013-002501] [Medline: 24503882]
8. Hoogenbosch B, Postma J, de Man-van Ginkel JM, Tiemessen NA, van Delden JJ, van Os-Medendorp H. Use and the users of a patient portal: cross-sectional study. J Med Internet Res 2018 Sep 17;20(9):e262 [FREE Full text] [doi: 10.2196/jmir.9418] [Medline: 30224334]
9. Chronic conditions. Centers for Medicare and Medicaid Services. 2021 Jan 15. URL: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/CC_Main [accessed 2021-03-15]
10. Smith S, O'Conor R, Aitken W, Curtis L, Wolf M, Goel M. Disparities in registration and use of an online patient portal among older adults: findings from the LitCog cohort. J Am Med Inform Assoc 2015 Jul;22(4):888-895 [FREE Full text] [doi: 10.1093/jamia/ocv025] [Medline: 25914099]
11. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, López A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011 May 01;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]
12. Walker D, Hefner J, Fareed N, Huerta T, McAlearney A. Exploring the digital divide: age and race disparities in use of an inpatient portal. Telemed J E Health 2020 May;26(5):603-613 [FREE Full text] [doi: 10.1089/tmj.2019.0065] [Medline: 31313977]
13. Graetz I, Gordon N, Fung V, Hamity C, Reed M. The digital divide and patient portals: internet access explained differences in patient portal use for secure messaging by age, race, and income. Med Care 2016 Aug;54(8):772-779. [doi: 10.1097/MLR.0000000000000560] [Medline: 27314262]
14. Perrin A, Duggan M. Americans' internet access: 2000-2015. Pew Research Center. 2015 Jun 26. URL: https://www.pewresearch.org/internet/2015/06/26/americans-internet-access-2000-2015/ [accessed 2020-09-24]
15. Prestin A, Vieux SN, Chou WS. Is online health activity alive and well or flatlining? Findings from 10 years of the health information national trends survey. J Health Commun 2015 Jun 04;20(7):790-798. [doi: 10.1080/10810730.2015.1018590] [Medline: 26042588]
16. Capurro D, Ganzinger M, Perez-Lu J, Knaup P. Effectiveness of eHealth interventions and information needs in palliative care: a systematic literature review. J Med Internet Res 2014 Mar 07;16(3):e72 [FREE Full text] [doi: 10.2196/jmir.2812] [Medline: 24610324]
17. Morrison RS, Meier DE. Palliative care. N Engl J Med 2004 Jun 17;350(25):2582-2590. [doi: 10.1056/nejmcp035232]
18. Institute of Medicine. Dying in America: Improving quality and honoring individual preferences near the end of life. In: Institute of Medicine. Washington, D.C: National Academies Press; 2014.
19. Kelley AS, Morrison RS. Palliative care for the seriously ill. N Engl J Med 2015 Aug 20;373(8):747-755. [doi: 10.1056/nejmra1404684]
20. Hawley P. Barriers to access to palliative care. Palliat Care 2017;10:1178224216688887 [FREE Full text] [doi: 10.1177/1178224216688887] [Medline: 28469439]
21. Shalev A, Phongtankuel V, Kozlov E, Shen M, Adelman R, Reid M. Awareness and misperceptions of hospice and palliative care: a population-based survey study. Am J Hosp Palliat Care 2018 Mar;35(3):431-439 [FREE Full text] [doi: 10.1177/1049909117715215] [Medline: 28631493]
22. Lum HD, Brungardt A, Jordan SR, Phimphasone-Brady P, Schilling LM, Lin C, et al. Design and implementation of patient portal-based advance care planning tools. J Pain Symptom Manage 2019 Jan;57(1):112-117 [FREE Full text] [doi: 10.1016/j.jpainsymman.2018.10.500] [Medline: 30595147]
23. Jordan SR, Brungardt A, Phimphasone-Brady P, Lum HD. Patient perspectives on advance care planning via a patient portal. Am J Hosp Palliat Care 2019 Aug 25;36(8):682-687 [FREE Full text] [doi: 10.1177/1049909119832820] [Medline: 30803245]

24. Bajracharya A, Crotty B, Kowaloff H, Safran C, Slack W. Improving health care proxy documentation using a web-based interview through a patient portal. J Am Med Inform Assoc 2016 May;23(3):580-587 [FREE Full text] [doi: 10.1093/jamia/ocv133] [Medline: 26568608]
25. What are palliative care and hospice care? National Institute on Aging. 2021. URL: https://www.nia.nih.gov/health/what-are-palliative-care-and-hospice-care#hospice [accessed 2021-07-01]
26. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
27. Tricco A, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
28. Meghani S, MacKenzie M, Morgan B, Kang Y, Wasim A, Sayani S. Clinician-targeted mobile apps in palliative care: a systematic review. J Palliat Med 2017 Oct;20(10):1139-1147. [doi: 10.1089/jpm.2017.0070] [Medline: 28557549]
29. Portz JD, Elsbernd K, Plys E, Ford KL, Zhang X, Gore MO, et al. Elements of social convoy theory in mobile health for palliative care: scoping review. JMIR Mhealth Uhealth 2020 Jan 06;8(1):e16060 [FREE Full text] [doi: 10.2196/16060] [Medline: 31904581]
30. Covidence. URL: https://www.covidence.org/ [accessed 2021-07-01]
31. American Academy of HospicePalliative Medicine, Center to Advance Palliative Care, HospicePalliative Nurses Association, Last Acts Partnership, National HospicePalliative Care Organization. National Consensus Project for Quality Palliative Care: Clinical Practice Guidelines for quality palliative care, executive summary. J Palliat Med 2004 Oct;7(5):611-627 [FREE Full text] [doi: 10.1089/jpm.2004.7.611] [Medline: 15588352]
32. Kelley A, Bollens-Lund E. Identifying the population with serious illness: The "Denominator" challenge. J Palliat Med 2018 Mar;21(S2):7-16 [FREE Full text] [doi: 10.1089/jpm.2017.0548] [Medline: 29125784]
33. Bose-Brill S, Feeney M, Prater L, Miles L, Corbett A, Koesters S. Validation of a novel electronic health record patient portal advance care planning delivery system. J Med Internet Res 2018 Jun 26;20(6):e208 [FREE Full text] [doi: 10.2196/jmir.9203] [Medline: 29945860]
34. Bose-Brill S, Kretovics M, Ballenger T, Modan G, Lai A, Belanger L, et al. Testing of a tethered personal health record framework for early end-of-life discussions. Am J Manag Care 2016 Jul 01;22(7):258-263 [FREE Full text] [Medline: 27442309]
35. Tieu C, Chaudhry R, Schroeder D, Bock F, Hanson G, Tung E. Utilization of patient electronic messaging to promote advance care planning in the primary care setting. Am J Hosp Palliat Care 2017 Aug;34(7):665-670. [doi: 10.1177/1049909116650237] [Medline: 27188759]
36. Brungardt A, Daddato AE, Parnes B, Lum HD. Use of an ambulatory patient portal for advance care planning engagement. J Am Board Fam Med 2019 Nov 08;32(6):925-930 [FREE Full text] [doi: 10.3122/jabfm.2019.06.190016] [Medline: 31704762]
37. Portz JD, Lum HD, Bull S, Boxer RS, Bekelman DB, Ford KL, et al. Perceptions of patient portal use for advance directive documentation among older adults with multiple chronic conditions. J Soc Work End Life Palliat Care 2020 Jun 01;16(3):238-249. [doi: 10.1080/15524256.2020.1771806] [Medline: 32482160]
38. Dalal A, Dykes P, Collins S, Lehmann L, Ohashi K, Rozenblum R, et al. A web-based, patient-centered toolkit to engage patients and caregivers in the acute care setting: a preliminary evaluation. J Am Med Inform Assoc 2016 Jan;23(1):80-87 [FREE Full text] [doi: 10.1093/jamia/ocv093] [Medline: 26239859]
39. Dalal AK, Dykes P, Samal L, McNally K, Mlaver E, Yoon CS, et al. Potential of an electronic health record-integrated patient portal for improving care plan concordance during acute care. Appl Clin Inform 2019 May 29;10(3):358-366 [FREE Full text] [doi: 10.1055/s-0039-1688831] [Medline: 31141830]
40. Kuijpers W, Groen W, Oldenburg H, Wouters M, Aaronson N, van Harten WH. eHealth for breast cancer survivors: use, feasibility and impact of an interactive portal. JMIR Cancer 2016 May 10;2(1):e3 [FREE Full text] [doi: 10.2196/cancer.5456] [Medline: 28410178]
41. Groen W, Kuijpers W, Oldenburg H, Wouters M, Aaronson N, van Harten WH. Supporting lung cancer patients with an interactive patient portal: feasibility study. JMIR Cancer 2017 Aug 08;3(2):e10 [FREE Full text] [doi: 10.2196/cancer.7443] [Medline: 28790025]
42. Wagner L, Schink J, Bass M, Patel S, Diaz M, Rothrock N, et al. Bringing PROMIS to practice: brief and precise symptom screening in ambulatory cancer care. Cancer 2015 Mar 15;121(6):927-934 [FREE Full text] [doi: 10.1002/cncr.29104] [Medline: 25376427]
43. Garcia SF, Wortman K, Cella D, Wagner LI, Bass M, Kircher S, et al. Implementing electronic health record-integrated screening of patient-reported symptoms and supportive care needs in a comprehensive cancer center. Cancer 2019 Nov 15;125(22):4059-4068 [FREE Full text] [doi: 10.1002/cncr.32172] [Medline: 31373682]
44. Brant J, Hirschman K, Keckler S, Dudley W, Stricker C. Patient and Provider Use of Electronic Care Plans Generated From Patient-Reported Outcomes. Oncol Nurs Forum 2019 Nov 1;46(6):715-726. [doi: 10.1188/19.onf.715-726]

45. Hazara A, Bhandari S. Barriers to patient participation in a self-management and education website Renal PatientView: A questionnaire-based study of inactive users. Int J Med Inform 2016 Mar;87:10-14. [doi: 10.1016/j.ijmedinf.2015.12.004] [Medline: 26806707]
46. Hudson C, Darking M, Cox J. Understanding the value of patientview for enabling self-care practice in chronic kidney disease. J Ren Care 2020 Mar 16;46(1):13-24. [doi: 10.1111/jorc.12300] [Medline: 31621197]
47. Pai H, Lau F, Barnett J, Jones S. Meeting the health information needs of prostate cancer patients using personal health records. Curr Oncol 2013 Dec 01;20(6):561-569 [FREE Full text] [doi: 10.3747/co.20.1584] [Medline: 24311957]
48. Nahm E, Son H, Yoon JM. Older adults' use of patient portals: Experiences, challenges, and suggestions shared through discussion board forums. Geriatr Nurs 2020 Jul;41(4):387-393 [FREE Full text] [doi: 10.1016/j.gerinurse.2019.12.001] [Medline: 31899003]
49. Nahm E, Miller K, McQuaige M, Corbitt N, Jaidar N, Rosenblatt P, et al. Testing the impact of a cancer survivorship patient engagement toolkit on selected health outcomes. Oncol Nurs Forum 2019 Sep 1;46(5):572-584. [doi: 10.1188/19.onf.572-584]
50. Pho K, Lu R, Gates S, Xie Y, Lee SJ, Gerber DE. Characteristics of patients using patient portals in oncology. JAMA Oncol 2018 Mar 01;4(3):416-418 [FREE Full text] [doi: 10.1001/jamaoncol.2017.5257] [Medline: 29372242]
51. Portz JD, Powers JD, Baldwin M, Bekelman DB, Casillas A, Kutner JS, et al. Patient portal use near the end-of-life. J Gen Intern Med 2021 Jan 27:A (forthcoming). [doi: 10.1007/s11606-020-06333-9] [Medline: 33506403]
52. Kavalieratos D, Corbelli J, Zhang D, Dionne-Odom J, Ernecoff N, Hanmer J, et al. Association between palliative care and patient and caregiver outcomes: a systematic review and meta-analysis. J Am Med Assoc 2016 Nov 22;316(20):2104-2114 [FREE Full text] [doi: 10.1001/jama.2016.16840] [Medline: 27893131]
53. Morrison R, Augustin R, Souvanna P, Meier D. America's care of serious illness: a state-by-state report card on access to palliative care in our nation's hospitals. J Palliat Med 2011 Oct;14(10):1094-1096 [FREE Full text] [doi: 10.1089/jpm.2011.9634] [Medline: 21923412]
54. Lynch S. Hospice and palliative care access issues in rural areas. Am J Hosp Palliat Care 2013 Mar 24;30(2):172-177. [doi: 10.1177/1049909112444592] [Medline: 22531150]
55. Calton B, Abedini N, Fratkin M. Telemedicine in the time of coronavirus. J Pain Symptom Manage 2020 Jul;60(1):12-14 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.03.019] [Medline: 32240756]
56. Wosik J, Fudim M, Cameron B, Gellad Z, Cho A, Phinney D, et al. Telehealth transformation: COVID-19 and the rise of virtual care. J Am Med Inform Assoc 2020 Jun 01;27(6):957-962 [FREE Full text] [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]
57. Nekhlyudov L, Duijts S, Hudson SV, Jones JM, Keogh J, Love B, et al. Addressing the needs of cancer survivors during the COVID-19 pandemic. J Cancer Surviv 2020 Oct 25;14(5):601-606 [FREE Full text] [doi: 10.1007/s11764-020-00884-w] [Medline: 32335850]
58. Wallace C, Wladkowski S, Gibson A, White P. Grief During the COVID-19 Pandemic: Considerations for palliative care providers. J Pain Symptom Manage 2020 Jul;60(1):70-76 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.04.012] [Medline: 32298748]
59. Nelson B. Covid-19 is shattering US cancer care. Br Med J 2020 Apr 17;369:m1544. [doi: 10.1136/bmj.m1544] [Medline: 32303501]
60. Greer D, Mor V, Morris J, Sherwood S, Kidder D, Birnbaum H. An alternative in terminal care: Results of the national hospice study. J Chron Dis 1986 Jan;39(1):9-26. [doi: 10.1016/0021-9681(86)90103-7]
61. Cagle J, Pek J, Clifford M, Guralnik J, Zimmerman S. Correlates of a good death and the impact of hospice involvement: findings from the national survey of households affected by cancer. Support Care Cancer 2015 Mar;23(3):809-818 [FREE Full text] [doi: 10.1007/s00520-014-2404-z] [Medline: 25194877]
62. Brumley R, Enguidanos S, Jamison P, Seitz R, Morgenstern N, Saito S, et al. Increased satisfaction with care and lower costs: results of a randomized trial of in-home palliative care. J Am Geriatr Soc 2007 Jul;55(7):993-1000. [doi: 10.1111/j.1532-5415.2007.01234.x] [Medline: 17608870]
63. Tate C, Venechuk G, Brereton E, Ingle P, Allen L, Morris M, et al. "It's Like a Death Sentence but It Really Isn't" What patients and families want to know about hospice care when making end-of-life decisions. Am J Hosp Palliat Care 2020 Sep;37(9):721-727. [doi: 10.1177/1049909119897259] [Medline: 31888342]
64. Cagle J, Van Dussen DJ, Culler K, Carrion I, Hong S, Guralnik J, et al. Knowledge about hospice: exploring misconceptions, attitudes, and preferences for care. Am J Hosp Palliat Care 2016 Feb;33(1):27-33. [doi: 10.1177/1049909114546885] [Medline: 25138649]
65. Cicolello K, Anandarajah G. Multiple stakeholders' perspectives regarding barriers to hospice enrollment in diverse patient populations: a qualitative study. J Pain Symptom Manage 2019 May;57(5):869-879. [doi: 10.1016/j.jpainsymman.2019.02.012] [Medline: 30790720]
66. Meichsner F, O'Connor M, Skritskaya N, Shear M. Grief before and after bereavement in the elderly: an approach to care. Am J Geriatr Psychiatry 2020 May;28(5):560-569. [doi: 10.1016/j.jagp.2019.12.010] [Medline: 32037292]
67. Irizarry T, Shoemake J, Nilsen M, Czaja S, Beach S, Dabbs A. Patient portals as a tool for health care engagement: a mixed-method study of older adults with varying levels of health literacy and prior patient portal use. J Med Internet Res 2017 Mar 30;19(3):e99 [FREE Full text] [doi: 10.2196/jmir.7099] [Medline: 28360022]

68. Czaja SJ, Boot WR, Charness N, Rogers WA. Designing for Older Adults: Principles and Creative Human Factors Approaches. Boca Raton: CRC press; 2019:1-351.

Abbreviations

ACP: advance care planning
MDPOA: Medical Durable Power of Attorney
PCEOL: palliative care and end-of-life
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROMIS: Patient-Reported Outcomes Measurement Information System

Edited by R Kukafka; submitted 15.03.21; peer-reviewed by M Puts, R Chan; comments to author 10.05.21; revised version received 02.07.21; accepted 27.07.21; published 16.09.21.

Please cite as:
Ingle MP, Valdovinos C, Ford KL, Zhou S, Bull S, Gornail S, Zhang X, Moore S, Portz J
Patient Portals to Support Palliative and End-of-Life Care: Scoping Review
J Med Internet Res 2021;23(9):e28797
URL: https://www.jmir.org/2021/9/e28797
doi:10.2196/28797
PMID:34528888

Original Paper

Building Primary Health Care Personnel's Support for a Patient Portal While Alleviating eHealth-Related Stress: Survey Study

Iiris Hörhammer[1], DSC, MSc; Sari Kujala[2], PhD; Pirjo Hilama[3], MSc; Tarja Heponiemi[4], PhD

[1]Department of Industrial Engineering and Management, Aalto University, Espoo, Finland

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

[3]Social Health Care Joint Authority of South-Savo, Mikkeli, Finland

[4]Social and Health System Research Unit, National Institute for Health and Welfare, Helsinki, Finland

Corresponding Author:
Iiris Hörhammer, DSC, MSc
Department of Industrial Engineering and Management
Aalto University
Otaniementie 1
Espoo, 02150
Finland
Phone: 358 503430063
Email: iiris.riippa@aalto.fi

Abstract

Background: Health care personnel's (HCP) engagement in patient portal implementation is necessary in embedding the use of the portal in everyday practices of a health care organization. While portal implementation may raise personnel's positive expectations of the benefits in patient care, it is often also stressful for them due to increased workloads and disruptions in clinical workflows. An understanding of social and technical factors that build personnel's support for patient portal implementation and alleviate their eHealth-related stress is therefore needed to realize the full potential of portals.

Objective: The aim of this study was to explore the influence of managerial implementation practices, information technology (IT) usability, and personnel's eHealth competences on support for patient portal implementation and eHealth-related stress among primary HCP.

Methods: The data were collected through a survey of 919 members at 2 health organizations in Finland. Linear and logistic regression models were fitted to study the associations between the variables.

Results: Professionals' eHealth competence (β=.15, *P*<.001), usability (β=.11, *P*<.001), and implementation practices (β=.07, *P*<.001) were positively associated with professionals' support and negatively associated with professionals eHealth-related stress (β=−.07, *P*=.010; β=−.27, *P*<.001; and β=−.14, *P*<.001, respectively). Professionals' support was associated with their promotion of the portal to the patients (odds ratio 1.22, 95% CI 1.07-1.40).

Conclusions: The adoption of appropriate implementation practices and the usability of the technology can build personnel's support for a patient portal and alleviate their stress related to eHealth. Personnel's support is manifested in their promotion of the portal to patients. Health care managers are encouraged to consider the usability of the technology and the good implementation practices, such as proper informing, engagement of the personnel in planning the services, and allocation of resources to improve eHealth competence, as prerequisites for meaningful and sustainable use of patient portals.

(J Med Internet Res 2021;23(9):e28976) doi:10.2196/28976

KEYWORDS

patient portal; implementation; adoption; health care personnel; eHealth-related stress

Introduction

Background

Patient portals provide online access to personal health information and interactions with health care providers. Thus, they can support patient self-management and enhance shared decision making [1-3]. However, as a precondition to the benefits of patient portals, studies have emphasized the importance of substantial initial investments required of health care organizations in the implementation [4-6]. Management

must embed the use of portals in daily clinical practice to support their meaningful and sustained use [6]. Success of the implementation ultimately relies on how the health care personnel (HCP), together with the patients, use the portal [6,7]. HCP's effort in making sense of the portal utility together with colleagues and patients is necessary for the normalization of the use [7-9]. Local champions persuade their peers and patients of the "professional appropriateness" of the new services [7]. While HCP's endorsement of patient portals is critical for the adoption and continued use by the patients [4,10,11], studies [12,13], even in controlled pilot settings [14], have identified the failure to engage frontline employees in the promotion of digital tools.

Management and technical factors may explain inadequate engagement among personnel. Poor managerial practices [15,16] and technical difficulties [4,10] have been shown to inhibit health care professionals' support. By contrast, good usability [6,15,17] and implementation practices, such as involving personnel in planning [18], and supporting the effective use of portals in clinical work [5,19-21], are essential to the successful adoption of health information technology (HIT). A study of the preimplementation phase of a patient portal found an association between health care professionals' expectations of their organization's implementation practices and their support for the portal [15]. However, little is known about this association in the postimplementation phase or whether personnel's support for the portal is manifested in their collaborations with patients.

Most studies have discussed the opportunities provided by patient portals; however, few have focused on their adverse effects on personnel and their work. The adoption of new information technology (IT) has raised concerns among health care professionals regarding increased workloads and disrupted clinical workflows [22,23]. A recent study in Finland found that poorly functioning, time-consuming, and inadequate information systems have emerged as a significant source of stress in HCP's work, and that stress related to information systems had increased between 2006 and 2015 [24]. Stress has been described to appear in a relationship between a person and the environment that is appraised as important for an individual but exceeds their coping resources [25]. While excessive stress experienced by HCPs may lead to decrease in their well-being, dissatisfaction at work, and even intent to leave the organization [26], the upward trend in HIT-related stress raises concerns about the sustainability of the health care system. HIT usability deficiencies have been associated with HCP's high stress related to information systems [27] and general self-rated stress [28]. However, little is known about the influence of the implementation practices of HIT on personnel's eHealth-related stress.

Building on recent sociotechnical research on the technical and organizational factors in successful HIT adoption and use [7,29-31], this study investigated the associations of (1) HCP's perception of *organizational implementation practices* and (2) *patient portal* usability with *their support* for the technology and their *eHealth-related stress*. Moreover, to identify the possible manifestation of HCP's support through their *interactions with the patients*, the association between personnel's support for and promotion of the portal was examined.

Research Model

Organizational practices surrounding health technology implementation can be determinative of personnel's successful adoption. Such practices include proper informing of the new services before implementation [15,32], engagement of the personnel in planning [15], adequate training [19-21], and enough time for learning [18]. Expectations of good implementation practices have been found to be associated with health care professionals' support for eHealth in the preimplementation phase [15]. Therefore, we hypothesize that:

> *H1: HCP's positive perception of the implementation practices in their unit is positively associated with their support for a patient portal.*

Studies have shown associations between HCP's perception of HIT usability and their acceptance of the technology [33-35]. Therefore, we hypothesize:

> *H2: Usability is positively associated with HCP's support for a patient portal.*

Usability has been shown to relate to the skills needed to use a new technology [33,36]. Therefore, we also propose that personnel's eHealth skills are associated with their support for the portal. In order to use eHealth in a meaningful way, personnel need new competences, not only in computer skills and literacy but also in the application of HIT in patient interactions, and the promotion of the technology to facilitate patient self-management [10,37]. Our hypothesis is:

> *H3: HCP's self-rated eHealth competences are positively associated with their support for a patient portal.*

HCP's endorsement has been deemed essential to patients' adoption and continued use of digital health tools [10,11]. Personnel who support a portal might be more likely to promote its use to their patients. We therefore hypothesize:

> *H4: HCP's support for patient portal is positively associated with their promotion of the portal to patients.*

Studies have identified HIT usability problems that disrupt HCP's routine work [38]. Usability issues have been associated with elevated stress related to information systems, which might be alleviated by experience in using the technology [24,27]. Therefore, we hypothesize the following:

> *H5: Usability is negatively associated with HCP's eHealth-related stress and*
>
> *H6: HCP's self-rated eHealth competence is negatively associated with their eHealth-related stress.*

In addition to personnel's competences, organizational practices in the development and implementation, such as end-user involvement in system development and work-procedure planning, have been suggested as potential alleviators of IT-related stress [24]. We hypothesize that:

H7: HCP's positive perception of the implementation practices in their unit is negatively associated with HCP's eHealth-related stress.

Methods

Study Setting

The survey study was conducted in 2 Finnish regional health and social care authorities (Organizations A and B) in 2017. Both authorities are public social and health care providers serving populations of approximately 100,000 (Organization A) and 40,000 (Organization B). At the time of the survey, the organizations had already implemented eHealth services, and aimed to increase their usage.

The functionalities of the patient portals in both organizations were online appointment booking, access to personal medical records, patient-reported medical history, and electronic messaging with HCP. In Organization A, digital symptom questionnaires, self-management instructions, and remote health care appointments had recently become available to certain patient groups. In both organizations, different functionalities had been gradually adopted over the 13 years preceding the survey; however, the pace had accelerated during the preceding 5 years.

An invitation to the online survey was sent to HCP via work email. The survey was also introduced on the health and social care authorities' intranet news page. No exclusion of a subgroup of employees was applied. Participation was anonymous; however, the respondents had an opportunity to win movie tickets and 2 wireless computer mice. The study protocol was reviewed and approved by Aalto University's Ethical Review Board.

Questionnaire

The questionnaire contained questions about demographics, support for the patient portal implemented by the regional health and social care authority, self-perceived eHealth competences, and HIT implementation practices in the respondent's unit. Personnel were also asked about usability and stress related to IT and the patient portal (Multimedia Appendix 1). All items, except those related to demographics and stress measurement, were answered on a 5-point Likert scale ranging from 1 (fully disagree) to 4 (fully agree), with the fifth option (don't know) omitted from the analysis.

eHealth competences were assessed with 8 statements from a previously used scale [37] encompassing the use of eHealth tools in the personnel's work. The use of the *recommended implementation practices* was assessed through 4 previously used items [15,39] regarding the resources and information given to the personnel for adopting new services and opportunities to participate in the planning of new services. *Personnel's support* for the patient portal was measured with 5 items from the previously validated I-SEE questionnaire [15,39-41]. The items address a respondent's personal and perceived collective support by colleagues and supervisors for the patient portal implementation.

IT *usability* was measured with 4 items as in our previous study [15]. We used 3 items from the Usability Metric for User Experience (UMUX) scale concerning utility, ease of use, and frustration related to the IT [42], and added a fourth item on general user satisfaction. A previously used measure [43-45] was used to assess *eHealth- and information systems–related stress*. This measure has been developed in Finland when examining the health and well-being of physicians [24,44]. It has previously been associated with, for example, experience in using information systems, cognitive workload, distress, and electronic health record usability [27,44]. The participants were asked how often during the previous 3 months they had been distracted by, worried about, or stressed about information systems, or IT equipment or software. A third item about stress related to the patient portal (eHealth services) was added to satisfy the scope of the study. The answers were rated on a 5-point Likert scale ranging from 1 (never) to 5 (very often).

To assess the frequency of HCP's *promotion of the portal to patients*, the participants were asked to choose one of the following options: "never," "1-4 times," "5-9 times," and "10 or more times." Because of the small proportion of responses in the categories other than "never," the 3 other response categories were combined to create a variable related to the promotion of the portal to a patient at least once (Yes/No).

Statistical Analysis

Descriptive statistics and reliability analyses were performed and mean sum scores were computed for all study variables (Table 1 and Multimedia Appendix 2). The scale reliabilities (Cronbach α) of the Likert scale measures were all over .70, and therefore at an acceptable level [46]. Although the lowest reliability of .71 for eHealth-related stress was not very high, it can still be considered acceptable due to its very short length with only 3 items [47]. The hypotheses were tested through regression analysis. Two linear regression models were fitted with (1) personnel's support for the patient portal and (2) their eHealth-related stress as the dependent variable. Usability, eHealth competence, eHealth-related stress, age, gender, work experience, and organization (Organization A or B) were the independent variables. A logistic regression model was fitted to assess the association between personnel's support for the portal and their promotion of the portal to patients. The model was adjusted for age, gender, work experience, and profession. We fitted the models stepwise and report estimates from the univariate models, multivariate model with independent variables of interest only, and multivariate model with adjustments. To test for multicollinearity, the variance inflation factors were calculated for the regression variables. All the variance inflation factors were lower than 1.7, thus indicating that multicollinearity was not a concern [48].

Results

Respondents

A majority of the respondents were women (800/919, 87.1%) and nurses or assistant nurses (589/919, 64.1%) with an average age of 44 years (SD 11.7) and 11.6 years (SD 10.2) of work experience in their current duties. The other personnel were secretaries, social workers, doctors, and psychologists or other

therapists (Table 1). The 919 respondents comprised approximately 20% of all health and social care personnel in the target organizations. The respondents well represented the Finnish health and social care professionals in terms of gender and age.

Table 1. Descriptive characteristics of participants (N=919).

Characteristic	Value
Gender, n (%)	
Man	98 (10.7)
Woman	800 (87.1)
Not reported	21 (2.3)
Age in years, mean (SD)	44.6 (11.7)
Profession, n (%)	
Nurse/midwife/public health nurse	310 (33.7)
Assistant/other nurse	279 (30.4)
Social worker	61 (6.6)
Ward/department secretary	55 (6.0)
Doctor/dentist	52 (5.7)
Psychologist/physiotherapist and other therapist	51 (5.6)
Administrator	45 (4.9)
Maintenance and technical support	41 (4.5)
Dental nurse or hygienist	13 (1.4)
Other	12 (1.3)
Organization, n (%)	
A	209 (22.7)
B	710 (77.3)
Years of work experience in similar tasks, mean (SD)	11.6 (10.2)
Has promoted the portal to a patient, n (%)	
Yes	391 (42.5)
No	528 (57.5)

Factors Associated With Support for the Portal Implementation

The results of the regression analysis (Table 2) revealed associations between the independent variables and the personnel's support for the patient portal. Thus, Hypotheses 1-3 were supported. The results of the univariate linear regression models (not shown in Table 2) indicated that the personnel's perceptions of their units' implementation practices (β=.12, P<.001), usability of eHealth tools (β=.20, P<.001), and their eHealth-related competence (β=.21, P<.001) were positively associated with their support. These variables together explained 17% of the variation in the support for the portal (Model A). The associations persisted after adjustments for age, gender, work experience, organization, and profession (Model B).

Table 2. Linear regression results showing association between independent variables and health care personnel's support for patient portal.

Variables	Model A		Model B	
	β (standard error)	*P* value	β (standard error)	*P* value
Implementation practices	.07 (0.02)	<.001	.06 (0.02)	.003
Usability	.11 (0.02)	<.001	.10 (0.02)	<.001
eHealth competence	.15 (0.02)	<.001	.15 (0.02)	<.001
Age	—[a]	—	.04 (0.03)	.12
Gender (Category reference: Man)	—	—	–.07 (0.07)	.26
Work experience	—	—	–.04 (0.02)	.13
Organization (Category reference: A)	—	—	–.12 (0.05)	.01
Profession (Category reference: nurse/midwife/public health nurse)	—	—		
Assistant/other nurse			–.01 (0.05)	.79
Ward/department secretary			–.04 (0.09)	.66
Social worker			–.03 (0.09)	.77
Doctor/dentist			.10 (0.10)	.33
Psychologist/physio and other therapist			.09 (0.09)	.30
Maintenance and technical support			–.17 (0.11)	.12
Administrator			.15 (0.08)	.06
Dental nurse or hygienist			.10 (0.14)	.45
Other			–.06 (0.13)	.67
R^2	0.17	—	0.17	—

[a]Not available.

The Association Between Support and Promotion to the Patients

Table 3 presents the results of the logistic regression regarding the association between personnel's support for and promotion of the portal to patients. Hypothesis 4 was supported. The personnel's support was positively associated with their promotion of the portal, and the association persisted after adjusting for age, gender, work experience, organization, and profession. The odds ratio regarding support for the portal was 1.18 after adjustments. This suggests that 1 SD improvement in support was associated with an 18% increase in the likelihood of promoting the portal to patients. Assistant nurses, social workers, doctors, and therapists were less likely to have promoted the portal to patients than were nurses.

Table 3. Logistic regression results showing predictors of promotion of the portal to patients, odds ratios, and 95% CIs.[a]

Predictors	Model A		Model B	
	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
Support for patient portal	1.22 (1.07-1.40)	.004	1.18 (1.02-1.38)	.03
Age	__[b]	—	1.13 (0.94-1.36)	.19
Gender	—	—		
Man			1	—
Woman			1.02 (0.63-1.65)	.95
Work experience	—	—	1.01 (0.84-1.21)	.90
Organization	—	—		
A			1	—
B			0.48 (0.34-0.68)	<.001
Profession	—	—		
Nurse/midwife/public health nurse			1	—
Assistant/other nurse			0.22 (0.15-0.32)	<.001
Ward/department secretary			0.80 (0.43-1.48)	.48
Social worker			0.41 (0.23-0.75)	.004
Doctor/dentist			0.32 (0.16-0.62)	.001
Psychologist/physiotherapist and other therapist			0.31 (0.16-0.61)	.001
Maintenance and technical support			0.35 (0.17-0.73)	.005
Administrator			0.54 (0.27-1.06)	.07
Dental nurse or hygienist			1.39 (0.41-4.71)	.60
Other			0.60 (0.16-2.94)	.44

[a]Continuous variables were used as continuous standardized variables.

[b]Not available.

The Factors Associated With Personnel's eHealth-Related Stress

The regression results (Table 4) revealed the associations between the independent variables and the personnel's eHealth-related stress. Hypotheses 5-7 were supported. The results of the univariate models indicated that the personnel's positive perceptions of their units' implementation practices (β=−.22, *P*<.001), IT usability (β=−.35, *P*<.001), and their eHealth-related competence (β=.21, *P*<.001) were associated with lower levels of eHealth-related stress. These variables explained 23% of the variance in eHealth-related stress (Model A). The associations persisted after adjustments (Model B). Older age was significantly associated with higher eHealth-related stress (*P*=.01).

Table 4. Regression results presenting association between independent variables and personnel's eHealth-related stress.

	Model A		Model B	
	β (standard error)	*P* value	β (standard error)	*P* value
Implementation practices	–.14 (0.03)	<.001	–.14 (0.03)	<.001
Usability	–.27 (0.03)	<.001	–.28 (0.03)	<.001
eHealth competence	–.07 (0.03)	.010	–.06 (0.03)	.02
Age	—[a]	—	.08 (0.03)	.01
Gender (Category reference: Man)	—	—	–.01 (0.08)	.94
Work experience	—	—	.03 (0.03)	.34
Organization (Category reference: A)	—	—	.09 (0.06)	.12
Profession (Category reference: nurse/midwife/public health nurse)	—	—		
Assistant/other nurse			–.02 (0.06)	.71
Ward/department secretary			.08 (0.11)	.48
Social worker			–.21 (0.11)	.07
Doctor/dentist			–.11 (0.10)	.30
Psychologist/physiotherapist and other therapist			–.32 (0.10)	.002
Maintenance and technical support			.06 (0.11)	.59
Administrator			–.15 (0.14)	.29
Dental nurse or hygienist			–.29 (0.17)	.09
Other			–.24 (0.22)	.27
R^2	.22	—	.27	—

[a]Not available.

Discussion

Principal Findings

Successful implementation of patient portals depends on HCP's adoption of the portals in their daily work routines. This survey study among primary HCP set out to explore the factors associated with their support for a patient portal and the adverse effect of the portal on HCP in terms of increased eHealth-related stress.

In this survey study with a representative and rather large sample of responses (n=919) from the primary care personnel, we found support for all of our 7 hypotheses. First, in line with a previous study from the preimplementation phase of a patient portal [15], we found that HCP's perception of the good implementation practices in their unit was associated with their high support for the patient portal. Second, consistent with several previous findings [33-35], an association between good usability and portal support was found.

Third, we found a positive association between HCP's eHealth competence and their support for the patient portal. We are not aware of previous evidence on this association in the eHealth context. However, in the context of quality improvement in health care, Damush and colleagues [49] found that professionals' confidence in their ability to perform behaviors required in the new practice was crucial for their acceptance of the improvement.

Fourth, this study shows an association between HCP's support and their endorsement of the portal to the patients. Previous studies have elaborated on a plethora of factors that inhibit and facilitate professionals' endorsement [10,50]. However, this study is to our knowledge the first to show quantitative evidence of the association between professionals' support for the portal and their endorsement of it to patients.

Fifth, in this study, good usability and high self-rated eHealth competence were found to alleviate eHealth-related stress. This is in line with previous studies showing that usability issues increase stress related to information systems, and that these issues may be alleviated by experience in using the technology [24,25].

Finally, we found that, in addition to good usability and high eHealth competence, HCP's perception of good implementation practices applied in their unit alleviate HCP's eHealth-related stress. To our knowledge, this association has not been examined before in the eHealth context. Similar associations between implementation practices and employee stress concerning new technology have, however, been found in other organizational contexts [51,52].

Strengths and Limitations

This study contributes to the research on patient portals by providing quantitative evidence of the roles of (1) managerial practices in the successful implementation, and (2) HCP's support toward a patient portal. First, studies on good eHealth implementation practices have relied mainly on qualitative

evidence [19]. This study aimed to quantify the influence of these practices. While the association between good implementation practices and HCP's support for eHealth services has been shown before [15], this study suggests that the same practices may also alleviate eHealth-related stress experienced by the personnel. Second, HCP's positive attitude toward the patient portal was measured with items reflecting respondent's personal and perceived collective support for new HIT. This concept of user's positive attitude is not limited to the act of using specific features of the portal but extends to the attitudes toward broader and longer-term changes that the implementation is perceived to entail. We propose that this concept is well-suited to depict HCP's attitudes in a context where the solution is not expected to remain unchanged but rather to be iteratively developed and molded in a social process to serve the purposes shared by the user, co-workers, and supervisors.

The limitations of this study are related to the cross-sectional, single-informant design, and omission of likely influential contextual factors in the regression models. First, this study relied on self-reported measures. This could lead to problems associated with common method variance and the inflation of the strength of relationships. To minimize the problems with self-reports, measures that exhibited good reliability in previous studies were applied. Second, as we wanted to keep the questionnaire at a suitable length, we were not able to include all relevant contextual factors in our analysis. This is a limitation especially in our model predicting professional support, in which relatively low proportion of variance (17%) could be explained. In particular, the survey did not include questions on the perceived usefulness of the patient portal. Although the functionalities in the studied patient portals were similar to those adopted by other Finnish health care organizations and represented well the functionalities that patients consider useful [48,49], it is possible that personnel's perceptions of their utility in patient care varies. Studies show that clinicians' perceptions of HIT utility in terms of improvements in patient care and personnel's work are among the most important factors in their support [32,50].

The data for this study were collected in 2017. In the field of fast developing eHealth, a delay in reporting the findings on a specific technological application may compromise the timeliness of the observations [53]. However, issues related to users' adoption of new IT seem to persist over time in the different contexts of technological applications. For example, in Finland, health care professionals' stress related to information systems has been shown to increase over time regardless of the changes in the applications [44]. We therefore maintain that our findings on the factors that contribute to professionals' support and stress related to eHealth endure over time and can be generalized to adoption and use of patient portals with different features than what was studied here.

Practical Implications

The findings of this study have several implications for health care managers and frontline leaders. In order to build HCP support and alleviate employees' stress related to new HIT, management needs to ensure good usability of eHealth tools, engage personnel in the planning, and provide adequate information on the tools and resources to normalize their use in the daily practices. In the procurement of new HIT, managers are encouraged to acknowledge the proper implementation as a prerequisite for meaningful use. The findings call for careful consideration of the resources needed in the adoption and maintenance phases to balance the investments required and the pace of adoption of new HIT. Experiences beyond the health care industry show that the investments in the implementation and normalization of new HIT are often overlooked in the procurement of new HIT [54].

Conclusions

This survey study suggests that health care organizations' implementation practices and good usability have a twofold impact on meaningful and sustainable use of patient portals: first, health care professionals' stress is relieved, and second, their support for the patient portal increases. Higher support is manifested in professionals' increased endorsement of the portal to the patients.

Acknowledgments

The authors thank the participating organizations and respondents. We are grateful to Head Nurse Ulla Kemppainen for her assistance in distributing the survey. This work was supported by the Strategic Research Council at the Academy of Finland (Grant Nos 327145 and 327147) and NordForsk through the funding to Nordic eHealth for Patients: Benchmarking and Developing for the Future (NORDeHEALTH), project number 100477.

Authors' Contributions

All the authors contributed to the literature review, revisions, and manuscript approval. SK, IH, and PH contributed to the study design, and SK was primarily responsible for the questionnaire development. IH was primarily responsible for drafting the manuscript and performing the statistical analyses, and TH and SK contributed to the interpretation of the analyses.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire.

[DOCX File , 19 KB - jmir_v23i9e28976_app1.docx]

Multimedia Appendix 2
The means and standard deviations of the study variables.
[DOCX File , 15 KB - jmir_v23i9e28976_app2.docx]

References

1. Lin C, Wittevrongel L, Moore L, Beaty BL, Ross SE. An Internet-based patient-provider communication system: randomized controlled trial. J Med Internet Res 2005 Aug 05;7(4):e47 [FREE Full text] [doi: 10.2196/jmir.7.4.e47] [Medline: 16236699]
2. Calvillo J, Román I, Roa LM. How technology is empowering patients? A literature review. Health Expect 2015 Oct;18(5):643-652 [FREE Full text] [doi: 10.1111/hex.12089] [Medline: 23711169]
3. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. How outcomes are achieved through patient portals: a realist review. J Am Med Inform Assoc 2014 Jul;21(4):751-757. [doi: 10.1136/amiajnl-2013-002501] [Medline: 24503882]
4. Urowitz S, Wiljer D, Dupak K, Kuehner Z, Leonard K, Lovrics E, et al. Improving diabetes management with a patient portal: a qualitative study of diabetes self-management portal. J Med Internet Res 2012;14(6):e158 [FREE Full text] [doi: 10.2196/jmir.2265] [Medline: 23195925]
5. Wakefield DS, Mehr D, Keplinger L, Canfield S, Gopidi R, Wakefield BJ, et al. Issues and questions to consider in implementing secure electronic patient-provider web portal communications systems. Int J Med Inform 2010 Jul;79(7):469-477. [doi: 10.1016/j.ijmedinf.2010.04.005] [Medline: 20472495]
6. Hulter P, Pluut B, Leenen-Brinkhuis C, de Mul M, Ahaus K, Weggelaar-Jansen AM. Adopting Patient Portals in Hospitals: Qualitative Study. J Med Internet Res 2020 May 19;22(5):e16921 [FREE Full text] [doi: 10.2196/16921] [Medline: 32427110]
7. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
8. May C, Mort M, Williams T, Mair F, Gask L. Health technology assessment in its local contexts: studies of telehealthcare. Social Science & Medicine 2003 Aug;57(4):697-710. [doi: 10.1016/s0277-9536(02)00419-7]
9. Finch TL, Rapley T, Girling M, Mair FS, Murray E, Treweek S, et al. Improving the normalization of complex interventions: measure development based on normalization process theory (NoMAD): study protocol. Implement Sci 2013;8:43 [FREE Full text] [doi: 10.1186/1748-5908-8-43] [Medline: 23578304]
10. Irizarry T, DeVito Dabbs A, Curran CR. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
11. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
12. Elwyn G, Rix A, Holt T, Jones D. Why do clinicians not refer patients to online decision support tools? Interviews with front line clinics in the NHS. BMJ Open 2012;2(6):e001530 [FREE Full text] [doi: 10.1136/bmjopen-2012-001530] [Medline: 23204075]
13. Fix GM, Hogan TP, Amante DJ, McInnes DK, Nazi KM, Simon SR. Encouraging Patient Portal Use in the Patient-Centered Medical Home: Three Stakeholder Perspectives. J Med Internet Res 2016 Nov 22;18(11):e308 [FREE Full text] [doi: 10.2196/jmir.6488] [Medline: 27876686]
14. Agarwal P, Kithulegoda N, Bouck Z, Bosiak B, Birnbaum I, Reddeman L, et al. Feasibility of an Electronic Health Tool to Promote Physical Activity in Primary Care: Pilot Cluster Randomized Controlled Trial. J Med Internet Res 2020 Feb 14;22(2):e15424 [FREE Full text] [doi: 10.2196/15424] [Medline: 32130122]
15. Kujala S, Hörhammer I, Kaipio J, Heponiemi T. Health professionals' expectations of a national patient portal for self-management. Int J Med Inform 2018 Sep;117:82-87 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.06.005] [Medline: 30032968]
16. Wald JS. Variations in patient portal adoption in four primary care practices. AMIA Annu Symp Proc 2010;2010:837-841 [FREE Full text] [Medline: 21347096]
17. Hoogenbosch B, Postma J, de Man-van Ginkel JM, Tiemessen NA, van Delden JJ, van Os-Medendorp H. Use and the Users of a Patient Portal: Cross-Sectional Study. J Med Internet Res 2018 Sep 17;20(9):e262 [FREE Full text] [doi: 10.2196/jmir.9418] [Medline: 30224334]
18. Cresswell K, Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. Int J Med Inform 2013 May;82(5):e73-e86. [doi: 10.1016/j.ijmedinf.2012.10.007] [Medline: 23146626]

19. Ingebrigtsen T, Georgiou A, Clay-Williams R, Magrabi F, Hordern A, Prgomet M, et al. The impact of clinical leadership on health information technology adoption: systematic review. Int J Med Inform 2014 Jun;83(6):393-405. [doi: 10.1016/j.ijmedinf.2014.02.005] [Medline: 24656180]
20. Ludwick DA, Doucette J. Adopting electronic medical records in primary care: lessons learned from health information systems implementation experience in seven countries. Int J Med Inform 2009 Jan;78(1):22-31. [doi: 10.1016/j.ijmedinf.2008.06.005] [Medline: 18644745]
21. Esmaeilzadeh P, Sambasivan M. Health Information Exchange (HIE): A literature review, assimilation pattern and a proposed classification for a new policy approach. J Biomed Inform 2016 Dec;64:74-86 [FREE Full text] [doi: 10.1016/j.jbi.2016.09.011] [Medline: 27645322]
22. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert. What do we know about developing patient portals? a systematic literature review. J Am Med Inform Assoc 2016 Apr;23(e1):e162-e168. [doi: 10.1093/jamia/ocv114] [Medline: 26335985]
23. Ratwani R, Fairbanks T, Savage E, Adams K, Wittie M, Boone E, et al. Mind the Gap. A systematic review to identify usability and safety challenges and practices during electronic health record implementation. Appl Clin Inform 2016 Nov 16;7(4):1069-1087 [FREE Full text] [doi: 10.4338/ACI-2016-06-R-0105] [Medline: 27847961]
24. Heponiemi T, Hyppönen H, Vehko T, Kujala S, Aalto A, Vänskä J, et al. Finnish physicians' stress related to information systems keeps increasing: a longitudinal three-wave survey study. BMC Med Inform Decis Mak 2017 Oct 17;17(1):147 [FREE Full text] [doi: 10.1186/s12911-017-0545-y] [Medline: 29041971]
25. Lazarus R, Folkman S. Stress, Appraisal, and Coping. Berlin, Germany: Springer; 1984.
26. Babbott S, Manwell LB, Brown R, Montague E, Williams E, Schwartz M, et al. Electronic medical records and physician stress in primary care: results from the MEMO Study. J Am Med Inform Assoc 2014 Feb;21(e1):e100-e106 [FREE Full text] [doi: 10.1136/amiajnl-2013-001875] [Medline: 24005796]
27. Heponiemi T, Kujala S, Vainiomäki S, Vehko T, Lääveri T, Vänskä J, et al. Usability Factors Associated With Physicians' Distress and Information System-Related Stress: Cross-Sectional Survey. JMIR Med Inform 2019 Nov 05;7(4):e13466 [FREE Full text] [doi: 10.2196/13466] [Medline: 31687938]
28. Vainiomäki S, Heponiemi T, Vänskä J, Hyppönen H. Tailoring EHRs for Specific Working Environments Improves Work Well-Being of Physicians. Int J Environ Res Public Health 2020 Jun 30;17(13):4715 [FREE Full text] [doi: 10.3390/ijerph17134715] [Medline: 32630043]
29. Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Qual Saf Health Care 2010 Oct;19 Suppl 3:i68-i74 [FREE Full text] [doi: 10.1136/qshc.2010.042085] [Medline: 20959322]
30. Greenhalgh T, Stramer K, Bratan T, Byrne E, Russell J, Potts H. Adoption and non-adoption of a shared electronic summary record in England: a mixed-method case study. BMJ 2010 Jun 16;340:c3111. [doi: 10.1136/bmj.c3111] [Medline: 20554687]
31. Beglaryan M, Petrosyan V, Bunker E. Development of a tripolar model of technology acceptance: Hospital-based physicians' perspective on EHR. Int J Med Inform 2017 Dec;102:50-61. [doi: 10.1016/j.ijmedinf.2017.02.013] [Medline: 28495348]
32. Flynn D, Gregory P, Makki H, Gabbay M. Expectations and experiences of eHealth in primary care: a qualitative practice-based investigation. Int J Med Inform 2009 Sep;78(9):588-604. [doi: 10.1016/j.ijmedinf.2009.03.008] [Medline: 19482542]
33. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
34. Gagnon M, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012 Feb;36(1):241-277 [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
35. Chen R, Hsiao J. An investigation on physicians' acceptance of hospital information systems: a case study. Int J Med Inform 2012 Dec;81(12):810-820. [doi: 10.1016/j.ijmedinf.2012.05.003] [Medline: 22652011]
36. Yu P, Li H, Gagnon M. Health IT acceptance factors in long-term care facilities: a cross-sectional survey. Int J Med Inform 2009 Apr;78(4):219-229. [doi: 10.1016/j.ijmedinf.2008.07.006] [Medline: 18768345]
37. Kujala S, Rajalahti E, Heponiemi T, Hilama P. Health Professionals' Expanding eHealth Competences for Supporting Patients' Self-Management. Stud Health Technol Inform 2018;247:181-185 [FREE Full text] [Medline: 29677947]
38. Viitanen J, Hyppönen H, Lääveri T, Vänskä J, Reponen J, Winblad I. National questionnaire study on clinical ICT systems proofs: physicians suffer from poor usability. Int J Med Inform 2011 Oct;80(10):708-725. [doi: 10.1016/j.ijmedinf.2011.06.010] [Medline: 21784701]
39. Kujala S, Hörhammer I, Heponiemi T, Josefsson K. The Role of Frontline Leaders in Building Health Professional Support for a New Patient Portal: Survey Study. J Med Internet Res 2019 Mar 22;21(3):e11413 [FREE Full text] [doi: 10.2196/11413] [Medline: 30901003]
40. Wakefield D, Halbesleben J, Ward M, Qiu Q, Brokel J, Crandall D. Development of a measure of clinical information systems expectations and experiences. Med Care 2007 Sep;45(9):884-890. [doi: 10.1097/MLR.0b013e3180653625] [Medline: 17712259]

41. Kirkendall ES, Goldenhar LM, Simon JL, Wheeler DS, Spooner AS. Transitioning from a computerized provider order entry and paper documentation system to an electronic health record: expectations and experiences of hospital staff. Int J Med Inform 2013 Nov;82(11):1037-1045. [doi: 10.1016/j.ijmedinf.2013.08.005] [Medline: 24041453]
42. Finstad K. The Usability Metric for User Experience. Interacting with Computers 2010 Sep;22(5):323-327. [doi: 10.1016/j.intcom.2010.04.004]
43. Heponiemi T, Aalto A, Pekkarinen L, Siuvatti E, Elovainio M. Are there high-risk groups among physicians that are more vulnerable to on-call work? Am J Emerg Med 2015 May;33(5):614-619. [doi: 10.1016/j.ajem.2015.01.034] [Medline: 25680563]
44. Heponiemi T, Hyppönen H, Kujala S, Aalto A, Vehko T, Vänskä J, et al. Predictors of physicians' stress related to information systems: a nine-year follow-up survey study. BMC Health Serv Res 2018 Dec 13;18(1):284 [FREE Full text] [doi: 10.1186/s12913-018-3094-x] [Medline: 29653530]
45. Kuusio H, Heponiemi T, Aalto A, Sinervo T, Elovainio M. Differences in well-being between GPs, medical specialists, and private physicians: the role of psychosocial factors. Health Serv Res 2012 Feb;47(1 Pt 1):68-85. [doi: 10.1111/j.1475-6773.2011.01313.x] [Medline: 22091688]
46. Fornell C, Larcker DF. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. Journal of Marketing Research 1981 Feb;18(1):39. [doi: 10.2307/3151312]
47. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. J Pers Assess 2003 Feb;80(1):99-103. [doi: 10.1207/S15327752JPA8001_18] [Medline: 12584072]
48. O'brien RM. A Caution Regarding Rules of Thumb for Variance Inflation Factors. Qual Quant 2007 Mar 13;41(5):673-690. [doi: 10.1007/s11135-006-9018-6]
49. Damush T, Penney L, Miech E, Rattray N, Baird S, Cheatham A, et al. Acceptability of a complex team-based quality improvement intervention for transient ischemic attack: a mixed-methods study. BMC Health Serv Res 2021 May 12;21(1):453 [FREE Full text] [doi: 10.1186/s12913-021-06318-2] [Medline: 33980224]
50. Nazi KM. The personal health record paradox: health care professionals' perspectives and the information ecology of personal health record systems in organizational and clinical settings. J Med Internet Res 2013 Apr;15(4):e70 [FREE Full text] [doi: 10.2196/jmir.2443] [Medline: 23557596]
51. Korunka C, Weiss A, Karetta B. Effects of new technologies with special regard for the implementation process per se. J. Organiz. Behav 1993 Jul;14(4):331-348. [doi: 10.1002/job.4030140405]
52. Chen S, Westman M, Eden D. Impact of enhanced resources on anticipatory stress and adjustment to new information technology: a field-experimental test of conservation of resources theory. J Occup Health Psychol 2009 Jul;14(3):219-230. [doi: 10.1037/a0015282] [Medline: 19586218]
53. Baker TB, Gustafson DH, Shah D. How can research keep up with eHealth? Ten strategies for increasing the timeliness and usefulness of eHealth research. J Med Internet Res 2014 Feb;16(2):e36 [FREE Full text] [doi: 10.2196/jmir.2925] [Medline: 24554442]
54. Ryan S, Harrison D. Considering Social Subsystem Costs and Benefits in Information Technology Investment Decisions: A View from the Field on Anticipated Payoffs. Journal of Management Information Systems 2015 Dec 02;16(4):11-40. [doi: 10.1080/07421222.2000.11518264]

Abbreviations

HCP: health care personnel
HIT: health information technology
IT: information technology
UMUX: Usability Metric for User Experience

Edited by R Kukafka; submitted 20.03.21; peer-reviewed by L Kremer, F Ghezelbash, F Ghezelbash; comments to author 30.05.21; revised version received 18.06.21; accepted 05.07.21; published 22.09.21.

Please cite as:
Hörhammer I, Kujala S, Hilama P, Heponiemi T
Building Primary Health Care Personnel's Support for a Patient Portal While Alleviating eHealth-Related Stress: Survey Study
J Med Internet Res 2021;23(9):e28976
URL: https://www.jmir.org/2021/9/e28976
doi:10.2196/28976
PMID:34550087

Original Paper

Exploring Test-Retest Reliability and Longitudinal Stability of Digital Biomarkers for Parkinson Disease in the m-Power Data Set: Cohort Study

Mehran Sahandi Far[1,2], MA; Simon B Eickhoff[1,2], Prof Dr; Maria Goni[1,2*], PhD; Juergen Dukart[1,2*], PhD

[1]Institute of Neuroscience and Medicine, Brain & Behaviour (INM-7), Research Centre Jülich, Jülich, Germany

[2]Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

[*]these authors contributed equally

Corresponding Author:
Juergen Dukart, PhD
Institute of Neuroscience and Medicine, Brain & Behaviour (INM-7)
Research Centre Jülich
Wilhelm-Johnen-Strasse
Jülich, 52425
Germany
Phone: 49 1632874330
Fax: 49 2461611880
Email: juergen.dukart@gmail.com

Abstract

Background: Digital biomarkers (DB), as captured using sensors embedded in modern smart devices, are a promising technology for home-based sign and symptom monitoring in Parkinson disease (PD).

Objective: Despite extensive application in recent studies, test-retest reliability and longitudinal stability of DB have not been well addressed in this context. We utilized the large-scale m-Power data set to establish the test-retest reliability and longitudinal stability of gait, balance, voice, and tapping tasks in an unsupervised and self-administered daily life setting in patients with PD and healthy controls (HC).

Methods: Intraclass correlation coefficients were computed to estimate the test-retest reliability of features that also differentiate between patients with PD and healthy volunteers. In addition, we tested for longitudinal stability of DB measures in PD and HC, as well as for their sensitivity to PD medication effects.

Results: Among the features differing between PD and HC, only a few tapping and voice features had good to excellent test-retest reliabilities and medium to large effect sizes. All other features performed poorly in this respect. Only a few features were sensitive to medication effects. The longitudinal analyses revealed significant alterations over time across a variety of features and in particular for the tapping task.

Conclusions: These results indicate the need for further development of more standardized, sensitive, and reliable DB for application in self-administered remote studies in patients with PD. Motivational, learning, and other confounders may cause variations in performance that need to be considered in DB longitudinal applications.

(J Med Internet Res 2021;23(9):e26608) doi:10.2196/26608

KEYWORDS

health sciences; medical research; biomarkers; diagnostic markers; neurological disorders; Parkinson disease; mobile phone

Introduction

Parkinson disease (PD) is primarily characterized by motor signs and symptoms, including tremor at rest, rigidity, akinesia, and postural instability [1]. Although standard in-clinic assessments such as the Unified Parkinson's Disease Rating Scale (UPDRS) are popular, they are influenced by interrater variability by relying on self-reporting by patients and caregivers or clinicians' judgement [2]. In addition, they are costly and limited with respect to observation frequency.

The emergence of new technologies has led to a variety of sensors (ie, acceleration, gyroscope, GPS, etc) embedded in

smart devices for daily use (ie, smartphone, smartwatch). Such sensor data, alongside other digital information recorded passively or when executing prespecified tasks, may provide valuable insight into health-related information. Such applications are now commonly referred to as digital biomarkers (DB) [3-5]. DB being collected frequently over a long period of time can provide an objective, ecologically valid, and more detailed understanding of the inter- and intra-individual variability in disease manifestation in daily life.

Numerous DB have been proposed for PD diagnosis as well as for assessing agreement between clinical rating scales such as UPDRS and sensor-driven data to quantify disease severity or intervention effects [4,6-9]. Despite these various proof of concept studies, many technical challenges with respect to DB deployment remain unaddressed. DB measures are prone to large variation caused by technical and procedural differences, including but not limited to placement/orientation, recording frequency of the devices, and environmental and individual variation (ie, due to motivation, medication, or other aspects) [10-12]. Other factors such as the effect of users' familiarity with technology and the impact of learning on the performance of measured DB in remote and self-administered PD assessment are other important sources of variation that have not been addressed so far. All of these factors may limit the sensitivity and reliability of DB measurements for any of the above PD clinical applications. DB longitudinal variation is therefore an important attribute that should be quantified and addressed. The reliability of DB assessment has been broadly studied for gait, balance, voice, and tapping data [13-18]. However, the existing studies typically focused on a single or a few aspects of PD, and most of them established the test-retest reliability in a standardized clinical setting, limiting the translatability of their findings to at-home applications. Among the studies that evaluated DB assessments for remote monitoring of PD, only one reported the test-retest reliability [4]. No PD studies systematically evaluated the test-retest reliability and longitudinal sensitivity of DB in a fully unsupervised and self-administered PD longitudinal setting.

Although various factors such as medication, disease severity, learning effects, bias from self-reporting, inconsistent disease severity, motivational impacts, and design protocols in self-administered studies can affect the long-term stability of DB, little attention has been paid to evaluating the reliability and longitudinal stability of DB in loosely controlled self-administered settings in daily life. Here, we aimed to address these open questions by assessing the test-retest reliability and longitudinal stability of gait, balance, speech, and tapping tasks in patients with PD and a control cohort consisting of healthy volunteers (HC) in an unsupervised and self-administered daily life setting using the large-scale m-Power data set [19].

Methods

Study Cohort

To address the open questions on the performance of DB measures in PD when collected in a self-administered setting in daily life, we first performed a comprehensive literature search identifying 773 DB features reported in previous studies to cover PD-related alterations in gait characteristics, tremor, postural instability, voice, and finger dexterity. We evaluated the longitudinal stability and test-retest reliability of these features as collected using 4 commonly applied PD tasks (gait, balance, voice, and tapping) in daily life using smartphone in a large cohort of self-reported patients with PD and healthy controls, the m-Power study [19-22]. In addition, we evaluated their sensitivity to learning and medication effects.

Enrolment in the m-Power study was open to adult participants who own an iPhone, are living in the United States, and are comfortable enough with English to read the instructions in the app. Participants were asked to download the app and complete a one-time demographic survey during registration. Demographic data include but are not limited to age, sex, health history, and previous PD clinical diagnosis. They also were asked to fill out a survey with selected questions from the UPDRS Section I (nonmotor experience) and Section II (motor experience), as well as the Parkinson's Disease Questionnaire (PDQ-8). All the participants were suggested to complete each task (walking, tapping, voice, and memory) up to 3 times a day for up to 6 months. In addition, self-reported patients with PD were asked to complete the task before medication, after medication, and at another time when they were feeling at their best.

Ethical oversight of the m-Power study was obtained from the Western Institutional Review Board. Prior to signing an electronically rendered traditional informed consent form, prospective participants had to pass a 5-question quiz evaluating their understanding of the study aims, participant rights, and data sharing options. After completing the e-consent process and electronically signing the informed consent form, participants were asked for an email address to which their signed consent form was sent and allowing for verification of their enrolment in the study. Participants were given the option to share their data only with the m-Power study team and partners ("share narrowly") or to share their data more broadly with qualified researchers worldwide, and they had to make an active choice to complete the consent process (no default choice was presented). The data used in our study consist of all individuals who chose to have their data shared broadly.

Data Preprocessing

The m-Power data set is assessed outside of a clinical environment with limited quality control and supervision. All information, including the health history, disease diagnosis, duration, treatment, and survey outcomes, are self-reported. To address these, we excluded participants who did not specify their age, sex, and information on professional diagnosis (if they belong to the PD or HC group) and those with empty, null, or corrupted files. The participants are assigned to the PD or HC group according to their response to the question "Have you been diagnosed by a medical professional with Parkinson disease?" There was a significant difference in the age and sex distribution between HC and PD groups. Particularly, age slanted toward younger and male individuals in HC. To reduce the impact of age, we restricted the age range for our analysis to between 35 and 75 years. The demographic details are

provided in Table 1, and the overall overview of preprocessing steps is displayed in Figure 1A.

Figure 1. Overview of statistical analyses and the preprocessing scheme. (A) Flowchart of preprocessing steps. (B) Flowchart of statistical analyses. (C) Flowchart of number of features at each selection step. HC: healthy controls; ICC: intraclass correlation coefficients; PD: Parkinson disease; rm-ANOVA: repeated-measures analysis of variance.

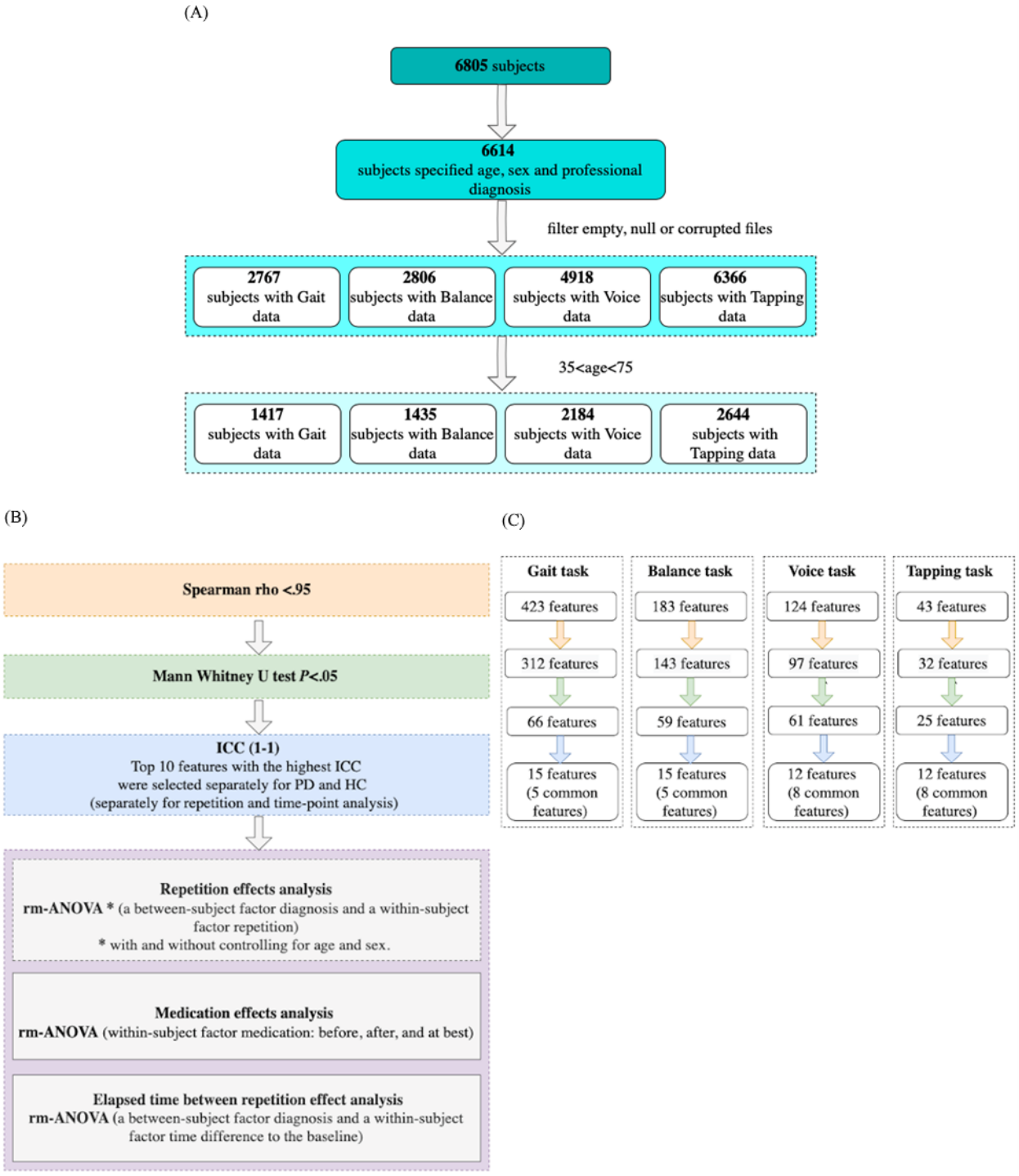


Table 1. Characteristics of study cohorts after data cleaning.

Characteristic	Gait		Balance		Voice		Tapping	
	HC[a]	PD[b]	HC	PD	HC	PD	HC	PD
Sex,[c] n								
Male	655	399	668	401	1042	571	1370	630
Female	152	211	155	211	249	322	304	340
Age (years),[c] mean (SD)	49 (10.60)	60.3 (8.90)	48.9 (10.70)	60.3 (8.90)	47.7 (10.40)	60.1 (9)	46.9 (10.1)	59.9 (9)
UPDRS,[d] mean (SD)	N/A[e]	12.60 (7.11)	N/A	12.53 (7.07)	N/A	12.58 (7.70)	N/A	12.54 (7.73)
UPDRS I, mean (SD)	N/A	4.90 (3.12)	N/A	4.9 (3.11)	N/A	4.93 (3.25)	N/A	4.95 (3.27)
UPDRS II, mean (SD)	N/A	7.76 (5.41)	N/A	7.7 (5.40)	N/A	7.61 (5.70)	N/A	7.56 (5.70)
PDQ-8,[f] mean (SD)	N/A	5.13 (4.72)	N/A	7.07 (4.70)	N/A	5.28 (5.01)	N/A	5.3 (4.96)

[a]HC: healthy controls.

[b]PD: Parkinson disease.

[c]$P<.001$ (two-sample, two-tailed t test for age and chi-square test for sex with 95% confidence) for all tasks.

[d]UPDRS: Unified Parkinson's Disease Rating Scale.

[e]N/A: not applicable.

[f]PDQ: Parkinson's Disease Questionnaire.

Feature Extraction

To identify features that are commonly used for the walking, voice, and tapping tasks for PD applications, we performed a comprehensive literature search in PubMed with the following terms: ((Parkinson's disease) AND (walking OR gait OR balance OR voice OR tapping) AND (wearables OR smartphones)). Based on this search, we identified a total of 773 features related to gait (N=423), balance (N=183), finger dexterity (N=43), and speech impairment (N=124). All of these features were computed for the m-Power study [23]. A detailed explanation of the extracted features, including the respective references, is provided in Tables S1-S4 in Multimedia Appendix 1. For features sharing the same variance (high pairwise correlation: Spearman $\rho>0.95$), only one of the features was selected randomly for further analyses to reduce the amount of redundant information for each task. Figure 1C summarizes the feature extraction process and the number of features at each selection step.

Gait and Balance

Impairments in gait speed, stride length, and stride time variability are common changes that are linked to PD [24-27]. Instability in postural balance is also considered to be one of the well-reported characteristics associated with PD [15,28-30]. Both were assessed by a walking task. The gait part consisted of 20 steps walking in a straight line, followed by the balance part of a 30-second stay still period. Given a heterogeneity of gait signal lengths across participants, we used a fixed length signal of 10 seconds and selected data from participants who met this criterion, which resulted in 28,150 records from 1417 unique participants. In addition to the accelerometer signals (x, y, and z), their average, the step series, position along the three axes by double integration, and velocity and acceleration along the path were used for feature extraction [31,32] (Table S1 in Multimedia Appendix 1). For balance, we used a 15-second time window, trimming the first 5 and the last 10 seconds of the 30-second records to reduce the noise due to the between-task transition period, resulting in 29,050 records from 1435 unique participants. Feature extraction covered signals related to tremor acceleration predicted to fall in the 4-7 Hz band and postural acceleration (nontremor) falling in the 0-3.5 Hz band [33] (Table S2 in Multimedia Appendix 1).

Voice

PD may also affect breathing and results in alterations in speech and voice. Reduced volume, hoarse quality, and vocal tremor are commonly reported for PD using voice analysis [16,34,35]. In this task, participants said "aaaah" for about 10 seconds. For voice, 49,676 records were selected, belonging to 2184 unique participants. Voice features were computed from fundamental frequency, amplitude, and period signals, trimming the first and the last 2 seconds of the 10-second interval (Table S3 in Multimedia Appendix 1).

Tapping

Impairment in finger dexterity is another sign associated with PD [36,37]. In the m-Power study, participants were asked to tap as fast as possible for 20 seconds with the index and middle fingers on the screen of their phone (positioned on a flat surface). Screen pixel coordinate (x, y) and timestamp of taped points plus acceleration sensor data were collected for this task. Overall, 55,894 recordings were selected, belonging to 2644 unique participants. Features were computed based on the intertapping distance and interval (Table S4 in Multimedia Appendix 1).

Statistical Analysis

For features to be considered usable for biomarker purposes in longitudinal studies, several criteria are important, including sensitivity to disease signs and symptoms, good test-retest reliability, and robustness against the effects of learning and

other longitudinal confounders. To address these criteria, we adopted a stepwise statistical procedure (see Figure 1B for a summary of statistical analyses).

As DB measures are frequently not normally distributed, Mann-Whitney *U* tests were used to identify all features that significantly differ between PD and HC at the first administration (baseline) ($P<.05$). Effect sizes (Cohen *d*) were computed for these features to provide an estimate of the magnitude of differentiation between PD and HC.

Next, intraclass correlation coefficients (ICC, type 1-1) were used to determine the test-retest reliability of features showing a significant differentiation between PD and HC. We used ICC type 1-1 in our study because individuals were not tested under the same conditions (ie, same device), and reliability was determined from a single measurement. ICC values of 0-0.40 were considered to be poor, 0.40-0.59 to be fair, 0.60-0.74 to be good, and 0.75-1.00 to be excellent [38]. To assess the reliability of each feature, ICC values were computed for different time points versus baseline (one hour [0-6 hours], one day [calendric day], one week [7 calendric days], or one month apart [30 calendric days]), as well as for different repeats versus baseline (baseline vs second, third, fourth, and fifth repeat). We then focused our analyses on the top 10 features (as they provide a representative subset of the best performing features) with the highest median ICC values for each group (PD, HC) and tested for their longitudinal stability over time. Results for all features are reported in Multimedia Appendix 1. Features from the PD group are further referred to as "PD features," those from the HC group only as "HC features," and overlapping features from both groups as "common features." We computed repeated-measures analyses of variance (rm-ANOVA) using a mixed factorial design with a between-subject factor diagnosis and a within-subject factor repetition (first, second, third, fourth, and fifth) including their interaction (Equation S1 in Multimedia Appendix 1). Participants who had at least 4 repetitions after baseline (463 for gait, 597 for balance, 1085 for voice, and 1333 for tapping) were included in these analyses. To assess the effects of age and sex on the longitudinal stability of the most reliable features, we repeated all analyses while controlling for age and sex as covariates (Equation S2 in Multimedia Appendix 1). Also, we assessed the impact of elapsed time between repetitions by computing rm-ANOVA using a mixed factorial design with a between-subject factor diagnosis and a within-subject factor elapsed time (calculated as a time difference of each repetition from the baseline in hours) and controlling for age and sex (Equation S3 in Multimedia Appendix 1).

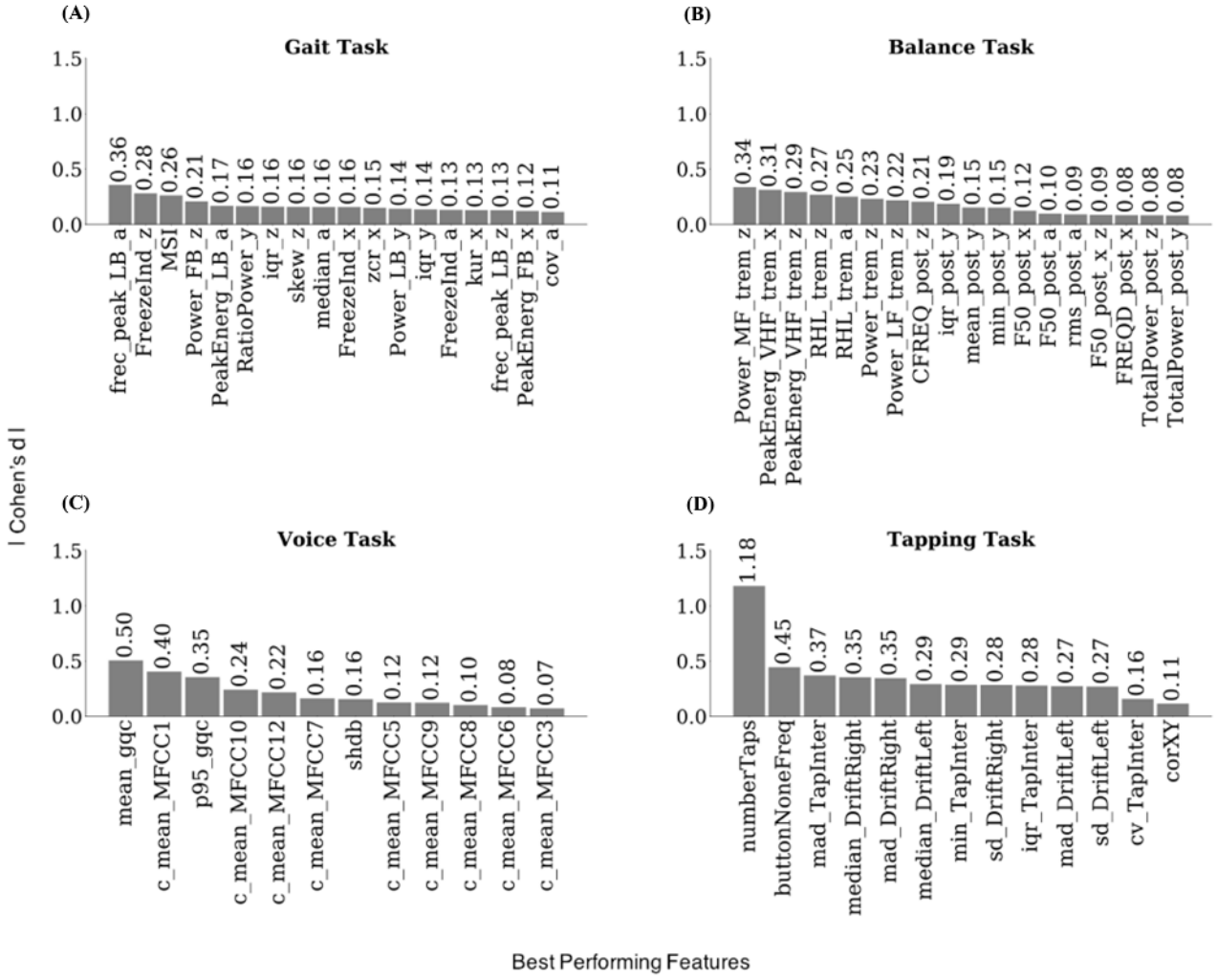
Lastly, we assessed the impact of PD medication by computing rm-ANOVA in the PD group with the within-subject factor medication (ie, before, after, and at best) (Equation S4 in Multimedia Appendix 1). Participants with PD who had at least one marked task for each of the 3 PD medication conditions (ie, before, after, and at best) were included in treatment effect analysis (188 for gait, 189 for balance, 280 for voice, and 338 for tapping).

Results

Differentiation Between PD and HC

First, we aimed to restrict the test-retest reliability analyses of the initial 773 features to those which significantly differ between PD (N=610 to 970 depending on the task, Table 1) and HC (N=807 to 1674). For this, we performed group comparisons for all computed features for gait, balance, voice, and tapping tasks. Overall, 66 out of 423 gait, 59 out of 183 balance, 60 out of 124 voice, and 25 out of 43 tapping features differed significantly (all Ps<.05) between PD and HC at baseline (Figure 1C) with small (gait and balance) to medium effect sizes for gait, balance, and voice and small to large effect sizes for the tapping task (Figure 2 and Tables S5-S8 in Multimedia Appendix 1).

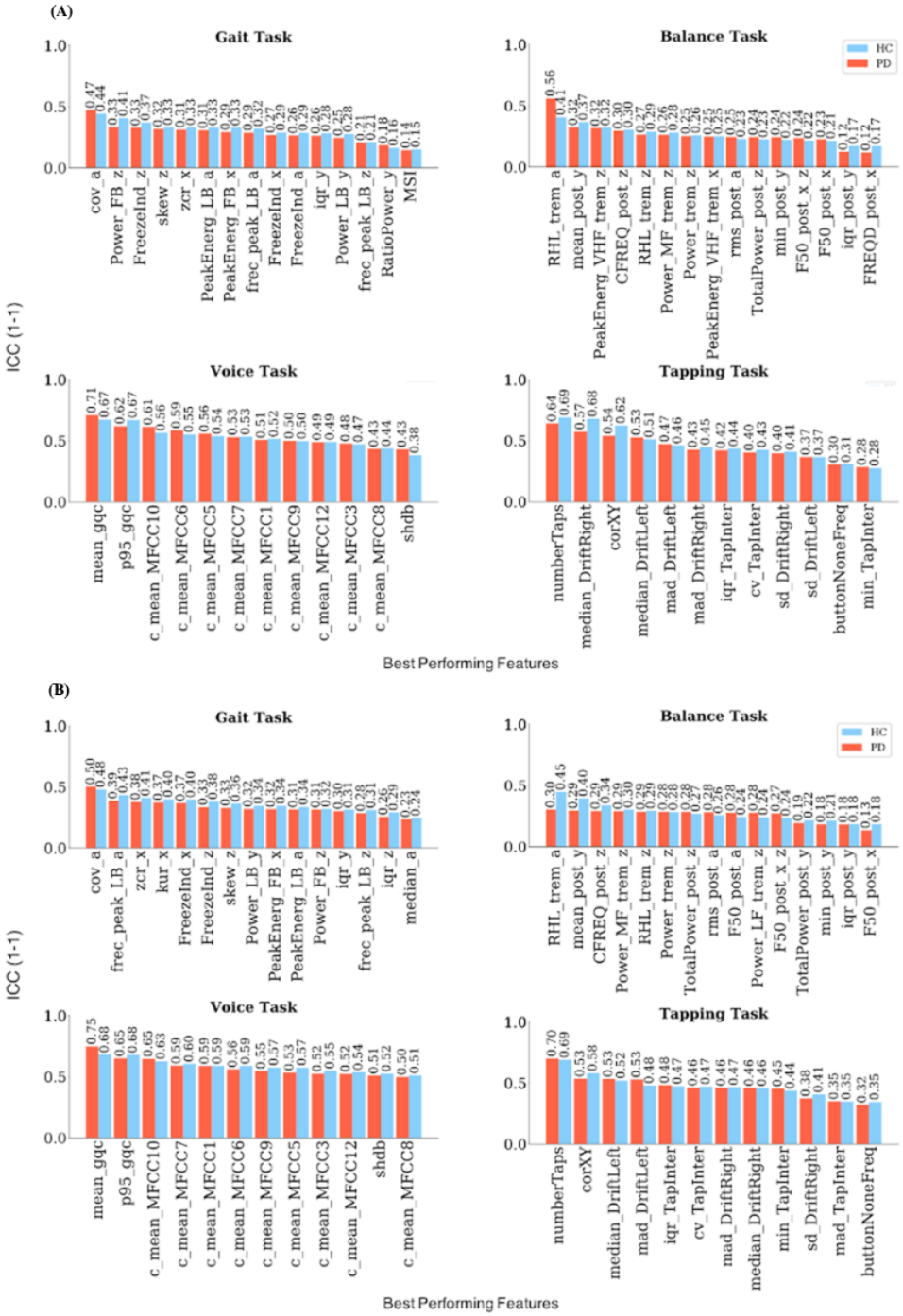
Figure 2. Effect size (Cohen d) for the most reliable features in the Parkinson disease and healthy control groups selected from different time points and repetitions. a: accelerometer average signal; iqr: interquartile range; min: minimum value; PeakEnerg: peak of energy; x: accelerometer mediolateral signal; y: accelerometer vertical signal; z: accelerometer anteroposterior signal. (A) Gait task. cov: coefficient of variation; FB: freezing band; frec_peak: frequency at the peak of energy; FreezeInd: freeze index; kur: kurtosis; LB: locomotor band; MSI: mean stride interval; RatioPower: sum of the power in the freezing and locomotor band; skew: skewness; zcr: zero-crossing rate. (B) Balance task. buttonNoneFreq: frequency of tapping outside the button; CFREQ: centroidal frequency; F50: frequency containing 50% of total power; FRQD: frequency of dispersion of the power spectrum; HF: high frequency (>4 Hz); LF: low frequency (0.15-3.5 Hz); MF: medium frequency (4-7 Hz); post: postural; Power: energy between 3.5-15 Hz; RHL: ratio between power in high frequency and low frequency; rms: root mean square; TotalPower: energy between 15-3.5 Hz; trem: tremor; VHF: very high frequency (>7 Hz). (C) Voice task. c_mean: mean of the MFCC; gqc: glottis quotient close; log: energy of the signal and the first and second derivatives of the MFCC; MFCC: Mel-frequency cepstral coefficients; p95: 95th percentile; shbd: shimmer. (D) Tapping task. corXY: correlation of X and Y positions; cv: coefficient; DriftLeft: left drift; DriftRight: right drift; mad: median absolute deviation; numberTaps: number of taps; sd: standard deviation; TapInter: tap interval.



Test-Retest Reliability

Next, we identified the top 10 features with highest median test-retest reliability (as measured using ICC) separately for PD and HC across different time points (one hour, one day, one week, or one month apart) and repetitions (all participants with 5 repetitions of the task) (Tables S5-S8 in Multimedia Appendix 1, Figure 1B). This procedure resulted in 12 to 15 features (including shared ones) being selected for each task (Figure 3, Figures S1 and S2 in Multimedia Appendix 1). ICC analyses revealed poor to good test-retest reliability for these most reliable features from the gait and balance tasks and good to excellent reliability for features from voice and tapping tasks (Figure 3). The average ICC across the best performing features selected from different repetitions was lower at the fifth repetition compared to the first; it dropped from 0.11 to 0.09 for gait, from 0.21 to 0.13 for balance, from 0.39 to 0.24 for voice, and from 0.3 to 0.23 for tapping. The average ICC across the best performing features selected from different time points was also lower at one month compared to one hour apart, decreasing from 0.13 to 0.07 for gait, from 0.2 to 0.12 for balance, from 0.33 to 0.26 for voice, and from 0.32 to 0.19 for tapping.

Figure 3. Median ICC values for the most reliable features in the Parkinson disease and healthy control groups. a: accelerometer average signal; ICC: intraclass correlation coefficient; iqr: interquartile range; min: minimum value; PeakEnerg: peak of energy; x: accelerometer mediolateral signal; y: accelerometer vertical signal; z: accelerometer anteroposterior signal. (A) Median ICC values across different time points for the best performing features. (B) Median ICC values across different repetitions for the best performing features. Gait task—cov: coefficient of variation; FB: freezing band; frec_peak: frequency at the peak of energy; FreezeInd: freeze index; kur: kurtosis; LB: locomotor band; MSI: mean stride interval; RatioPower: sum of the power in the freezing and locomotor band; skew: skewness; zcr: zero-crossing rate. Balance task—buttonNoneFreq: frequency of tapping outside the button; CFREQ: centroidal frequency; F50: frequency containing 50% of total power; FRQD: frequency of dispersion of the power spectrum; HF: high frequency (>4 Hz); LF: low frequency (0.15-3.5 Hz); MF: medium frequency (4-7 Hz); post: postural; Power: energy between 3.5-15 Hz; RHL: ratio between power in high frequency and low frequency; rms: root mean square; TotalPower: energy between 15-3.5 Hz; trem: tremor; VHF: very high frequency (>7 Hz). Voice task—c_mean: mean of the MFCC; gqc: glottis quotient close; log: energy of the signal and the first and second derivatives of the MFCC; MFCC: Mel-frequency cepstral coefficients; p95: 95th percentile; shbd: shimmer. Tapping task—corXY: correlation of X and Y positions; cv: coefficient; DriftLeft: left drift; DriftRight: right drift; mad: median absolute deviation; numberTaps: number of taps; sd: standard deviation; TapInter: tap interval.



Repetition Effects

Next, we evaluated the longitudinal stability of these most reliable features. Using rm-ANOVA, we tested for the main effects of diagnosis, repetition (first, second, third, fourth, and fifth), and their interaction (Figures 4 and 5, Tables S9 and S10-S13 in Multimedia Appendix 1). A significant main effect of diagnosis across all time points was observed for 6 out of 15 gait features, 11 out of 15 balance features, 8 out of 12 voice features, and 11 out of 12 tapping features. A significant effect of repetition was found for 8 out of 15 gait features, 8 out of 15 balance features, 4 out of 12 voice features, and 10 out of 12 tapping features. A significant diagnosis-by-repetition interaction effect was identified for 3 out of 15 gait features, 0 out of 15 balance features, 3 out of 12 voice features, and 9 out of 12 tapping features. Further, we tested for the main effects of the elapsed time between repetitions and its interaction with diagnosis (Tables S18-S21 in Multimedia Appendix 1). A significant main effect of elapsed time was observed for 1 out of 15 gait features, 2 out of 15 balance features, 5 out of 12 voice features, and 5 out of 12 tapping features. A significant diagnosis-by-time interaction effect was observed only in 1 out of 15 balance features and 3 out of 12 tapping features.

Figure 4. Mean value of the best performing baseline features across different time points, calculated for PD and HC separately. a: accelerometer average signal; HC: healthy controls; iqr: interquartile range; min: minimum value; PD: Parkinson disease; PeakEnerg: peak of energy; x: accelerometer mediolateral signal; y: accelerometer vertical signal; z: accelerometer anteroposterior signal. (A) Gait task. cov: coefficient of variation; FB: freezing band; frec_peak: frequency at the peak of energy; FreezeInd: freeze index; LB: locomotor band; MSI: mean stride interval; RatioPower: sum of the power in the freezing and locomotor band; skew: skewness; zcr: zero-crossing rate. (B) Balance task. buttonNoneFreq: frequency of tapping outside the button; CFREQ: centroidal frequency; F50: frequency containing 50% of total power; FRQD: frequency of dispersion of the power spectrum; HF: high frequency (>4 Hz); LF: low frequency (0.15-3.5 Hz); MF: medium frequency (4-7 Hz); post: postural; Power: energy between 3.5-15 Hz; RHL: ratio between power in high frequency and low frequency; rms: root mean square; TotalPower: energy between 15-3.5 Hz; trem: tremor; VHF: very high frequency (>7 Hz). (C) Voice task. c_mean: mean of the MFCC; gqc: glottis quotient close; log: energy of the signal and the first and second derivatives of the MFCC; MFCC: Mel-frequency cepstral coefficients; p95: 95th percentile; shbd: shimmer. (D) Tapping task. corXY: correlation of X and Y positions; cv: coefficient; DriftLeft: left drift; DriftRight: right drift; mad: median absolute deviation; numberTaps: number of taps; sd: standard deviation; TapInter: tap interval.

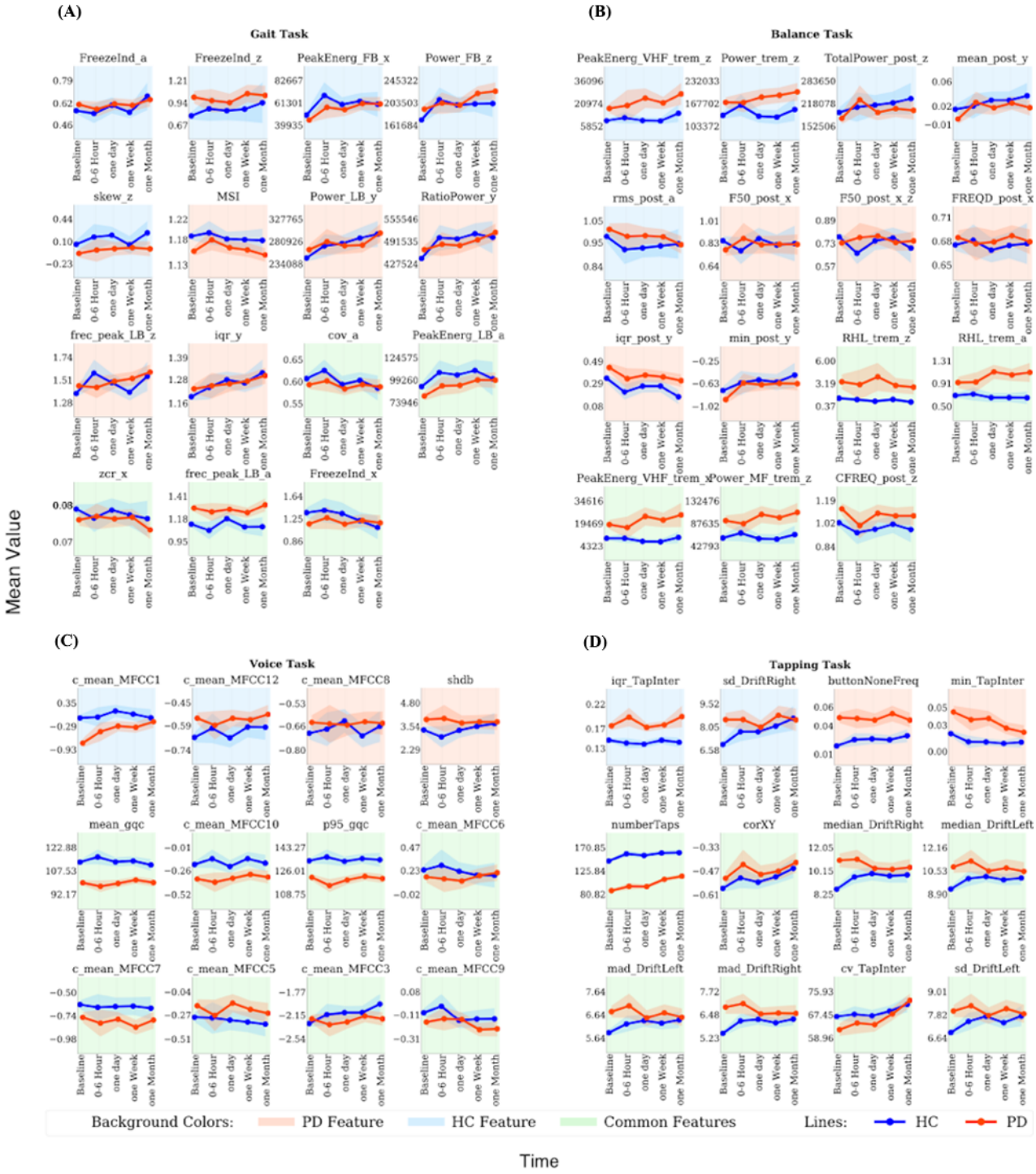
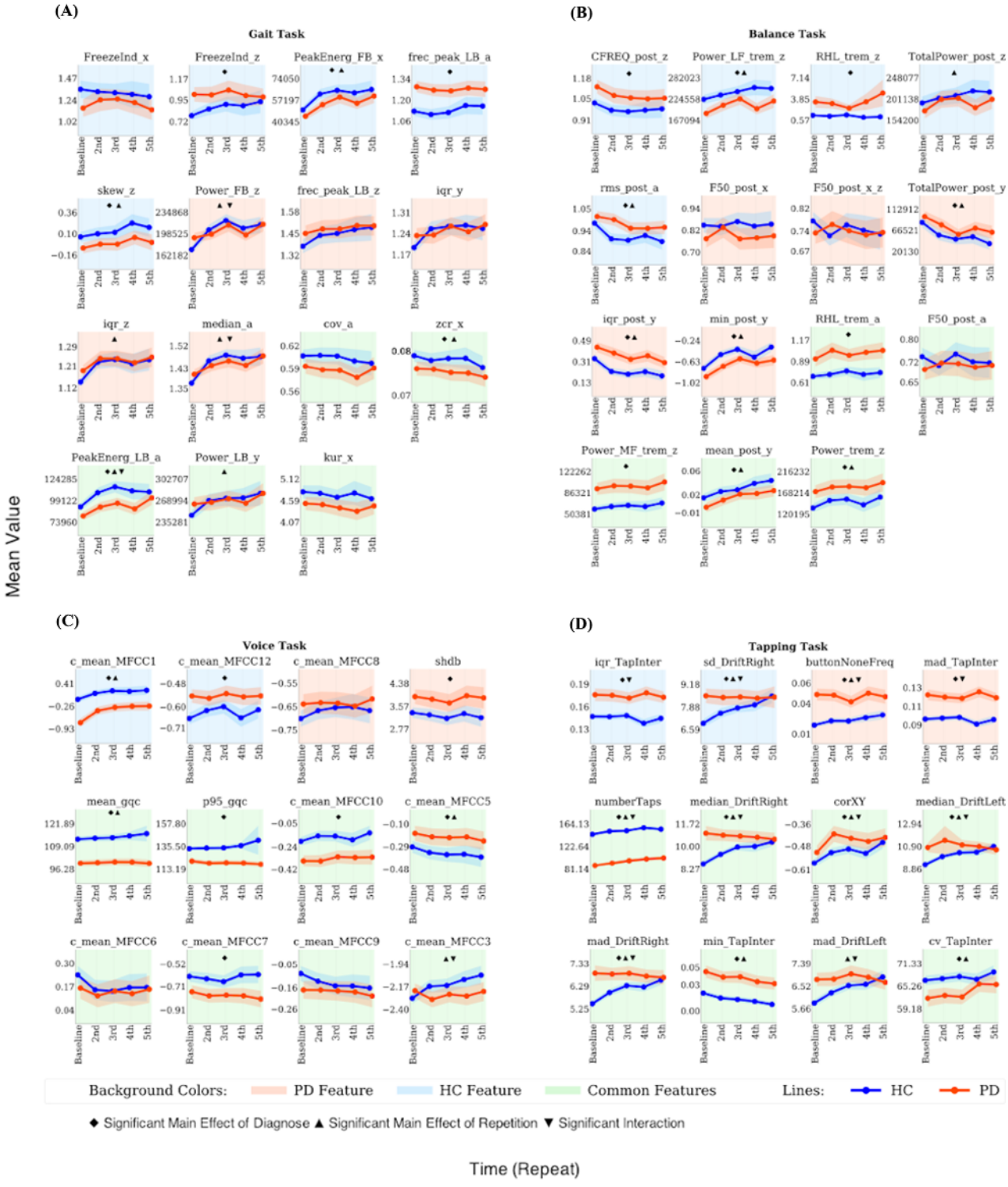


Figure 5. Mean value of the best performing baseline features across repetitions, calculated for PD and HC separately. a: accelerometer average signal; HC: healthy controls; iqr: interquartile range; min: minimum value; PD: Parkinson disease; PeakEnerg: peak of energy; x: accelerometer mediolateral signal; y: accelerometer vertical signal; z: accelerometer anteroposterior signal. (A) Gait task. cov: coefficient of variation; FB: freezing band; frec_peak: frequency at the peak of energy; FreezeInd: freeze index; kur: kurtosis; LB: locomotor band; MSI: mean stride interval; RatioPower: sum of the power in the freezing and locomotor band; skew: skewness; zcr: zero-crossing rate. (B) Balance task. buttonNoneFreq: frequency of tapping outside the button; CFREQ: centroidal frequency; F50: frequency containing 50% of total power; FRQD: frequency of dispersion of the power spectrum; HF: high frequency (>4 Hz); LF: low frequency (0.15-3.5 Hz); MF: medium frequency (4-7 Hz); post: postural; Power: energy between 3.5-15 Hz; RHL: ratio between power in high frequency and low frequency; rms: root mean square; TotalPower: energy between 15-3.5 Hz; trem: tremor; VHF: very high frequency (>7 Hz). (C) Voice task. c_mean: mean of the MFCC; gqc: glottis quotient close; log: energy of the signal and the first and second derivatives of the MFCC; MFCC: Mel-frequency cepstral coefficients; p95: 95th percentile; shbd: shimmer. (D) Tapping task. corXY: correlation of X and Y positions; cv: coefficient; DriftLeft: left drift; DriftRight: right drift; mad: median absolute deviation; numberTaps: number of taps; sd: standard deviation; TapInter: tap interval.



In an additional sensitivity analysis, we further tested if the between-group differences and group-by-repetition interaction remain significant when controlling for age and sex. The results (Tables S14-S17 in Multimedia Appendix 1) show that a significant effect of diagnosis was still identified for 2 out of 6 gait features, 8 out of 11 balance features, 1 out of 8 voice features, and 10 out of 11 tapping features. A significant effect of repetition was still found for 6 out of 8 gait features, 7 out of 8 balance features, 3 out of 4 voice features, and 10 out of 10 tapping features. Also, a significant main effect of diagnosis-by-repetition was still observed for 1 out of 3 gait features, 1 out of 1 balance feature, and 8 out of 10 tapping features.

Medication Effects

Lastly, we tested which of the most reliable features identified above also display sensitivity to PD medication. For this we compared the conditions reported by the patients as being before PD medication, after PD medication, or at best. A significant effect of PD medication was only observed for 2 out of 15 gait features, 1 out of 15 balance features, 2 out of 12 voice features, and 1 out of 12 tapping features (Figure S3, Tables S9 and S10-S13, medication column, in Multimedia Appendix 1).

Discussion

Principal Findings

Here we assessed the longitudinal test-retest reliability and stability of DB measures related to gait, balance, voice, and finger dexterity impairments in PD. We found a wide range of test-retest reliabilities across tasks and features ranging from poor to excellent, with highest reliabilities observed for voice followed by the tapping task. Only a few features had medium to large effect sizes for differentiation between PD and HC. For all tasks, a substantial percentage of features displayed significant longitudinal alterations in their mean values over time.

Overall, tapping and voice tasks revealed a better performance compared to gait and balance tasks with respect to test-retest reliability and observed effect sizes. Balance and gait tasks displayed consistently poor test-retest reliabilities as well as low effect sizes for differentiation between PD and HC, calling into question their usability for home-based applications. In contrast, best performing voice features displayed fair to excellent test-retest reliabilities across repetitions but also over weeks and months.

Unlike some previous studies that showed good performance and moderate to excellent correlation of gait and balance features with clinical score [4,39], the overall poor performance of these tasks in the m-Power study may be explained by the nature of these tasks, which requires strict supervision and monitoring. Both may not be sufficiently achieved in the self-administered setting of the m-Power study. Overall, acceleration-related features in the gait task and tremor-related features and those selected from frequency domain in the balance task displayed the best performance for the respective task [23,40]. The features related to Mel-frequency cepstral coefficients for the voice task displayed the highest effect sizes for this task, which is in line with previous studies showing its ability in identifying pathological speech [41,42]. In line with previous studies, features related to intertapping interval and precision of the tapping task (eg, number of taps, taps drift) displayed the best performance among all [43,44].

Most features showed a decrease in test-retest reliability with longer periods of time. This may reflect a consequence of the repetition effects and the group-by-repetition interaction observed in the analyses of variance for a substantial proportion of the features. Features selected from the tapping task were less sensitive to the effect of age and sex compared to other tasks. Overall, the effects of age and sex were not significant for most of the features. The analysis of elapsed time between repetitions also revealed that the time difference between repetitions did not have a significant effect on most of the features. ICC values obtained from the PD and HC groups were largely similar, suggesting that other non-PD related sources of variation may have played a larger role in the observed low ICC values. Determining these reasons requires more controlled experiments than provided by the m-Power study.

Despite a significant difference at baseline, several features did not differentiate PD and HC when using data from all time points. This effect became most pronounced for the gait task, likely due to its poor test-retest reliability performance. Differential learning, variation in motivation, medication, reduced adherence to task instructions, and other physical and environmental parameters may contribute to this loss of differentiation [2,10,12]. While a clear differentiation of motivation versus learning effects on the often-abstract DB features is difficult in an observational study design, a possible way to provide inference on this issue is to compare the direction of alterations in PD and HC. Assuming that alterations in PD relative to HC reflect impairment, movement of a feature state toward PD is likely to reflect worsening due to reduced motivation, disease progression, or other similar factors. In contrast, movements toward HC is likely to reflect improvement and is therewith compatible with a learning effect. We find a mixture of both effects for most tasks, suggesting the presence of both aspects in DB longitudinal data. These observations are also in line with previous studies showing that training may reduce motor impairment in PD [45-47]. In particular, for the tapping task the difference between PD and HC disappears for several features, which is primarily due to a shift in performance in HC. These findings may point to a differential change in motivation across groups. While differential learning has been previously reported [45,48-52], the differential change in motivation is an important novel aspect to consider when comparing DB measures between PD patients and HC. Understanding the sources leading to this variability of DB measures over time is a vital and open question that needs to be systematically addressed to enable their application for specific clinical questions.

Most patients with PD take dopaminergic medication to alleviate their motor functions. However, the responsiveness to PD medication highly varies between patients. Besides good reliability and the ability to differentiate PD and HC, another important and desired quality of an effective DB is therefore to monitor PD medication response. Among the most reliable

features from each task, only a few displayed significant but weak sensitivity to different medication conditions. One possible reason for this poor performance of DB measures in our study, as compared to some previous reports [20], might be the self-reported nature of the medication status in the m-Power data set, which likely introduced some noise variation (ie, different drugs and differences in time after administration). Nonetheless, our findings point to the need for further optimization of DB measures to increase their sensitivity to PD medication effects.

The self-administered design of the m-Power data set is also the major limitation of our study. In such an uncontrolled setting, accuracy in reporting the diagnosis and demographics, defining the medication status, and ensuring correct understanding of and compliance with the instructions may all have introduced variation into the study measures. The reported ballpark estimates for test-retest reliability and ability of the respective measures to differentiate between PD and HC therefore need to be carefully considered when interpreting our results. Another limitation of our study is the moderate adherence of participants in the m-Power study, which limited the number of participants who could be included in our analyses. Differences in age as well as lack of standardization of the time of day when the assessments were conducted are further sources of variation that may affect the generalizability of our findings [53]. Future studies may make inferences about the impact of different confounders such as comorbidities and disease severity on the longitudinal stability of DB. Also, further research is needed to establish the longitudinal stability of DB in the context of their relationship to clinical rating scales such as UPDRS.

Nonetheless, our findings clearly demonstrate the need for further optimization of DB tasks as well as for introducing careful monitoring and quality control procedures to enable integration of DB measures into clinically relevant applications.

Data Availability

The m-Power data set used for this paper is available upon registration from Synapse [54].

Acknowledgments

This study was supported by the Human Brain Project, funded from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement 785907 (Human Brain Project SGA2).

The data used in this study were contributed by users of the Parkinson m-Power mobile app as part of the m-Power study developed by Sage Bionetworks and described in Synapse [54].

Authors' Contributions

MSF performed analyses and wrote the manuscript. MG performed feature extraction. MSF, JD, and MG contributed to study design and writing the manuscript. SBE and JD designed the overall study and contributed to interpretation of the results. All authors reviewed and commented on the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplement.
[DOCX File , 1606 KB - jmir_v23i9e26608_app1.docx]

References

1. Jankovic J. Parkinson's disease: clinical features and diagnosis. J Neurol Neurosurg Psychiatry 2008 Apr;79(4):368-376. [doi: 10.1136/jnnp.2007.131045] [Medline: 18344392]
2. Prince J, Arora S, de Vos M. Big data in Parkinson's disease: using smartphones to remotely detect longitudinal disease phenotypes. Physiol Meas 2018 Apr 26;39(4):044005. [doi: 10.1088/1361-6579/aab512] [Medline: 29516871]
3. Insel TR. Digital Phenotyping: Technology for a New Science of Behavior. JAMA 2017 Oct 03;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
4. Lipsmeier F, Taylor KI, Kilchenmann T, Wolf D, Scotland A, Schjodt-Eriksen J, et al. Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial. Mov Disord 2018 Aug;33(8):1287-1297 [FREE Full text] [doi: 10.1002/mds.27376] [Medline: 29701258]
5. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. NPJ Digit Med 2019;2(1):14 [FREE Full text] [doi: 10.1038/s41746-019-0090-4] [Medline: 30868107]
6. Mahadevan N, Demanuele C, Zhang H, Volfson D, Ho B, Erb MK, et al. Development of digital biomarkers for resting tremor and bradykinesia using a wrist-worn wearable device. NPJ Digit Med 2020;3:5 [FREE Full text] [doi: 10.1038/s41746-019-0217-7] [Medline: 31970290]

7. Shah VV, McNames J, Mancini M, Carlson-Kuhta P, Nutt JG, El-Gohary M, et al. Digital Biomarkers of Mobility in Parkinson's Disease During Daily Living. J Parkinsons Dis 2020;10(3):1099-1111 [FREE Full text] [doi: 10.3233/JPD-201914] [Medline: 32417795]
8. Schlachetzki JCM, Barth J, Marxreiter F, Gossler J, Kohl Z, Reinfelder S, et al. Wearable sensors objectively measure gait parameters in Parkinson's disease. PLoS One 2017;12(10):e0183989 [FREE Full text] [doi: 10.1371/journal.pone.0183989] [Medline: 29020012]
9. Tracy JM, Özkanca Y, Atkins DC, Hosseini Ghomi R. Investigating voice as a biomarker: Deep phenotyping methods for early detection of Parkinson's disease. J Biomed Inform 2020 Apr;104:103362 [FREE Full text] [doi: 10.1016/j.jbi.2019.103362] [Medline: 31866434]
10. Espay AJ, Bonato P, Nahab FB, Maetzler W, Dean JM, Klucken J, Movement Disorders Society Task Force on Technology. Technology in Parkinson's disease: Challenges and opportunities. Mov Disord 2016 Sep;31(9):1272-1282 [FREE Full text] [doi: 10.1002/mds.26642] [Medline: 27125836]
11. Moore ST, Yungher DA, Morris TR, Dilda V, MacDougall HG, Shine JM, et al. Autonomous identification of freezing of gait in Parkinson's disease from lower-body segmental accelerometry. J Neuroeng Rehabil 2013 Feb 13;10:19 [FREE Full text] [doi: 10.1186/1743-0003-10-19] [Medline: 23405951]
12. Fisher JM, Hammerla NY, Rochester L, Andras P, Walker RW. Body-Worn Sensors in Parkinson's Disease: Evaluating Their Acceptability to Patients. Telemed J E Health 2016 Jan;22(1):63-69 [FREE Full text] [doi: 10.1089/tmj.2015.0026] [Medline: 26186307]
13. Rahlf AL, Petersen E, Rehwinkel D, Zech A, Hamacher D. Validity and Reliability of an Inertial Sensor-Based Knee Proprioception Test in Younger vs. Older Adults. Front Sports Act Living 2019;1:27 [FREE Full text] [doi: 10.3389/fspor.2019.00027] [Medline: 33344951]
14. Orlowski K, Eckardt F, Herold F, Aye N, Edelmann-Nusser J, Witte K. Examination of the reliability of an inertial sensor-based gait analysis system. Biomed Tech (Berl) 2017 Nov 27;62(6):615-622. [doi: 10.1515/bmt-2016-0067] [Medline: 28099115]
15. Hasegawa N, Shah VV, Carlson-Kuhta P, Nutt JG, Horak FB, Mancini M. How to Select Balance Measures Sensitive to Parkinson's Disease from Body-Worn Inertial Sensors-Separating the Trees from the Forest. Sensors (Basel) 2019 Jul 28;19(15):3320 [FREE Full text] [doi: 10.3390/s19153320] [Medline: 31357742]
16. Skodda S, Grönheit W, Mancinelli N, Schlegel U. Progression of voice and speech impairment in the course of Parkinson's disease: a longitudinal study. Parkinsons Dis 2013;2013:389195 [FREE Full text] [doi: 10.1155/2013/389195] [Medline: 24386590]
17. Aghanavesi S, Nyholm D, Senek M, Bergquist F, Memedi M. A smartphone-based system to quantify dexterity in Parkinson's disease patients. Informatics in Medicine Unlocked 2017;9:11-17. [doi: 10.1016/j.imu.2017.05.005]
18. Wissel B, Mitsi G, Dwivedi A, Papapetropoulos S, Larkin S, López Castellanos JR, et al. Tablet-Based Application for Objective Measurement of Motor Fluctuations in Parkinson Disease. Digit Biomark 2017;1(2):126-135 [FREE Full text] [doi: 10.1159/000485468] [Medline: 32095754]
19. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 2016 Mar 03;3:160011 [FREE Full text] [doi: 10.1038/sdata.2016.11] [Medline: 26938265]
20. Zhan A, Little M, Harris D, Abiola S, Dorsey E, Saria S. High Frequency Remote Monitoring of Parkinson’s Disease via Smartphone: Platform Overview and Medication Response Detection. arXiv. Preprint posted online on January 5, 2016 [FREE Full text]
21. Schwab P, Karlen W. PhoneMD: Learning to Diagnose Parkinson’s Disease from Smartphone Data. 2019 Jul 17 Presented at: Thirty-Third AAAI Conference on Artificial Intelligence; January 27-February 1, 2019; Honolulu, HI p. 1118-1125. [doi: 10.1609/aaai.v33i01.33011118]
22. Arora S, Venkataraman V, Zhan A, Donohue S, Biglan KM, Dorsey ER, et al. Detecting and monitoring the symptoms of Parkinson's disease using smartphones: A pilot study. Parkinsonism Relat Disord 2015 Jun;21(6):650-653. [doi: 10.1016/j.parkreldis.2015.02.026] [Medline: 25819808]
23. Goñi M, Eickhoff S, Far M, Patil K, Dukart J. Limited diagnostic accuracy of smartphone-based digital biomarkers for Parkinson’s disease in a remotely-administered setting. medRxiv. Preprint posted online on January 15, 2021 [FREE Full text] [doi: 10.1101/2021.01.13.21249660]
24. Mirelman A, Heman T, Yasinovsky K, Thaler A, Gurevich T, Marder K, LRRK2 Ashkenazi Jewish Consortium. Fall risk and gait in Parkinson's disease: the role of the LRRK2 G2019S mutation. Mov Disord 2013 Oct;28(12):1683-1690. [doi: 10.1002/mds.25587] [Medline: 24123150]
25. Blin O, Ferrandez AM, Serratrice G. Quantitative analysis of gait in Parkinson patients: increased variability of stride length. J Neurol Sci 1990 Aug;98(1):91-97. [doi: 10.1016/0022-510x(90)90184-o] [Medline: 2230833]
26. Hausdorff JM. Gait dynamics in Parkinson's disease: common and distinct behavior among stride length, gait variability, and fractal-like scaling. Chaos 2009 Jun;19(2):026113 [FREE Full text] [doi: 10.1063/1.3147408] [Medline: 19566273]
27. Miller Koop M, Ozinga SJ, Rosenfeldt AB, Alberts JL. Quantifying turning behavior and gait in Parkinson's disease using mobile technology. IBRO Rep 2018 Dec;5:10-16 [FREE Full text] [doi: 10.1016/j.ibror.2018.06.002] [Medline: 30135951]

28. Martinez-Mendez R, Sekine M, Tamura T. Postural sway parameters using a triaxial accelerometer: comparing elderly and young healthy adults. Comput Methods Biomech Biomed Engin 2012;15(9):899-910. [doi: 10.1080/10255842.2011.565753] [Medline: 21547782]
29. Prieto TE, Myklebust JB, Hoffmann RG, Lovett EG, Myklebust BM. Measures of postural steadiness: differences between healthy young and elderly adults. IEEE Trans Biomed Eng 1996 Sep;43(9):956-966. [doi: 10.1109/10.532130] [Medline: 9214811]
30. Palakurthi B, Burugupally SP. Postural Instability in Parkinson's Disease: A Review. Brain Sci 2019 Sep 18;9(9):239 [FREE Full text] [doi: 10.3390/brainsci9090239] [Medline: 31540441]
31. Pittman B, Ghomi RH, Si D. Parkinson's Disease Classification of mPower Walking Activity Participants. Annu Int Conf IEEE Eng Med Biol Soc 2018 Jul;2018:4253-4256. [doi: 10.1109/EMBC.2018.8513409] [Medline: 30441293]
32. Seifert K, Camacho O. Implementing Positioning Algorithms Using Accelerometers. 2007 Feb. URL: https://www.nxp.com/docs/en/application-note/AN3397.pdf [accessed 2021-02-01]
33. Palmerini L, Rocchi L, Mellone S, Valzania F, Chiari L. Feature selection for accelerometer-based posture analysis in Parkinson's disease. IEEE Trans Inf Technol Biomed 2011 May;15(3):481-490. [doi: 10.1109/TITB.2011.2107916] [Medline: 21349795]
34. Little MA, McSharry PE, Roberts SJ, Costello DAE, Moroz IM. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. Biomed Eng Online 2007 Jun 26;6:23 [FREE Full text] [doi: 10.1186/1475-925X-6-23] [Medline: 17594480]
35. Chiaramonte R, Bonfiglio M. Acoustic analysis of voice in Parkinson's disease: a systematic review of voice disability and meta-analysis of studies. Rev Neurol 2020 Jun 01;70(11):393-405 [FREE Full text] [doi: 10.33588/rn.7011.2019414] [Medline: 32436206]
36. Rao G, Fisch L, Srinivasan S, D'Amico F, Okada T, Eaton C, et al. Does this patient have Parkinson disease? JAMA 2003 Jan 15;289(3):347-353. [doi: 10.1001/jama.289.3.347] [Medline: 12525236]
37. Jobbágy A, Harcos P, Karoly R, Fazekas G. Analysis of finger-tapping movement. J Neurosci Methods 2005 Jan 30;141(1):29-39. [doi: 10.1016/j.jneumeth.2004.05.009] [Medline: 15585286]
38. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment 1994 Dec;6(4):284-290. [doi: 10.1037/1040-3590.6.4.284]
39. Horak FB, Mancini M. Objective biomarkers of balance and gait for Parkinson's disease using body-worn sensors. Mov Disord 2013 Sep 15;28(11):1544-1551 [FREE Full text] [doi: 10.1002/mds.25684] [Medline: 24132842]
40. Sejdić E, Lowry KA, Bellanca J, Redfern MS, Brach JS. A comprehensive assessment of gait accelerometry signals in time, frequency and time-frequency domains. IEEE Trans Neural Syst Rehabil Eng 2014 May;22(3):603-612 [FREE Full text] [doi: 10.1109/TNSRE.2013.2265887] [Medline: 23751971]
41. Khan T. Running-speech MFCC are better markers of Parkinsonian speech deficits than vowel phonation and diadochokinetic. DiVA. Preprint posted online on July 15, 2015 [FREE Full text]
42. Tsanas A, Little MA, McSharry PE, Spielman J, Ramig LO. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. IEEE Trans Biomed Eng 2012 May;59(5):1264-1271. [doi: 10.1109/TBME.2012.2183367] [Medline: 22249592]
43. Memedi M, Khan T, Grenholm P, Nyholm D, Westin J. Automatic and objective assessment of alternating tapping performance in Parkinson's disease. Sensors (Basel) 2013 Dec 09;13(12):16965-16984 [FREE Full text] [doi: 10.3390/s131216965] [Medline: 24351667]
44. Lee CY, Kang SJ, Hong S, Ma H, Lee U, Kim YJ. A Validation Study of a Smartphone-Based Finger Tapping Application for Quantitative Assessment of Bradykinesia in Parkinson's Disease. PLoS One 2016;11(7):e0158852 [FREE Full text] [doi: 10.1371/journal.pone.0158852] [Medline: 27467066]
45. Olson M, Lockhart TE, Lieberman A. Motor Learning Deficits in Parkinson's Disease (PD) and Their Effect on Training Response in Gait and Balance: A Narrative Review. Front Neurol 2019;10:62 [FREE Full text] [doi: 10.3389/fneur.2019.00062] [Medline: 30792688]
46. Steib S, Wanner P, Adler W, Winkler J, Klucken J, Pfeifer K. A Single Bout of Aerobic Exercise Improves Motor Skill Consolidation in Parkinson's Disease. Front Aging Neurosci 2018;10:328 [FREE Full text] [doi: 10.3389/fnagi.2018.00328] [Medline: 30405397]
47. Bryant M, Workman CD, Jamal F, Meng H, Jackson GR. Feasibility study: Effect of hand resistance exercise on handwriting in Parkinson's disease and essential tremor. J Hand Ther 2018;31(1):29-34. [doi: 10.1016/j.jht.2017.01.002] [Medline: 28389133]
48. Krebs HI, Hogan N, Hening W, Adamovich SV, Poizner H. Procedural motor learning in Parkinson's disease. Exp Brain Res 2001 Dec;141(4):425-437. [doi: 10.1007/s002210100871] [Medline: 11810137]
49. Sehm B, Taubert M, Conde V, Weise D, Classen J, Dukart J, et al. Structural brain plasticity in Parkinson's disease induced by balance training. Neurobiol Aging 2014 Jan;35(1):232-239. [doi: 10.1016/j.neurobiolaging.2013.06.021] [Medline: 23916062]

50. Foreman KB, Sondrup S, Dromey C, Jarvis E, Nissen S, Dibble LE. The effects of practice on the concurrent performance of a speech and postural task in persons with Parkinson disease and healthy controls. Parkinsons Dis 2013;2013:987621 [FREE Full text] [doi: 10.1155/2013/987621] [Medline: 23841022]
51. Agostino R, Currà A, Soldati G, Dinapoli L, Chiacchiari L, Modugno N, et al. Prolonged practice is of scarce benefit in improving motor performance in Parkinson's disease. Mov Disord 2004 Nov;19(11):1285-1293. [doi: 10.1002/mds.20247] [Medline: 15390058]
52. Behrman AL, Cauraugh JH, Light KE. Practice as an intervention to improve speeded motor performance and motor learning in Parkinson's disease. J Neurol Sci 2000 Mar 15;174(2):127-136. [doi: 10.1016/s0022-510x(00)00267-7] [Medline: 10727698]
53. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020;3:21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
54. mPower Public Researcher Portal. URL: https://www.synapse.org/#!Synapse:syn4993293 [accessed 2021-02-01]

Abbreviations

DB: digital biomarkers
HC: healthy controls
ICC: intraclass correlation coefficients
PD: Parkinson disease
rm-ANOVA: repeated-measures analyses of variance
UPDRS: Unified Parkinson's Disease Rating Scale

Edited by R Kukafka; submitted 29.12.20; peer-reviewed by M Lindemann, O Chen, O Beiki, Z Feng, Y Cai; comments to author 03.03.21; revised version received 21.06.21; accepted 05.07.21; published 13.09.21.

Please cite as:
Sahandi Far M, Eickhoff SB, Goni M, Dukart J
Exploring Test-Retest Reliability and Longitudinal Stability of Digital Biomarkers for Parkinson Disease in the m-Power Data Set: Cohort Study
J Med Internet Res 2021;23(9):e26608
URL: https://www.jmir.org/2021/9/e26608
doi:10.2196/26608
PMID:34515645

Original Paper

Assessing Professionals' Adoption Readiness for eMental Health: Development and Validation of the eMental Health Adoption Readiness Scale

Milou A Feijt[1], MSc; Yvonne A W de Kort[1], Prof Dr; Joyce H D M Westerink[1,2], Prof Dr; Joyce J P A Bierbooms[3,4], PhD; Inge M B Bongers[3,4], Prof Dr; Wijnand A IJsselsteijn[1], Prof Dr

[1]Human-Technology Interaction Group, Department of Industrial Engineering & Innovation Sciences, Eindhoven University of Technology, Eindhoven, Netherlands

[2]Digital Engagement, Cognition & Behavior Group, Philips Research, Eindhoven, Netherlands

[3]Tranzo, Tilburg School of Social and Behavioural Sciences, Tilburg University, Tilburg, Netherlands

[4]Mental Healthcare Eindhoven, Eindhoven, Netherlands

Corresponding Author:
Milou A Feijt, MSc
Human-Technology Interaction Group
Department of Industrial Engineering & Innovation Sciences
Eindhoven University of Technology
De Rondom 70
Eindhoven, 5612 AP
Netherlands
Phone: 31 40 247 2889
Email: m.a.feijt@tue.nl

Abstract

Background: The last few decades have witnessed significant advances in the development of digital tools and applications for mental health care. Despite growing evidence for their effectiveness, acceptance and use of these tools in clinical practice remain low. Hence, a validated and easy-to-use instrument for assessing professionals' readiness to adopt eMental health (EMH) is necessary to gain further insights into the process of EMH adoption and facilitate future research on this topic.

Objective: The aim of this study is to develop and validate an instrument for assessing mental health care professionals' readiness to adopt EMH.

Methods: Item generation was guided by literature and inputs from mental health care professionals and experts in survey development. Exploratory factor analyses were conducted on an initial set of 29 items completed by a sample of mental health care professionals (N=432); thereafter, the scale was reduced to 15 items in an iterative process. The factor structure thus obtained was subsequently tested using a confirmatory factor analysis with a second sample of mental health care professionals (N=363). The internal consistency, convergent validity, and predictive validity of the eMental Health Adoption Readiness (eMHAR) Scale were assessed.

Results: Exploratory factor analysis resulted in a 3-factor solution with 15 items. The factors were analyzed and labeled as perceived benefits and applicability of EMH, EMH proactive innovation, and EMH self-efficacy. These factors were confirmed through a confirmatory factor analysis. The total scale and subscales showed a good internal consistency (Cronbach α=.73-.88) along with acceptable convergent and predictive relationships with related constructs.

Conclusions: The constructed eMHAR Scale showed a conceptually interpretable 3-factor structure having satisfactory characteristics and relationships with relevant concepts. Its ease of use allows for quick acquisition of data that can contribute to understanding and facilitating the process of adoption of EMH by clinical professionals.

(J Med Internet Res 2021;23(9):e28518) doi:10.2196/28518

KEYWORDS

eMental health; adoption of innovation; mental health care; scale development

Introduction

Over recent decades, a wide range of digital tools and technologies have been used in the practice of mental health care. There is a growing body of evidence for the effectiveness of eMental health (EMH) [1,2], which can be defined as "the use of information and communication technology (ICT)—in particular, the many technologies related to the Internet—when these technologies are used to support and improve mental health conditions and mental healthcare" [3]. Multiple benefits, such as an increased access to psychological treatment, convenience, and enhanced self-reflection and self-disclosure of the client, are associated with the use of these technologies [4-7]. Despite these advantages, studies consistently report that, before COVID-19, the use of EMH in daily practice was relatively low, finding that 80%-90% of practitioners never or only rarely made use of web-based tools in their routine [8-10].

Several studies have aimed to elucidate what determines this low use of EMH and have pointed to a variety of factors that can play a role, such as the characteristics of clients and professionals, technological aspects, and legal and managerial issues [11-13]. Initially, it was assumed that client factors such as their attitudes and skills were the most important barriers; therefore, studies tended to focus on clarifying these characteristics [14]. However, research suggests that the role of professionals is at least as important and that their adoption is crucial in the successful implementation of EMH [15]. Adoption has been described as a staged process that comprises required knowledge and skills, acceptance, implementation in daily procedures, actual use, and evaluation [16]. Reviews on the perspective of professionals demonstrate that a variety of factors influence adoption, such as (lack of) perceived benefits, applicability and conditions in daily practice, experience and know-how regarding EMH, and technical issues [6,7,17].

As conceptualized in the Levels of Adoption of eMental Health Model [5], each stage in the adoption process is associated with different barriers and drivers that professionals experience. Some of these are external factors, such as the organizational setting and supporting conditions in daily practice (eg, time and resources provided by an institution), characteristics of the EMH tool, and available technical infrastructure, whereas other factors concern individual characteristics of professionals, such as their beliefs and attitudes, knowledge, skills, and experience regarding EMH [12]. Research has shown that these individual characteristics are crucial to the adoption process [13]. Therefore, in this study, we focus on these personal characteristics and refer to these with the term adoption readiness—the extent to which a professional is ready to use EMH (ie, has a positive attitude, is motivated, and possesses the necessary skills and knowledge). By doing so, we distinguish individual adoption readiness from the broader context of adoption, as the latter also includes the actual use in practice and hence is highly influenced by external factors [12].

Studies that investigated relevant individual characteristics generally reported relatively low levels of acceptance, skills, knowledge, and experience regarding EMH, that is, low adoption readiness [6,11,13]. These studies also indicated that perceived benefits and applicability in daily practice, an innovative attitude, experience, and feelings of self-competency regarding EMH are important determinants of adoption readiness. Although providing valuable knowledge on adoption readiness, each of these studies used different measures and definitions; to date, no validated quantitative instrument comprehensively captures all the relevant aspects. This complicates research on this topic, and professionals' adoption readiness and its underlying factors have remained difficult to qualify and quantify, which also makes it more difficult to develop effective strategies to increase the uptake of EMH.

To improve our understanding, we need a measurement instrument that can reliably assess a professional's readiness to adopt EMH. The development of our scale benefits from previous studies in the same direction. The measurement instruments that were used in these studies focused on attitudes [18], comprised *ad hoc* measures for one specific study [19,20], or were measures that were originally developed to assess individuals' adoption of technology in general [6,9,10]. Other studies have used qualitative methods that have the strength of providing in-depth results, but are less suited for studying larger sample sizes and cross-group comparisons [21,22]. In addition, related studies have focused on the client's perception [23] or organizational structures and implementation strategies [24]. We build upon and extend existing measures by developing a validated scale that specifically focuses on gauging individuals' readiness to adopt new or existing EMH tools, independent of organizational settings or specific tools. A practical measurement instrument would facilitate studies with larger samples and allow scholars to systematically test the specific hypotheses about how professionals' adoption readiness relates to external factors, how the various individual and external factors interact, and what their relative roles are in the actual use of EMH in mental health care practice. In addition, it would allow cross-study comparisons to investigate the differences between demographic groups and longitudinal comparisons to examine developments over time—all important means to increase our understanding of the EMH adoption process.

This study presents the process of the development and validation of the eMental Health Adoption Readiness (eMHAR) Scale: a scale that assesses EMH adoption readiness of mental health care professionals. Specifically, this study describes the scale construction process (item generation and item selection), and validation of the scale through an exploratory factor analysis followed by a confirmatory factor analysis, using two sizable samples of mental health care professionals, along with an analysis of scale reliability as well as convergent and predictive validity. In the *Discussion* section, we will highlight the relevance of the scale and reflect on its theoretical and applied strengths and current limitations.

Methods

Scale Construction

The construction of the eMHAR Scale broadly involved the following steps: (1) item development, generating an item pool, and having survey experts review the items; (2) test with a sample of the target population to explore the underlying factor

structure, refine the scale, and obtain an indication of the psychometric properties of the resulting scale [25,26]; and (3) a second study with a different sample to confirm the factor structure and further establish its psychometric properties. This procedure of first conducting an exploratory factor analysis followed by a confirmatory factor analysis is generally recommended in scale development [27-29]. The next sections describe each of these steps in more detail. The study protocol was assessed and approved by the ethical review board of the Eindhoven University of Technology.

Item Development

To generate our item pool, we started with the results of Feijt et al [5]. This qualitative study reported factors associated with EMH adoption readiness that formed the basis for the items that should be covered by the scale. These factors included perceived benefits, applicability in daily clinical practice, personal affinity, proactive behavior, knowledge, experience, and skills. To ensure that we did not miss any relevant topics, we additionally searched PsycINFO and PubMed for articles on EMH adoption readiness that were published since the publication of this work [5]. Thus, we used the terms (*e(Mental)Health* OR *technology*) AND (*adoption* OR *attitude* OR *barriers* OR *drivers*) and limited the results to articles published from April 2017 to January 2018. This search supported our findings and did not yield additional topics.

We translated the topics into statements that used the language familiar to mental health care professionals. This procedure led to the first draft version of the eMHAR Scale. It comprised 25 statements with a 5-point Likert scale response format: strongly disagree, disagree, neutral, agree, and strongly agree, which were assigned the values 1-5, respectively. We collected feedback on this draft version from 20 mental health care professionals as part of a workshop that took place during an EMH event. With 3 of them, we subsequently used a (phone-based) think-aloud strategy while filling out the questionnaire. Thus, we obtained a better understanding of unclear or redundant items. On the basis of the feedback, several items were rephrased, and definitions were clarified: four new items were added (three regarding pioneering activities and one on perceived added value) and two items that were considered duplications (regarding the lack of applicability of EMH and experience with EMH) were removed. This second draft version was presented to 3 experts on the topic of EMH for a final check to ensure that all the identified aspects of EMH adoption readiness were adequately and sufficiently covered. On the basis of this expert review, another two items were added to skills and feelings of competency, further resulting in a final item pool of 29 items (Multimedia Appendix 1).

Validation Study 1

Web-Based Survey

The data for the first validation were collected via a web-based survey. In addition to the eMHAR Scale items, the survey comprised self-developed items on concepts related to adoption readiness of EMH to analyze convergent validity: perceived added value, feeling of competency regarding EMH in general, perceived proficiency for various EMH skills and for specific tools, and frequency of use. *The perceived added value* of EMH was measured for each of 13 EMH tools (eg, videocall, web-based modules, and virtual reality) on 5-point scales ranging from 1 (not valuable) to 5 (very valuable). *Feeling of competency* regarding EMH was measured using a 5-point scale describing incrementally increasing levels of skills and knowledge. Moreover, participants indicated *perceived proficiency* regarding nine EMH skills (eg, ability to establish an empathic interaction on the internet and having sufficient knowledge of privacy and security requirements) and *perceived proficiency* regarding the same 13 EMH tools on 5-point scales ranging from 1 (not at all competent) to 5 (very competent). Last, the *frequency of use of EMH* was also probed for the same 13 EMH tools on 5-point scales from 1 (almost never) to 5 (almost every day). The items for these convergent measures can be found in Multimedia Appendix 2. We expected that the eMHAR Scale or its potential subscales would show substantial positive correlations with these constructs.

Finally, the survey included basic demographic questions and several items regarding the characteristics of participants' everyday clinical practice (ie, which psychological disorders they treated, which kinds of psychological treatments they provided, years of professional experience, and whether they had received training in EMH). The time required to complete the survey was approximately 15 minutes.

Sample Size and Recruitment

For factor analysis, a sufficient observation-to-item ratio is required to ensure the reliability of the correlation coefficients that are used to compute the factors [30]. Although there are varying opinions on what exactly is an adequate ratio, an index of 10:1 is generally considered sufficient [29,30]. With our item pool of 29 items, our aim was to recruit at least 300 mental health care professionals. The target sample comprised mental health care professionals representing broader ranges in the use and experience of EMH, in the types of professions within mental health care, and the types of mental health care institutions to which they were affiliated. To achieve this, the participants were recruited by contacting several large mental health care organizations spread across the Netherlands. A total of six mental health care institutions disseminated an announcement with information and the survey link either through email or intranet. Furthermore, a similar announcement was placed in the newsletters and on the network webpages of three national professional associations of psychologists: the Dutch Institute for Psychology, the Dutch Association of Health Psychologists, and the Dutch Association of Independent Psychotherapists. In addition, several independent practitioners were contacted directly through the authors' network. Participants could sign up for a raffle in which 10 gift vouchers were allotted as a reward for their participation.

Data Collection and Sample

Data were collected between October 2018 and April 2019. In total, 432 participants (288/432, 66.7% female) completed the survey, with ages ranging from 20 to 69 years (mean 41.3, SD 12.1). The most frequently reported professions were clinical or counseling psychologists, psychiatric nurses, and social

workers. Table 1 presents further details of the demographic data of the sample.

Data Screening

Before the factor analysis, the items were examined for the accuracy of data entry, missing values, and their fit with multivariate assumptions by checking their distributions, descriptive statistics, frequencies, plots, and standardized residuals [30]. Second, the existence of multivariate outliers was examined with the Mahalanobis distance at $P<.001$, which was evaluated as chi-square with the degrees of freedom equal to the number of variables [30]. Identified cases were inspected to understand the cause of the significant value, and analyses were both run with and without the outliers to determine the effect of the outliers on the results, to decide upon their retention or removal.

Table 1. Demographic data for the samples of study 1 (N=432) and study 2 (N=363), including gender, age, and profession.

Characteristic	Study 1, n (%)	Study 2, n (%)
Gender		
Male	144 (33.3)	95 (26.2)
Female	288 (66.7)	268 (73.8)
Age (years)		
<25	31 (7.2)	24 (6.6)
26-35	144 (33.3)	149 (41.0)
36-45	83 (19.2)	83 (22.8)
46-55	106 (24.5)	65 (17.9)
56-65	63 (14.6)	40 (11.0)
65+	2 (0.5)	2 (0.6)
Profession		
Clinical or counseling psychologists and psychotherapists	126 (29.2)	135 (37.2)
Psychiatric nurses	147 (34.0)	118 (32.5)
Social work	115 (26.6)	78 (21.5)
Expressive therapists (eg, creative arts therapist and psychomotor therapist)	13 (3.0)	8 (2.2)
Physicians (eg, psychiatrist, general practitioner, and neurologist)	27 (6.3)	23 (6.3)
Other (eg, researcher and team manager)	4 (0.9)	1 (0.3)

Scale Analysis and Evaluation

Although we had expectations about the underlying factors of the eMHAR Scale based on a previous work [5], we conducted an exploratory factor analysis to allow for the detection of nonhypothesized factors, which is a recommended procedure in the early stages of scale development [27,28,31]. Owing to the ordinal nature of Likert scales, the factor analysis in this study was conducted on the polychoric correlation matrix to correct for statistical issues that could arise because of the attenuation of the correlations between ordinal variables [32]. Before conducting the actual analysis, we checked the factorability of the items. Therefore, we examined whether the items were sufficiently correlated to produce representative factors by evaluating each item's interitem correlation (>0.30, for at least three items), significance of Bartlett test of sphericity, and the Kaiser-Meyer-Olkin test (>0.80) [29,30].

We first conducted a principal component analysis to establish the number of factors to be extracted [27]. Then, we conducted a parallel analysis [33] with the eigenvalues as the leading criterion to determine the number of factors to extract, as this method is found to be the most robust and sophisticated [34]. As multiple criteria should be used to determine the number of factors to be extracted [29,35,36], we checked the results from the parallel analysis with other frequently used criteria: the Kaiser criterion (eigenvalue >1 [37]), the scree test (point on the scree plot where the eigenvalues seem to level off [38]), and the cumulative percentage of variance extracted (>60% explained variance [29]). However, this analysis did not yield different solutions.

For our common factor analysis, we applied the principal axis factoring method and Oblimin oblique rotation with Kaiser normalization. Oblique rotation is best suited in social sciences, as many constructs in this field have an empirically and theoretically based expectation to be correlated with each other [29,35,39]. Moreover, in the (unlikely) case that the factors are actually uncorrelated, orthogonal and oblique factors will produce similar results [40].

We cycled through an iterative analysis process [27,29,30]: (1) conduct a principal component analysis to determine the number of factors to retain; (2) conduct principal axis factoring with Oblimin rotation; (3) examine the rotated solution and assess the factor loadings for low loadings and cross-loadings; (4) evaluate potential items for deletion and delete the most problematic items. If an item was deleted, then we returned to step 1. This process was repeated until a satisfactory solution with a simple structure (ie, highest item loading >0.40,

cross-loadings <0.30, and conceptual convergence between items in one factor) was achieved.

We then reverse scored the negatively worded items and obtained a total score, which was computed as the unweighted mean of all items, and obtained three subscale scores constituting the unweighted mean of the items of each of the three respective factors. We calculated both Pearson correlations and attenuation-corrected correlations between the subscales [41]. Finally, we calculated Cronbach α coefficients to probe the internal consistency.

To assess the construct validity of the scale, we examined convergent validity by calculating correlations between the scales (both the total scale and the subscales) and the measures we included for probing convergent validity: perceived added value of various EMH tools, feeling of competency regarding EMH, perceived proficiency regarding EMH skills and tools, and actual use of these tools. We calculated Spearman rank-order correlations because these items were measured on ordinal scales. It was expected that the total score would positively correlate with these additional measures and that such correlations would also be found for the subscale scores. All analyses were conducted using the Stata statistical software package, version 14.2 (StataCorp) [42].

Validation Study 2

A second validation with a different sample was conducted to test whether the factor structure found in the first study could be confirmed and to find further support for the scale's psychometric properties.

Web-Based Survey

The second study used a web-based survey similar to the first study, comprising the resulting eMHAR Scale items from study 1 and the same convergent measures: perceived added value of EMH tools, feeling of competency regarding EMH, perceived proficiency in EMH skills and with specific tools, and frequency of use. We again expected that the eMHAR Scale and the subscales would correlate substantially with these concepts.

Furthermore, the survey included questions regarding barriers, drivers, and needs experienced by professionals while using EMH and questions on their experiences with the interaction with their clients through EMH. Owing to the outbreak of the COVID-19 pandemic, we also included some questions that probed differences in the use and perceived usefulness of EMH between the period before the COVID-19 pandemic and at the time of the survey (June-September 2020) during the first wave. The survey concluded with basic demographic questions, such as years of professional experience and whether they had received training in EMH. The required time to complete the survey was established at approximately 20 minutes.

Sample Size and Recruitment

To establish our required sample size, we followed a simulation study that compared the validity of confirmatory factor analysis solutions for various sample sizes [43]. For our 3-factor solution, the minimum required sample size to achieve an excellent agreement (ie, congruence values >0.98) between the sample and population solutions was determined to be 200. As in the first study, our recruitment was aimed at obtaining a sample representative of the population of mental health care professionals in the Netherlands. Four mental health care institutions disseminated an announcement with information and the survey link either through email or intranet. Furthermore, an announcement was placed in the newsletter of a national professional association of psychologists, that is, Dutch Mental Health care. The message contained information about the purpose of the study and the link to the web-based survey. Participants could sign up for a raffle in which 12 gift vouchers were allotted as a reward for their participation.

Data Collection and Sample

Data were collected between June 2020 and September 2020. It is important to note that this second phase of data collection occurred after the COVID-19 outbreak. Practitioners had been forced to switch to web-based treatments because face-to-face treatment was not, or only sparsely, possible during the first peak of the pandemic. This led to a sharp increase in the use of EMH, where 80%-90% of the practitioners never or only rarely made use of web-based tools in their routine [8-10]; in May 2020, 80%-90% reported using it almost daily [44]. In total, 363 participants (268/363, 73.8% female) completed the survey, with ages ranging from 18 to 70 years (mean 39.1, SD 11.5). Again, the most frequent professions were clinical or counseling psychologists, psychiatric nurses, and social workers. In total, 92.5% (336/363) of the respondents indicated that they did not participate in the previous study, and inclusion of the overlapping 7.4% (27/363) did not affect the results. Table 1 presents further details on the demographic data of the second sample.

Scale Analysis and Evaluation

After the data screening process as described for study 1, a confirmatory factor analysis was conducted on the data from the second sample with structural equation modeling using robust maximum likelihood estimation and the standardization of latent factors. The root mean square error of approximation (RMSEA) and the comparative fit index (CFI) were used to assess model fit; RMSEA<0.07 and CFI>0.92 indicated an acceptable model fit [29].

Then, we followed the same procedure as in study 1 to obtain the total score and subscale scores (ie, reverse scoring, computing the unweighted mean scores for the total scale and respective subscales), the correlations between the subscales, and Cronbach α coefficients. To assess the construct validity, we again calculated Spearman rank-order correlations between the scales and the associated measures to examine the convergent validity. Another important component of the scale's validity pertains to its ability to measure changes in adoption readiness. Due to the sharp increase in the use of EMH as a consequence of the COVID-19 outbreak that occurred between the data collections [44], we expected that the gained experience in EMH would lead to higher feelings of competency in EMH; thus, scores on the EMH self-efficacy subscale and total scale should be higher in the second sample than in the first. No differences were expected in the other two subscales. We analyzed this through two methods of analysis: multigroup confirmatory factor analysis and independent samples two-tailed

t tests. Although multigroup confirmatory factor analysis is a precise method to compare responses across groups, it depends on the particular samples that are being compared, which complicates cross-study comparisons, especially when complete data sets are not accessible [45]. Therefore, we also compared the unweighted means of the subscales with independent sample *t* tests. For the multigroup confirmatory factor analysis, the first step involved testing for various levels of measurement invariance, that is, whether the scale measured the same construct in the two samples, which is a precondition for comparing the sample means. We tested this using the free baseline approach, which means testing successive models that are gradually being restricted (a more elaborate description of this strategy is presented in the study by Kline [45]). A CFI decrease ≤0.01 was used as the criterion to indicate no significant decrement of model fit, that is, supporting measurement invariance across the samples [46]. After establishing measurement invariance, we compared the estimated mean differences. Here, we took study 1 as the reference group, setting the intercepts at 0, and the means of study 2 were free to estimate. The estimated mean differences are then assessed with the value of the critical ratio, which is calculated as the parameter estimate divided by its SE. This provides a *z* statistic that tests whether the estimate is statistically different from zero. Second, we also compared the unweighted mean scores on the total and subscales of the first and second sample using independent samples *t* tests (two-tailed, α=.05).

The (multigroup) confirmatory factor analysis was conducted with R, version 4.0 (R Foundation for Statistical Computing) [47], via the software package RStudio, version 1.3 (RStudio) using the lavaan library [48], which is a package for structural equation modeling. All other analyses were conducted using the Stata statistical software package, version 14.2 (StataCorp) [42].

Results

Validation Study 1

Exploratory Factor Analysis

Screening of the data showed that there were no missing values or violations of multivariate assumptions. We identified 11 cases out of the total sample of 432 that exceeded the critical value for the Mahalanobis distance. Upon inspection of these cases, none of them showed suspicious values (eg, only reporting very low or high values), suggesting that they misunderstood the items or had a particular response tendency. In addition, performing the factor analysis with and without these cases did not yield substantial differences in the found structure and factor loadings. Therefore, we decided to keep all the cases in our final analyses.

All 29 items had at least three interitem correlations >0.30. The Kaiser-Meyer-Olkin test of sampling adequacy had a value of 0.92 for the total item pool, and on the item level, all values exceeded 0.84. Bartlett test of sphericity (χ^2_{406}=7063.00) was significant (P<.001). These results indicated that the items were appropriate for factor analysis [29].

Initially, a 4-factor solution emerged from the iterative analysis procedure described above. During the process, we deleted seven items (in seven consecutive iterations) because the cross-loadings exceeded the recommended threshold of 0.30 [29]. From the remaining 21 items, four more items with relatively low loadings on all components were removed. One of the four factors then comprised only two items, both of which showed skewness values above the recommended value of 1.00 and were also difficult to interpret theoretically. For these reasons, these two items were also removed. The final 15-item solution with three factors explained 60.59% of the total variance. Table 2 shows the loadings of the items on the respective factors.

The first factor (items 1-7) concerns the perceived benefits and applicability of EMH; the extent to which the professional thinks EMH can have added value for clinical practice and fits the mental health care profession. The second factor (items 8-12) covers proactive innovation toward EMH; whether the professional encourages other colleagues to work with EMH and is involved in EMH development. The third factor (items 13-15) concerns perceived self-efficacy regarding EMH; the extent to which the professional feels competent and possesses the skills necessary to work with EMH.

Table 2. The 15 items in the final 3-factor solution and their factor loadings (>0.40)[a].

Item	Factor 1: perceived benefits of EMH[b]	Factor 2: EMH proactive innovation	Factor 3: EMH self-efficacy
1. eHealth fits well to my work as a health care professional	0.67	__[c]	—
2. Contact between health care professional and client always has to be face-to-face (R[d])	0.63	—	—
3. I expect that eHealth provides benefits to the care that I deliver	0.70	—	—
4. eHealth does not improve the care that I deliver (R)	0.71	—	—
5. eHealth does not fit the profession of a mental health care professional (R)	0.78	—	—
6. eHealth does not have any added value for my work as a mental health care professional (R)	0.69	—	—
7. eHealth is an indispensable part of the mental health care profession	0.54	—	—
8. I am involved in setting up initiatives for the development of new eHealth tools and applications	—	0.91	—
9. Compared with my colleagues, I use eHealth a lot	—	0.71	—
10. Compared with colleagues, I take a lot of initiative regarding eHealth	—	0.80	—
11. I have ideas about new eHealth tools and technologies that could be developed (eg, virtual reality, gaming, biofeedback)	—	0.59	—
12. In my work I try to stimulate colleagues to use eHealth	—	0.78	—
13. I have the skills that are necessary to apply eHealth in my work	—	—	0.71
14. Using eHealth tools comes easy to me	—	—	0.67
15. I have to learn new skills to start using eHealth (R)	—	—	0.63

[a]Extraction method; principal axis factoring; rotation method; Oblimin with Kaiser normalization.

[b]EMH: eMental health.

[c]Factor loading <0.3.

[d]R: reverse scored.

Psychometric Evaluation

Table 3 presents the descriptive statistics for the total scale and subscales. Multimedia Appendix 3 presents a histogram of the total scores.

Table 3. Descriptive statistics of the scores on the total eMHAR Scale and subscales in both studies.

Subscale	Mean (SD; range)	Median	Skewness	Kurtosis
Study 1 (N=432)				
Total eMHAR[a] Scale	3.22 (0.61; 1.60-5.00)	3.20	0.10	0.07
Perceived benefits of EMH[b]	3.62 (0.64; 1.43-5.00)	3.71	−0.42	0.41
EMH proactive innovation	2.65 (0.91; 1.00-5.00)	2.60	0.39	−0.32
EMH self-efficacy	3.27 (0.77; 1.00-5.00)	3.33	−0.20	0.11
Study 2 (N=363)				
Total eMHAR Scale	3.30 (0.60; 1.07-5.00)	3.33	−0.19	0.19
Perceived benefits of EMH	3.65 (0.66; 1.00-5.00)	3.71	−0.55	0.49
EMH proactive innovation	2.76 (0.88; 1.00-5.00)	2.80	0.08	−0.52
EMH self-efficacy	3.42 (0.73; 1.00-5.00)	3.33	−0.51	0.70

[a]eMHAR: eMental Health Adoption Readiness.

[b]EMH: eMental health.

Internal consistency, calculated by Cronbach α, yielded satisfactory results for the total scale (.88) and the respective subscales (.73-.87, Table 4). The subscales were significantly correlated with each other (Table 4), suggesting that they share an underlying construct. Table 4 shows the Cronbach α values of the subscales, Pearson correlations, and corrected correlations between subscales.

Table 4. Internal consistencies and intercorrelations of the subscales.

Subscale	Perceived benefits of EMH[a]	EMH proactive innovation	EMH self-efficacy
Study 1			
Perceived benefits of EMH			
Coefficient	.83[b]	0.57[c]	0.45[c]
P value	—[d]	<.001	<.001
EMH proactive innovation			
Coefficient	0.48[e]	.87[b]	0.58[c]
P value	<.001	—	<.001
EMH self-efficacy			
Coefficient	0.35[e]	0.47[e]	.73[b]
P value	<.001	<.001	—
Study 2			
Perceived benefits of EMH			
Coefficient	.86[b]	0.56[c]	0.37[c]
P value	—	<.001	<.001
EMH proactive innovation			
Coefficient	0.48[e]	.87[b]	0.56[c]
P value	<.001	—	<.001
EMH self-efficacy			
Coefficient	0.29[e]	0.45[e]	.74[b]
P value	<.001	<.001	—

[a]EMH: eMental health.

[b]Reliability coefficient: Cronbach α (on diagonal).

[c]Correlation coefficients corrected for attenuation (above diagonal).

[d]Not applicable.

[e]Pearson correlations (below diagonal).

Convergent and Predictive Validity

To establish convergent validity, we analyzed correlations between the eMHAR Scale scores and constructs that have been associated with the adoption of EMH. As expected, the total mean score was positively correlated with all the relevant measures (Table 5). More specifically, a higher score on the eMHAR Scale was associated with higher perceived added value of EMH, higher feelings of competency in EMH, and higher perceived proficiency regarding various EMH skills and specific EMH tools. The scale also showed a predictive validity for the more frequent use of EMH tools.

We also explored the correlations between these measures and the three eMHAR subscales (Table 5). The strength of the relationship of the subscales corresponded to their respective intuitively related constructs: compared with the other subscales, the perceived benefits of the EMH subscale correlated highest with the perceived added value. In addition, the EMH self-efficacy subscale correlated highest with feeling of competency, perceived proficiency in various EMH skills, and specific EMH tools. EMH proactive innovation correlated the strongest with frequency of use. These results support the convergent validity of the eMHAR Scale. Importantly, the correlations with the score on the total scale were always higher than those with the scores on the individual subscales.

Table 5. Spearman rho correlations between scores on the total eMental Health Adoption Readiness Scale and the subscales, and related constructs in studies 1 and 2.

Subscale	Perceived added value	Feelings of competency	Perceived proficiency general	Perceived proficiency tools	Frequency of use
Study 1					
Total scale					
Correlation coefficient	0.54	0.63	0.69	0.47	0.50
P value	<.001	<.001	<.001	<.001	<.001
Perceived benefits of EMH[a]					
Correlation coefficient	*0.48*[b]	0.38	0.48	0.26	0.38
P value	<.001	<.001	<.001	<.001	<.001
EMH proactive innovation					
Correlation coefficient	0.44	0.56	0.57	0.45	*0.47*
P value	<.001	<.001	<.001	<.001	<.001
EMH self-efficacy					
Correlation coefficient	0.31	*0.58*	*0.64*	*0.46*	0.30
P value	<.001	<.001	<.001	<.001	<.001
Study 2					
Total scale					
Correlation coefficient	0.43	0.57	0.66	0.38	0.39
P value	<.001	<.001	<.001	<.001	<.001
Perceived benefits of EMH					
Correlation coefficient	*0.40*	0.34	0.46	0.21	0.27
P value	<.001	<.001	<.001	<.001	<.001
EMH proactive innovation					
Correlation coefficient	0.37	0.51	0.55	0.35	*0.40*
P value	<.001	<.001	<.001	<.001	<.001
EMH self-efficacy					
Correlation coefficient	0.16	*0.55*	*0.64*	*0.36*	0.22
P value	.002	<.001	<.001	<.001	<.001

[a]EMH: eMental health.

[b]Italics indicate correlations of the highest scoring subscale.

Validation Study 2

Confirmatory Factor Analysis

Following the same data screening process as in study 1, no missing values or violations of assumptions were found. In total, 11 cases out of the total sample of 363 exceeded the critical value for the Mahalanobis distance, but their responses again showed no suspicious values. Performing the factor analysis with and without these cases yielded almost identical results for the factor structure and loadings. On the basis of this result, all cases were retained for our final analyses.

The confirmatory factor analysis confirmed the 3-factor model with 15 items found in study 1, showing relatively good fit indices (χ^2_{87}=200.4, *P*<.001; CFI=0.96; RMSEA=0.062). The model provided satisfactory standardized factor loadings (ie, >0.40) for all items on their respective factors, as shown in Table 6.

Table 6. Standardized factor loadings (>0.40) of the 15 items in the 3-factor model.

Item	Factor 1: Perceived benefits of EMH[a]	Factor 2: EMH proactive innovation	Factor 3: EMH self-efficacy
1. eHealth fits well to my work as a health care professional	0.81	—[b]	—
2. Contact between health care professional and client always has to be face-to-face (R[c])	0.68	—	—
3. I expect that eHealth provides benefits to the care that I deliver	0.71	—	—
4. eHealth does not improve the care that I deliver (R)	0.73	—	—
5. eHealth does not fit the profession of a mental health care professional (R)	0.73	—	—
6. eHealth does not have any added value for my work as a mental health care professional (R)	0.64	—	—
7. eHealth is an indispensable part of the mental health care profession	0.50	—	—
8. I am involved in setting up initiatives for the development of new eHealth tools and applications	—	0.71	—
9. Compared to my colleagues, I use eHealth a lot	—	0.82	—
10. Compared to colleagues, I take a lot of initiative regarding eHealth	—	0.87	—
11. I have ideas about new eHealth tools and technologies that could be developed (eg, Virtual Reality, gaming, biofeedback)	—	0.60	—
12. In my work I try to stimulate colleagues to use eHealth	—	0.80	—
13. I have the skills that are necessary to apply eHealth in my work	—	—	0.77
14. Using eHealth tools comes easy to me	—	—	0.99
15. I have to learn new skills to start using eHealth (R)	—	—	0.41

[a]EMH: eMental health.

[b]Factor loading <0.3.

[c]R: reverse scored.

Psychometric Evaluation

Table 3 presents the descriptive statistics for the total scale and subscales. Multimedia Appendix 4 presents a histogram of the total scores.

Regarding the reliability of the questionnaire, measured by Cronbach α, the values found for the total scale (.88) and the respective subscales (.74-.87, Table 4) were almost identical to those in study 1, again providing evidence for satisfactory internal consistency. In addition, similar values were found for the intercorrelations between the subscales (Table 4), which were all significantly correlated with each other, further confirming that they share an underlying construct.

Convergent and Predictive Validity

Convergent validity was again evaluated by analyzing the correlations between the eMHAR Scale scores and associated constructs. As expected, the total score correlated positively with all the relevant measures (Table 5). More specifically, a higher score on the eMHAR Scale was associated with a higher perceived added value of EMH, higher feeling of competency for EMH, and higher perceived proficiency in EMH skills and for EMH tools. The scale also showed predictive validity for the more frequent use of EMH tools.

Next, we analyzed the correlations between these measures and the three eMHAR subscales (Table 5), which yielded the expected results, similar to study 1: compared with the other subscales, the perceived benefits of the EMH subscale correlated highest with the perceived added value, the self-efficacy subscale correlated highest with the feeling of competency, proficiency in general EMH skills, and for specific tools. EMH proactive innovation correlated the strongest with the frequency of use. These results confirmed the validity of the subscales.

Finally, we assessed the scale's sensitivity to detect changes in adoption readiness by comparing the total and subscale scores between the two samples, as changes were expected because of the increased EMH use following the COVID-19 pandemic. After establishing measurement invariance (see Multimedia Appendix 5 for the goodness-of-fit test statistics), the multigroup confirmatory factor analysis showed a significant estimated mean increase of 0.25 units between study 1 and study 2 for EMH self-efficacy (z=3.342; P=.001) but not for the other factors (perceived benefits of EMH: z=0.543; P=.59; and EMH

proactive innovation: z=1.862; P=.06), all in line with our expectations. The same results were found while comparing the unweighted mean scores for the total scale and subscales; we found significant differences for the EMH self-efficacy subscale (t_{793}=−2.79; P=.005) and a trend for the total scale (t_{793}=−1.87; P=.06) but not for the perceived benefits of the EMH subscale (t_{793}=−0.60; P=.55) or the EMH proactive innovation subscale (t_{793}=−1.76; P=.08).

Discussion

Principal Findings

EMH has been shown to offer many promising possibilities for mental health care delivery [4-7], as also recognized by many clients [49]. Despite this, before the COVID-19 pandemic, the adoption of EMH technologies by mental health care professionals remained low [8-10]. To facilitate research on the adoption of EMH, we set out to develop a valid instrument that can reliably assess a professional's readiness to adopt EMH, is independent of organizational setting and specific tools, and can be easily applied to larger groups. To date, such an instrument did not exist. The eMHAR Scale was based on a thorough consultation of both literature and professionals in the field. Initial testing with a large sample of the target population provided a meaningful 3-factor structure with good internal consistency and convergent validity. The structure and satisfactory psychometric properties were confirmed in a second sample, thereby supporting the scale's ability to assess adoption readiness for the EMH of mental health care professionals.

Investigating the underlying structure of the eMHAR Scale indicated the existence of three factors: perceived benefits and applicability of EMH, EMH proactive innovation, and EMH self-efficacy. Identifying these underlying components allows us to look at the determinants that are important for the adoption of EMH in a new way, providing information on how they are connected to each other, which aspects are general to the adoption of technologies, and which aspects are more specific to the field of mental health care.

For the first factor, perceived benefits and applicability of EMH, some of the items are similar to items in factors of previous models on the adoption of novel technologies (eg, the factor Performance Expectancy in the unified theory of acceptance and use of technology model [50]), indicating that the perception that a new technology adds value is a fundamental factor in its adoption. These items on perceived benefits in the eMHAR Scale were highly related to the items regarding the extent to which practitioners feel EMH is appropriate for mental health care delivery. This is in line with other research on the adoption of innovations in health care that reported that the innovation's compatibility with the practitioner's profession is an important determinant of its adoption [51]. The emergence of this factor with items of both topics suggests that the applicability of the technologies to care delivery constitutes an essential part of perceiving benefits for (mental) health care professionals, which seems to differentiate the adoption of health care technologies from consumer technologies.

The second factor, EMH proactive innovation, includes some items that resemble the construct of personal innovativeness, defined as the willingness of individuals to try out a new technology [52]. However, in the eMHAR Scale, these items were related to the items of proactive behavior such as being involved in EMH initiatives and stimulating the use of EMH among colleagues. The social aspect of these items bears some resemblance with the social influence factor of the unified theory of acceptance and use of technology model [50], but those items are mainly focused on using a technology to adhere to a subjective norm. This could indicate that in mental health care, the subjective norm is less important for the adoption of technology, whereas adoption is more influenced by a person's individual degree of proactivity regarding technologies.

The third factor concerns EMH's self-efficacy, that is, the feeling that you possess the necessary skills and knowledge to use EMH. Self-efficacy has been identified as a common determinant in research on the adoption of innovations in health care [51] as well as in other areas (eg, learning technologies [53]). It has also been found to be an important predictor of technology use in general [54] and for specific technologies such as computers [55]. Furthermore, research indicates that self-efficacy influences both behavior and evaluation; when someone's perceived ability to use a new technology increases, this has a positive effect on both actual use and how this use is evaluated [56]. Thus, it can be expected that higher levels of EMH self-efficacy are related to a more frequent use of EMH tools and to experiencing more benefits, in line with our findings that the EMH self-efficacy subscale showed substantial positive correlations with the use of EMH tools and with the perceived benefits and applicability of the EMH subscale. It should be noted that although the three identified factors showed these intercorrelations, the total score showed the strongest correlations with the convergent and predictive measures (ie, use of EMH tools), which suggests that the total score for adoption readiness might be the most informative measure.

Implications

The main goal of the eMHAR Scale is to measure professionals' readiness to adopt EMH and its underlying factors (ie, perceived benefits of EMH, EMH proactive innovation, and EMH self-efficacy). As these individual characteristics of professionals play a crucial role in the adoption process [12], acquiring a deeper understanding of adoption readiness is essential to truly grasp the intricacies of this process. There are several ways in which the eMHAR Scale could be used to achieve this goal. First, the scale could be used for descriptive purposes, providing knowledge on the current level of adoption readiness and its specific factors. These measurements could be used at both the individual and group levels, for example, within a particular organization or country. In addition, such assessments could concern a single measurement at a particular moment or they could be used in longitudinal studies to assess changes over time by repeating measurements at predetermined intervals.

Second, the scale could facilitate the development and assessment of interventions aimed at changing EMH adoption, such as a specific training program. The scale—and its subscales—could indicate to organizations which specific factors

are most opportune to target and in this way inform strategies to increase use of EMH in daily practice, as opposed to the one-size-fits-all approach that is currently standard practice [57]. For example, finding a relatively low score on the EMH self-efficacy subscale suggests that it might be worthwhile to focus on increasing skills and knowledge by organizing training on EMH skills, whereas a low score on perceived benefits of EMH suggests that actions could be taken to increase the awareness of their merits and advantages, or that perhaps the available EMH tools require improvement or refinement. In addition to informing the development of interventions, it could also be used to examine the effects of these interventions by comparing pre- and postintervention measurements.

Finally, the scale could contribute to more explanatory studies that aim to obtain a more in-depth understanding of EMH adoption by investigating how adoption readiness relates to other constructs or how it relates to different clinical contexts and therapeutic interventions. For example, studies could investigate relationships between scores on the eMHAR Scale and its subscales with factors such as gender and age, client population, (digital) skills, education, type of mental health care organization and profession, and specific EMH tools. In addition, scholars could examine the relationship between adoption readiness and actual use and the extent to which demographic or external factors influence this relationship. It would also be interesting to investigate how adoption readiness scores can be related to experienced barriers, drivers, and needs that have been reported in previous studies. Knowing how these experiences are influenced by the determinants of adoption readiness and actual use might provide a better understanding of why and under which circumstances they arise. In addition, the inclusion of this instrument across multiple studies will allow for comparison across different groups and settings, which facilitates a more cumulative science of the adoption of EMH tools.

While applying the eMHAR Scale, it is important to note that the scale is not intended to be used as a normative scale; a higher score is not per definition a better score. It is meant to provide a view of professionals' current status regarding their readiness to adopt EMH. Furthermore, the level of adoption readiness does not correspond one-to-one with actual use in practice (and hence, is not a direct measure of adoption according to our terminology). Although we found a substantial positive correlation, meaning that practitioners with higher eMHAR scores were more likely to use EMH than those with lower scores, actual use in practice is also influenced by several external factors, such as technical infrastructure and organizational support [11-13]. Hence, this should be considered while aiming to increase the actual use of EMH.

Strengths and Limitations

Although there were no established instruments that could be used to test the validity of the scale, the convergent and predicted measures showed expected relationships with the eMHAR Scale, supporting its external validity. In addition, finding the expected increase in scores on the EMH self-efficacy subscale between the two samples (pre-COVID-19 and a few months after the first lockdown) provided evidence for the sensitivity of the scale. A drawback associated with the unfolding of the COVID-19 pandemic and its associated physical distancing measures (eg, lockdown) that occurred between the two studies is the fundamental change in daily clinical practice that happened as a consequence [44]; therefore, a valid measure of test-retest reliability could not be established. As it is expected that the position of EMH in mental health care practice will continue to evolve over the coming years, a strict test-retest assessment will probably remain difficult to conduct in future work. On the other hand, the fact that the factor structure remained unchanged, even during this radical change in care practice, speaks to the robustness of the scale. We are also aware that the correlational and descriptive nature of this study design limits us in drawing causal inferences. Performing a study that applies the scale before and after an intervention for EMH adoption (eg, a specific training or providing guidelines for web-based treatment) might provide a more solid test for its sensitivity.

An important additional asset of this instrument is that it can be applied relatively quickly and unobtrusively, both on the internet and offline, thereby facilitating the systematic study of EMH adoption readiness. This is especially important for application in mental health care practice, where practitioners generally work under high time pressure, and experience the challenge of continuously delivering high-quality care to a growing population of people in need of care, while dealing with changing job demands in mental health care. Another asset is the scale's independence in organizational setting and specific technologies. Therefore, it is not restricted to the mental health care systems of particular countries or practitioners working in institutions, and its applicability will not be compromised by future technological developments.

A strength of the current validation studies in support of the scale's validity is that the building and testing of the model were conducted using two substantial samples (exceeding the recommended 10 participants per item [29,30]), and particularly that the results were robust despite the great change in the daily practice of mental health care that occurred during the time frame of the research. Another strength of the validation studies is that both samples consisted of mental health care professionals with a high variation in use and experience with EMH, professions within mental health care, and types of mental health care institutions. This benefits the generalizability of the sample toward the entire population of mental health care professionals in the Netherlands, and perhaps to similar countries, and thereby the generalizability of the results of the study across clinical approaches and contexts. Currently, the scale has only been examined in the Dutch language. Translation to other languages would allow for testing the validity of the scale internationally. As translation might lead to subtle differences in participants' responses (eg, subtle differences in semantics of terms may lead to a different interpretation of items), we plan to investigate whether the same factor structure will emerge. These translated scales could then be used to study whether findings can be generalized across other countries, including those with significantly different mental health care systems and technological infrastructures.

Conclusions

This paper presents the construction and validation of the eMHAR Scale, a measurement instrument to assess the EMH adoption readiness of mental health care professionals. Overall, the scale showed satisfactory characteristics and relationships with relevant concepts. These results suggest that the eMHAR Scale is robust, valid, and reliable. With this work, we aim to stimulate future research and use in practice, which we hope will lead to improved insights into the individual characteristics of professionals in adopting EMH and facilitate well-informed solutions for the adoption process from which professionals, clients, and mental health care as a whole can benefit.

Acknowledgments

WAIJ, JJPAB, and MAF conceived the study. MAF and JJPAB set up and distributed the survey. MAF conducted the data analysis and interpretation of the data and drafted the manuscript. WAIJ, JHDMW, YAWdK, JJPAB, and IMBB contributed to the critical revision of the paper. Each author listed on the manuscript has seen and approved the submission of this version of the manuscript and takes full responsibility for the manuscript. The authors thank the mental health care professionals for their participation, and Antal Haans and Martijn Willemsen for contributing with their statistical expertise. This work was supported by the Dutch Research Council, grant number 055.16.141, project title Serious Games Supporting Mental Health Professionals in Skill Development and Empathic Interactions in Online Counselling, under the Professional Games for Professional Skills program.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The total item pool in Dutch and English translation (items included in the final 15-item solution are italicized; reversed scoring items are indicated with R).
[DOCX File , 5985 KB - jmir_v23i9e28518_app1.docx]

Multimedia Appendix 2
Survey items for convergent measures (translated from Dutch).
[DOCX File , 5106 KB - jmir_v23i9e28518_app2.docx]

Multimedia Appendix 3
Histogram of the total scores in study 1.
[PNG File , 76 KB - jmir_v23i9e28518_app3.png]

Multimedia Appendix 4
Histogram of the total scores in study 2.
[PNG File , 77 KB - jmir_v23i9e28518_app4.png]

Multimedia Appendix 5
Tests of measurement invariance for the multigroup measurement model of the eMHAR Scale across the two studies.
[DOCX File , 5443 KB - jmir_v23i9e28518_app5.docx]

References

1. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
2. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. J Technol Hum Serv 2008 Jul 03;26(2-4):109-160. [doi: 10.1080/15228830802094429]
3. Riper H, Andersson G, Christensen H, Cuijpers P, Lange A, Eysenbach G. Theme issue on e-mental health: a growing field in internet research. J Med Internet Res 2010 Dec 19;12(5):e74 [FREE Full text] [doi: 10.2196/jmir.1713] [Medline: 21169177]
4. Andersson G. Internet-delivered psychological treatments. Annu Rev Clin Psychol 2016;12:157-179. [doi: 10.1146/annurev-clinpsy-021815-093006] [Medline: 26652054]
5. Feijt MA, de Kort YA, Bongers IM, IJsselsteijn WA. Perceived drivers and barriers to the adoption of emental health by psychologists: the construction of the levels of adoption of eMental health model. J Med Internet Res 2018 Apr 24;20(4):e153 [FREE Full text] [doi: 10.2196/jmir.9485] [Medline: 29691215]

6. Connolly SL, Miller CJ, Lindsay JA, Bauer MS. A systematic review of providers' attitudes toward telemental health via videoconferencing. Clin Psychol Scie Pract 2020 Jun 06;27(2):e12311. [doi: 10.1111/cpsp.12311]
7. Stoll J, Müller JA, Trachsel M. Ethical issues in online psychotherapy: a narrative review. Front Psychiatry 2019;10:993 [FREE Full text] [doi: 10.3389/fpsyt.2019.00993] [Medline: 32116819]
8. Glueckauf RL, Maheu MM, Drude KP, Wells BA, Wang Y, Gustafson DJ, et al. Survey of psychologists' telebehavioral health practices: technology use, ethical issues, and training needs. Prof Psychol Res Pract 2018 Jun;49(3):205-219. [doi: 10.1037/pro0000188]
9. van der Vaart R, Atema V, Evers AW. Guided online self-management interventions in primary care: a survey on use, facilitators, and barriers. BMC Fam Pract 2016 Mar 09;17:27 [FREE Full text] [doi: 10.1186/s12875-016-0424-0] [Medline: 26961547]
10. Hennemann S, Beutel ME, Zwerenz R. Ready for eHealth? Health professionals' acceptance and adoption of eHealth interventions in inpatient routine care. J Health Commun 2017 Mar;22(3):274-284. [doi: 10.1080/10810730.2017.1284286] [Medline: 28248626]
11. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of ehealth interventions: systematic review of the literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
12. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
13. de Grood C, Raissi A, Kwon Y, Santana MJ. Adoption of e-health technology by physicians: a scoping review. J Multidiscip Healthc 2016;9:335-344 [FREE Full text] [doi: 10.2147/JMDH.S103881] [Medline: 27536128]
14. Musiat P, Goldstone P, Tarrier N. Understanding the acceptability of e-mental health - attitudes and expectations towards computerised self-help treatments for mental health problems. BMC Psychiatry 2014 Apr 11;14:109 [FREE Full text] [doi: 10.1186/1471-244X-14-109] [Medline: 24725765]
15. Whitten PS, Mackert MS. Addressing telehealth's foremost barrier: provider as initial gatekeeper. Int J Technol Assess Health Care 2005 Oct 25;21(4):517-521. [doi: 10.1017/S0266462305050725] [Medline: 16262977]
16. Rogers EM. Diffusion of Innovations. 5th Ed. New York, NY: The Free Press; 2003.
17. Davies F, Shepherd HL, Beatty L, Clark B, Butow P, Shaw J. Implementing web-based therapy in routine mental health care: systematic review of health professionals' perspectives. J Med Internet Res 2020 Jul 23;22(7):e17362 [FREE Full text] [doi: 10.2196/17362] [Medline: 32706713]
18. Tonn P, Reuter SC, Kuchler I, Reinke B, Hinkelmann L, Stöckigt S, et al. Development of a questionnaire to measure the attitudes of laypeople, physicians, and psychotherapists toward telemedicine in mental health. JMIR Ment Health 2017 Oct 03;4(4):e39 [FREE Full text] [doi: 10.2196/mental.6802] [Medline: 28974485]
19. Cipolletta S, Mocellin D. Online counseling: An exploratory survey of Italian psychologists' attitudes towards new ways of interaction. Psychother Res 2018 Nov 09;28(6):909-924. [doi: 10.1080/10503307.2016.1259533] [Medline: 28068875]
20. Wells M, Mitchell KJ, Finkelhor D, Becker-Blease KA. Online mental health treatment: concerns and considerations. Cyberpsychol Behav 2007 Jun;10(3):453-459. [doi: 10.1089/cpb.2006.9933] [Medline: 17594270]
21. Mol M, van Genugten C, Dozeman E, van Schaik DJ, Draisma S, Riper H, et al. Why uptake of blended internet-based interventions for depression is challenging: a qualitative study on therapists' perspectives. J Clin Med 2019 Dec 30;9(1):91 [FREE Full text] [doi: 10.3390/jcm9010091] [Medline: 31905834]
22. Simms DC, Gibson K, O'Donnell S. To use or not to use: Clinicians' perceptions of telemental health. Can Psychol 2011 Feb;52(1):41-51. [doi: 10.1037/a0022275]
23. Rochlen AB, Beretvas SN, Zack JS. The online and face-to-face counseling attitudes scales: a validation study. Measure Eval Counsel Devel 2017 Oct 19;37(2):95-111. [doi: 10.1080/07481756.2004.11909753]
24. Vis C, Mol M, Kleiboer A, Bührmann L, Finch T, Smit J, et al. Improving implementation of eMental health for mood disorders in routine practice: systematic review of barriers and facilitating factors. JMIR Ment Health 2018 Mar 16;5(1):e20 [FREE Full text] [doi: 10.2196/mental.9769] [Medline: 29549072]
25. DeVellis RF. Scale Development: Theory and Applications. 2nd Ed. Thousand Oaks, CA: SAGE Publications; 2003.
26. Carpenter S. Ten steps in scale development and reporting: a guide for researchers. Commun Method Meas 2017 Nov 22;12(1):25-44. [doi: 10.1080/19312458.2017.1396583]
27. Pett MA, Lackey NR, Sullivan JJ. Making Sense of Factor Analysis. Thousand Oaks, CA: SAGE Publications; 2003.
28. Worthington RL, Whittaker TA. Scale development research. Counsel Psychol 2016 Jun 30;34(6):806-838. [doi: 10.1177/0011000006288127]
29. Hair Jr JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis. 7th Ed. Essex, UK: Pearson; 2014.
30. Tabachnick BG, Fidell LS. Using Multivariate Statistics. 5th Ed. Boston, MA: Pearson; 2007.
31. Kelloway EK. Structural equation modelling in perspective. J Organiz Behav 1995 May;16(3):215-224. [doi: 10.1002/job.4030160304]
32. Holgado–Tello FP, Chacón–Moscoso S, Barbero–García I, Vila–Abad E. Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. Qual Quant 2008 Sep 7;44(1):153-166. [doi: 10.1007/s11135-008-9190-y] [Medline: 30311153]

33. Horn JL. A rationale and test for the number of factors in factor analysis. Psychometrika 1965 Jun;30(2):179-185. [doi: 10.1007/bf02289447]
34. Velicer WF, Eaton CA, Fava JL. Construct explication through factor or component analysis: a review and evaluation of alternative procedures for determining the number of factors or components. In: Goffin RD, Helmes E, editors. Problems and Solutions in Human Assessment. Boston, MA: Springer; 2000:41-71.
35. Costello AB, Osborne J. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Pract Assessment, Res Eval 2005;10:7 [FREE Full text] [doi: 10.7275/jyj1-4868]
36. Thompson B, Daniel LG. Factor analytic evidence for the construct validity of scores: a historical overview and some guidelines. Edu Psychol Measur 1996 Apr 01;56(2):197-208. [doi: 10.1177/0013164496056002001]
37. Kaiser HF. The varimax criterion for analytic rotation in factor analysis. Psychometrika 1958 Sep;23(3):187-200. [doi: 10.1007/BF02289233]
38. Cattell RB. The scree test for the number of factors. Multivariate Behav Res 1966 Apr 01;1(2):245-276. [doi: 10.1207/s15327906mbr0102_10] [Medline: 26828106]
39. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. Psychol Method 1999 Sep;4(3):272-299. [doi: 10.1037/1082-989X.4.3.272]
40. Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical assessment instruments. Psychol Assess 1995 Sep;7(3):286-299. [doi: 10.1037/1040-3590.7.3.286]
41. Charles EP. The correction for attenuation due to measurement error: clarifying concepts and creating confidence sets. Psychol Method 2005;10(2):206-226. [doi: 10.1037/1082-989x.10.2.206]
42. Stata Statistical Software: Release 14. StataCorp LP, College Station, TX. 2015. URL: https://www.stata.com/ [accessed 2021-07-31]
43. Mundfrom DJ, Shaw DG, Ke TL. Minimum sample size recommendations for conducting factor analyses. Int J Test 2005 Jun;5(2):159-168. [doi: 10.1207/s15327574ijt0502_4]
44. Feijt MA, de Kort YA, Bongers IM, Bierbooms JJ, Westerink JH, IJsselsteijn WA. Mental health care goes online: practitioners' experiences of providing mental health care during the COVID-19 pandemic. Cyberpsychol Behav Soc Netw 2020 Dec;23(12):860-864. [doi: 10.1089/cyber.2020.0370] [Medline: 32815742]
45. Kline RB. Principles and Practice of Structural Equation Modeling. 4th Ed. New York, NY: The Guilford Press; 2015.
46. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equat Model Multidiscipl J 2002 Apr;9(2):233-255. [doi: 10.1207/S15328007SEM0902_5]
47. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2020. URL: https://www.r-project.org/ [accessed 2021-07-31]
48. Rosseel Y. lavaan: An R package for structural equation modeling. J Stat Soft 2012;48(2):1-36 [FREE Full text] [doi: 10.18637/jss.v048.i02]
49. Kruse CS, Krowski N, Rodriguez B, Tran L, Vela J, Brooks M. Telehealth and patient satisfaction: a systematic review and narrative analysis. BMJ Open 2017 Aug 03;7(8):e016242 [FREE Full text] [doi: 10.1136/bmjopen-2017-016242] [Medline: 28775188]
50. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003 Sep;27(3):425-478. [doi: 10.2307/30036540]
51. Fleuren MA, Paulussen TG, Van Dommelen P, Van Buuren S. Towards a measurement instrument for determinants of innovations. Int J Qual Health Care 2014 Jun 20;26(5):501-510 [FREE Full text] [doi: 10.1093/intqhc/mzu060] [Medline: 24951511]
52. Agarwal R, Prasad J. A conceptual and operational definition of personal innovativeness in the domain of information technology. Inf Syst Res 1998 Jun;9(2):204-215. [doi: 10.1287/isre.9.2.204]
53. Buchanan T, Sainter P, Saunders G. Factors affecting faculty use of learning technologies: implications for models of technology adoption. J Comput High Educ 2013 Jan 30;25(1):1-11. [doi: 10.1007/s12528-013-9066-6]
54. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352 [FREE Full text] [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]
55. Igbaria M, Iivari J. The effects of self-efficacy on computer usage. Omega 1995 Dec;23(6):587-605. [doi: 10.1016/0305-0483(95)00035-6]
56. Ellen PS, Bearden WO, Sharma S. Resistance to technological innovations: an examination of the role of self-efficacy and performance satisfaction. J Acad Mark Sci 1991 Sep;19(4):297-307. [doi: 10.1007/bf02726504]
57. Mair FS, May C, O'Donnell C, Finch T, Sullivan F, Murray E. Factors that promote or inhibit the implementation of e-health systems: an explanatory systematic review. Bull World Health Organ 2012 May 1;90(5):357-364 [FREE Full text] [doi: 10.2471/BLT.11.099424] [Medline: 22589569]

Abbreviations

CFI: comparative fit index

EMH: eMental health
eMHAR: eMental Health Adoption Readiness
RMSEA: root mean square error of approximation

Edited by R Kukafka; submitted 05.03.21; peer-reviewed by N Schulze, SL Lee; comments to author 24.04.21; revised version received 02.06.21; accepted 05.07.21; published 17.09.21.

Please cite as:
Feijt MA, de Kort YAW, Westerink JHDM, Bierbooms JJPA, Bongers IMB, IJsselsteijn WA
Assessing Professionals' Adoption Readiness for eMental Health: Development and Validation of the eMental Health Adoption Readiness Scale
J Med Internet Res 2021;23(9):e28518
URL: https://www.jmir.org/2021/9/e28518
doi:10.2196/28518
PMID:34533469

Original Paper

Conceptual Ambiguity Surrounding Gamification and Serious Games in Health Care: Literature Review and Development of Game-Based Intervention Reporting Guidelines (GAMING)

Simon Warsinsky[1], MSc; Manuel Schmidt-Kraepelin[1], MSc; Sascha Rank[1], MSc; Scott Thiebes[1], MSc; Ali Sunyaev[1], PhD

Department of Economics and Management, Karlsruhe Institute of Technology, Karlsruhe, Germany

Corresponding Author:
Ali Sunyaev, PhD
Department of Economics and Management
Karlsruhe Institute of Technology
Kaiserstr. 89
Karlsruhe, 76133
Germany
Phone: 49 72160846037
Email: sunyaev@kit.edu

Abstract

Background: In health care, the use of game-based interventions to increase motivation, engagement, and overall sustainability of health behaviors is steadily becoming more common. The most prevalent types of game-based interventions in health care research are gamification and serious games. Various researchers have discussed substantial conceptual differences between these 2 concepts, supported by empirical studies showing differences in the effects on specific health behaviors. However, researchers also frequently report cases in which terms related to these 2 concepts are used ambiguously or even interchangeably. It remains unclear to what extent existing health care research explicitly distinguishes between gamification and serious games and whether it draws on existing conceptual considerations to do so.

Objective: This study aims to address this lack of knowledge by capturing the current state of conceptualizations of gamification and serious games in health care research. Furthermore, we aim to provide tools for researchers to disambiguate the reporting of game-based interventions.

Methods: We used a 2-step research approach. First, we conducted a systematic literature review of 206 studies, published in the *Journal of Medical Internet Research* and its sister journals, containing terms related to gamification, serious games, or both. We analyzed their conceptualizations of gamification and serious games, as well as the distinctions between the two concepts. Second, based on the literature review findings, we developed a set of guidelines for researchers reporting on game-based interventions and evaluated them with a group of 9 experts from the field.

Results: Our results show that less than half of the concept mentions are accompanied by an explicit definition. To distinguish between the 2 concepts, we identified four common approaches: implicit distinction, synonymous use of terms, serious games as a type of gamified system, and distinction based on the full game dimension. Our Game-Based Intervention Reporting Guidelines (GAMING) consist of 25 items grouped into four topics: conceptual focus, contribution, mindfulness about related concepts, and individual concept definitions.

Conclusions: Conceptualizations of gamification and serious games in health care literature are strongly heterogeneous, leading to conceptual ambiguity. Following the GAMING can support authors in rigorous reporting on study results of game-based interventions.

(J Med Internet Res 2021;23(9):e30390) doi:10.2196/30390

KEYWORDS

game-based interventions; gamification; serious games; literature review; reporting guidelines; conceptual ambiguity

Introduction

Health interventions that draw on games as their inspirational source (ie, game-based interventions) come with the promise of increasing motivation, engagement, and overall sustainability of health behaviors [1] by enabling new paths of interactions between health care providers and patients [2]. Although games are present in various forms and genres, they generally share four defining traits [3]: a goal players try to achieve, rules that limit how players can achieve the goal, a feedback system telling players how they can achieve the goal, and voluntary participation of the players.

Over time, a myriad of different labels for game-based interventions have emerged, including playification [4], educational games [5], game-based learning [6], active video games [7], fitnessification [8], fitness games [9], exergames or serious exergames [10], cognitive games [11], simulation games [12], games with a purpose [13,14], persuasive information systems [15], and persuasive games [16]. Among these approaches to game-based health interventions, gamification and serious games have emerged as the 2 prevailing concepts [13,17,18]. Both concepts have been reported to be successfully used in various areas of health care, such as promoting healthier lifestyles, supporting rehabilitation processes, or fostering the education of health professionals [19,20]. However, although serious games have been researched long before the proliferation of computer technology and can be traced back to the 1970s [21], the term gamification only became popular around the 2010s [22] and has since then rapidly gained interest from health care researchers and professionals alike [18].

As both concepts have become increasingly established, scholars have developed a variety of definitions for the terms gamification and serious games [23,24]. In addition, several researchers have discussed substantial conceptual differences between gamification and serious games. For example, Liu et al [25] argued that serious games are full-fledged games that are modeled after but independent and separate from real-world systems, whereas gamification can never exist on its own and is always a part of a real-world system that maintains its instrumental functionality. Such conceptual considerations have been supported by empirical studies that showed differences in the effects of gamification and serious games on specific health behaviors [26]. Consistent with the law of parsimony, the differences in conceptual and empirical levels indicate that gamification and serious games are, in fact, 2 distinct concepts that require separate scholarly consideration [27,28]. However, researchers frequently report cases in which terms related to gamification or serious games are used ambiguously (eg, due to a lack of definitions [29,30]) or even interchangeably [31]. Over the course of our own research endeavors, we have also seen first indications that there is an amalgamation of the terms in the form of authors using them interchangeably or ambiguously. The fuzzy use of the 2 terms can yield negative consequences, such as an impediment to cumulative knowledge development [28,31] or the prevention of collaboration between researchers and practitioners [32]. Some of these issues are already visible in the literature. For example, Koivisto and Hamari [33] reported in their well-cited literature review that they had to consider every paper that was labeled as gamification by the authors, despite being aware of and acknowledging conceptual unclarity between the terms gamification and serious games in some research communities. To ensure the feasibility of the review, they had to perform some substantial abstraction, which "has consequently caused some specifics of the studies to be lost" [33]. In health care, negative effects stemming from conceptual unclarity may also ripple through to practice, as it is not uncommon for game-based health interventions to be built around theoretical knowledge [34,35].

Within health care literature, the extent to which existing research explicitly distinguishes between gamification and serious games and whether it draws on existing conceptual considerations to do so remains unclear. A major challenge for authors may also be that they see themselves faced with several different definitions and conceptual differences of gamification and serious games, each operating on a different level of abstraction, focusing on different characteristics. This can lead to contradictory indications and may make it difficult for the authors to reconcile different views [23]. We argue that more practical guidelines for authors to avoid ambiguities between different types of game-based interventions can help achieve a more accurate attribution of cognitive, affective, and behavioral outcomes to either gamification or serious games and ultimately advance the development of effective interventions. In this study, we aim to provide researchers and practitioners with the necessary tools. To develop such tools, it is necessary to capture the status quo of how health care researchers understand, define, and use the terms gamification and serious games and whether they explicitly distinguish between the 2 concepts. Consequently, we formulated the following research questions (RQs):

RQ1: How does extant research conceptualize gamification and serious games in health care?

RQ2: How does extant research distinguish between gamification and serious games in health care?

RQ3: How can authors be guided to avoid conceptual ambiguity when reporting game-based intervention studies?

To answer our RQs, we use a 2-step research design. In step 1, we analyze the current understanding of the terms *gamification* and *serious games* in health care research by reviewing the literature in the *Journal of Medical Internet Research* and its sister journals. By doing so, we provide a comprehensive overview of existing conceptualizations of gamification and serious games in health care (RQ1) and provide rich insights into the nature of the conceptual unclarities surrounding gamification and serious games (RQ2). In step 2, we derive guidelines for authors of game-based intervention studies that may support them in avoiding conceptual ambiguity when reporting their results (RQ3).

Methods

Literature Review

To answer the first two RQs, we conducted a systematic web-based database search following the guidelines by Levy

and Ellis [36]. We limited the search to journals of JMIR Publications, as the *Journal of Medical Internet Research* itself is one of the leading medical informatics and health care sciences and services journal with an impact factor of 5.43 [37] and a broad area of topics is covered in more than 30 of its sister journals. To identify relevant literature, we used the search tool on the *JMIR* website [38] to search for papers containing terms related to either gamification (search terms: *gamification*, *gamified*, *gamifying*, or *gamify*) or serious games (search terms: *serious AND games*, *serious AND gaming*, or *serious AND game*) in any field. The search was performed on November 12, 2020, and yielded 271 studies, for which we screened their full text to assess their relevance. In doing so, we excluded 65 studies that did not focus on either gamification or serious games but, for example, only mention them in the keywords of the paper [39] or as a possible future research avenue [40]. We presumed that all articles published in a journal of JMIR Publications belong to the research area of health care (ie, are related to research on efforts made to maintain or restore physical, mental, or emotional well-being [41]). Hence, we were left with 206 studies for further analysis. Our data analysis followed a concept-centric approach informed by Webster and Watson [42]. For each relevant study, we conducted a full-text analysis to identify the applied research methods, health care context (based on a study by Stepanovic and Mettler [19]), and applied conceptualizations of the terms gamification or serious games. Regarding the conceptualization of gamification and serious games, we first analyzed whether and, if so, how each of the studies defined the terms. For the purpose of this study, we broadly construed a definition as a phrase that conveys the meaning of a term. Furthermore, we analyzed the inspiration sources of each provided definition (ie, whether they were based on extant literature or self-developed). In addition, we analyzed whether and how the studies differentiated between gamification and serious games. The full results of our concept-centric data analysis, including all 206 papers, can be found in Multimedia Appendix 1 [1,2,10,18,40,43-243].

Development of Reporting Guidelines

To develop the reporting guidelines, we used a 3-step approach inspired by Moher et al [244]. First, one of the authors generated an initial list of items for inclusion in the checklist based on the insights of the literature review, as well as our personal experience in the field. These items cover both best practices we selected from studies that clearly define and distinguish serious games and gamification, as well as common pitfalls, inconsistencies, and inaccuracies identified within the reviewed literature. The resulting list was grouped into sets of related items. Second, the initial list of items was discussed by all authors in an iterative process to ensure a common understanding. During this step, items were merged, divided, or specified more precisely as required to develop a comprehensive yet concise list of guidelines. In addition, we developed exemplary statements to provide future authors with concrete starting points for reporting game-based interventions. Third, to pilot test our checklist, we sent it to 9 experts in the field of game-based interventions and asked for feedback regarding comprehensibility and completeness. We then consolidated their feedback and incorporated them into guidelines, where feasible. This resulted in adjustments of 6 items (1a, 1b, 2a, 2b, 2c, and 3.1b), as well as the addition of 2 new items (2d and 2e).

Results

Conceptualizations of Gamification and Serious Games in Health Care Literature

Characteristics of the Included Studies

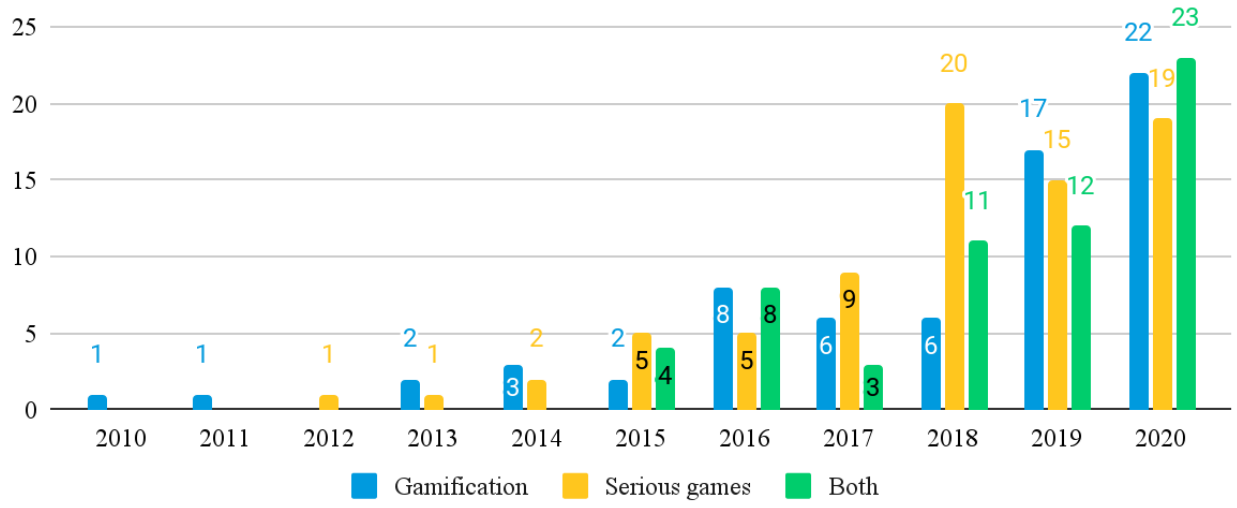
The reviewed articles were published in the *Journal of Medical Internet Research* and 10 of its sister journals. Table 1 provides an overview of the number of articles published in each journal. In total, 38.8% (80/206) of studies applied quantitative methods, with the most common individual method being intervention trials (67/206, 32.5%). Another 18.9% (39/206) of studies applied qualitative methods, whereas 30% (62/206) of studies were conceptual in nature. The remaining 12.1% (25/206) of studies used mixed methods. Regarding the health care context [19], a majority of studies (99/206, 48.1%) focused on individual lifestyle habits. The most prominent theme in this context is achieving an increase in the physical activity of users, as examined by 7.3% (15/206) of studies [10,43,44]. Chronic disease management and rehabilitation is covered by 38.3% (79/206) of studies, including studies on dementia [45,46], diabetes [47,48], or depression [49,50]. Furthermore, 23.3% (48/206) of studies focused on supporting health professionals and mostly dealt with education on various topics, such as emergency care [51], specific surgical procedures [52,53], or teaching ethics in medicine [54].

Table 1. Number of articles mentioning terms related to gamification, serious games, or both for each journal of JMIR Publications (N=206).

Outlet	Gamification, n (%)	Serious games, n (%)	Both, n (%)	Total, n (%)
JMIR Serious Games	21 (10.2)	38 (18.4)	34 (16.5)	93 (45.1)
Journal of Medical Internet Research	11 (5.3)	18 (8.7)	10 (4.9)	39 (18.9)
JMIR mHealth and uHealth	22 (10.7)	1 (0.5)	4 (1.9)	27 (13.1)
JMIR Research Protocols	14 (6.8)	8 (3.9)	3 (1.5)	25 (12.1)
JMIR Mental Health	1 (0.5)	1 (0.5)	6 (2.9)	8 (3.9)
JMIR Formative Research	5 (2.4)	0 (0)	1 (0.5)	6 (2.9)
JMIR Rehabilitation and Assistive Technologies	1 (0.5)	2 (1)	1 (0.5)	4 (1.9)
JMIR Pediatrics and Parenting	1 (0.5)	0 (0)	0 (0)	1 (0.5)
JMIR Human Factors	1 (0.5)	0 (0)	0 (0)	1 (0.5)
JMIR Medical Education	0 (0)	1 (0.5)	0 (0)	1 (0.5)
JMIR Public Health and Surveillance	1 (0.5)	0 (0)	0 (0)	1 (0.5)

The 206 studies we reviewed were dated from 2010 to 2020. Across these 206 studies, 137 (66.5%) feature terms related to gamification (*gamification*, *gamified*, *gamifying*, or *gamify*), whereas 128 (62.1%) feature terms related to serious games (*serious AND games*, *serious AND gaming*, or *serious AND game*) in their full text. There is an overlap of 28.6% (59/206) studies that mention terms related to both concepts. Figure 1 shows the distribution of articles that mention terms related to each concept over time. As can be seen in Figure 1, both concepts became increasingly featured in JMIR journals since 2010, with no concept having a clear edge over the other. Out of the 137 studies that feature gamification, 61 (44.5%) explicitly define the term, contrasted by 76 (55.5%) studies that do not. As for serious games, of the 128 studies, 60 (46.9%) that feature the term provide an explicit definition, whereas 68 (53.1%) do not.

Figure 1. Number of articles in JMIR journals mentioning terms related to gamification, serious games, or both per year.



References to Definitions From Extant Literature

Of the 61 studies that define gamification, a share of 52 (85%) studies explicitly base their definitions on extant literature, whereas 9 (15%) studies provide a self-developed definition. In contrast, out of the 60 studies that define serious games, 50 (83%) explicitly refer to extant literature to do so, whereas 10 (17%) studies do not. In defining gamification, our reviewed papers draw on 29 different sources, whereas for serious games, we found 27 different sources of definitions. Figures 2 and 3 illustrate the different sources our review papers draw on to define gamification or serious games, respectively. Each node represents one such source. The number of times a source has been cited by our review papers is indicated by both the color and size of the corresponding node. An arrow from one node to another indicates that a study draws on another, specifically when defining gamification or serious games.

Figure 2. Identified gamification definition papers.

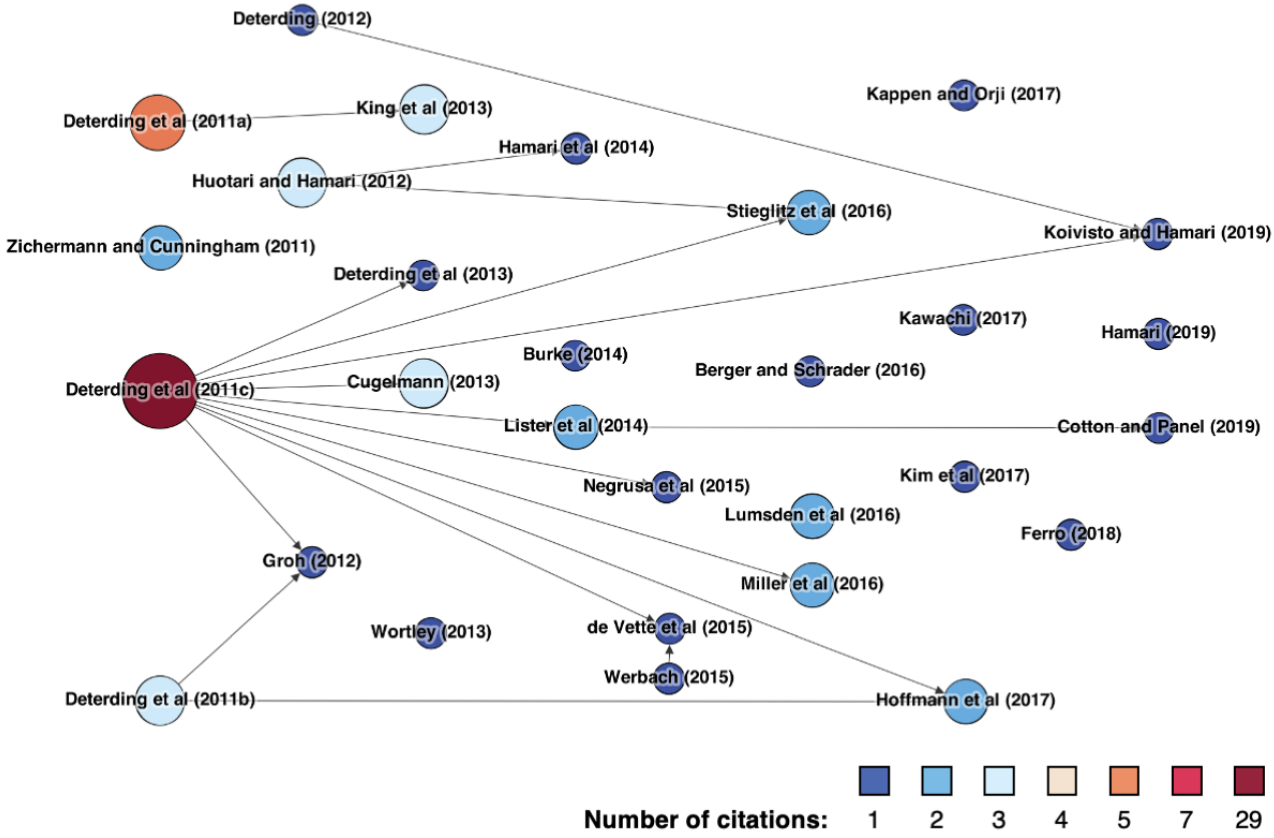
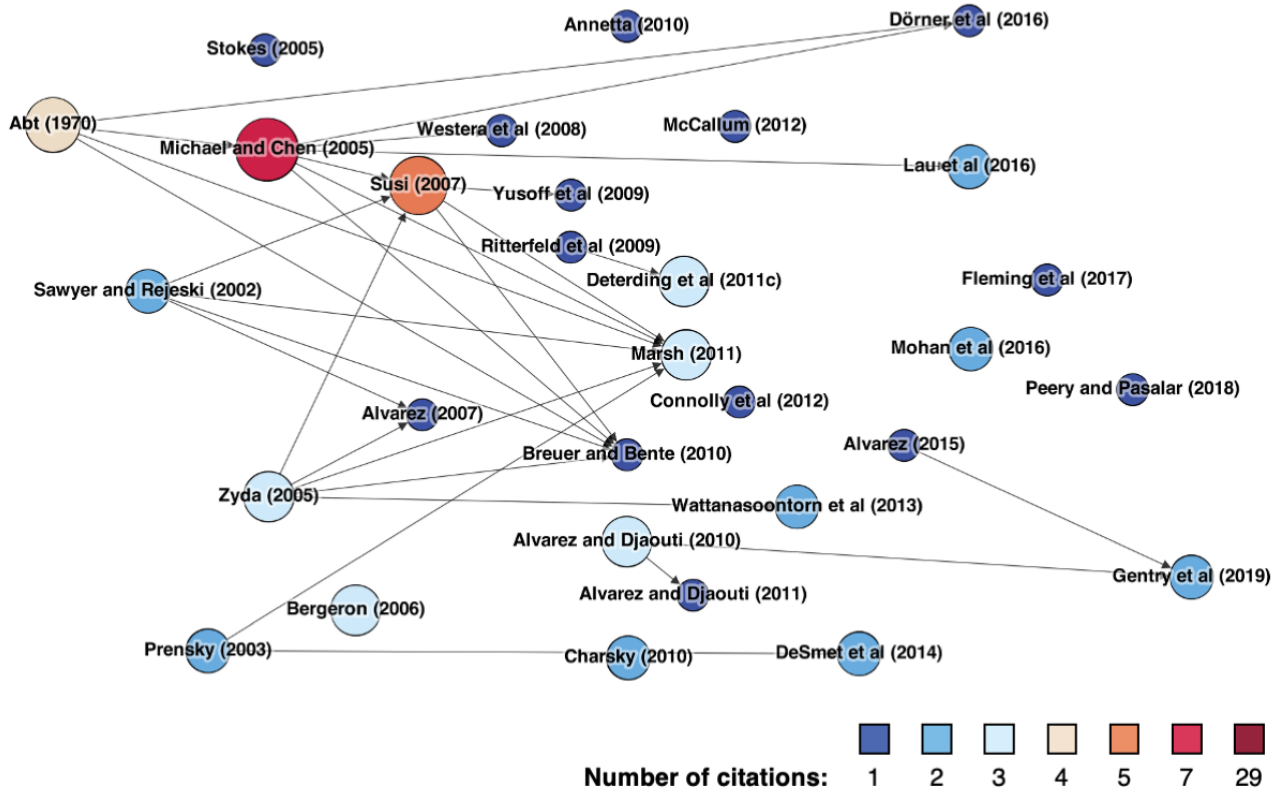


Figure 3. Identified serious games definition papers.



The most commonly cited source for a gamification definition, cited by 14.1% (29/206) of our reviewed papers, is the seminal paper by Deterding et al [245], who define gamification as "the use of game design elements in nongame contexts." Considering indirect citations (as indicated by arrows in Figure 2), the percentage value increases to 27.2% (56/206), thus forming the vast majority. Another definition by Deterding et al [246] was used by 2.4% (5/206) of the reviewed papers. Huotari and

Hamari [247], Cugelman [1], King et al [248], and Deterding et al [249] are each cited by three (3/206, 1.5%) of the reviewed papers, whereas Zichermann and Cunningham [250], Lister et al [55], Miller et al [251], Lumsden et al [56], Stieglitz et al [252], and Hoffmann et al [57] are each referenced twice (2/206, 1%).

In contrast to gamification, the sources of serious games definitions in our reviewed literature seem to be more fractured and heterogeneous. The most cited paper is that by Michael and Chen [14], cited by 3.4% (7/206) of our reviewed papers. They define serious games as "games that do not have entertainment, enjoyment, or fun as their primary purpose." In addition to the seven direct citations, Michael and Chen [14] are also cited indirectly 6.3% (13/206) of the time in the definitions of our review papers of serious games (as indicated by arrows in Figure 3). Other prominent sources (based on citation count) include studies by Susi et al [24] (5/206, 2.4%) and Abt [253] (4/206, 1.9%), as well as Deterding et al [245], Zyda [254], Alvarez and Djaouti [255], Bergeron [256], and Marsh [257] (3/206, 1.5%). The remaining sources were cited only once (1/206, 0.5%) or twice (2/206, 1%).

Conceptions of Gamification in Health Care

We found that the content of the definitions provided for gamification in the reviewed papers represent several different conceptions. Of the 61 papers defining gamification, 58 (95%) contained only a single definition of the term. Among these articles, 74% (43/58) shared the basic notion of Deterding et al [245,246,249]. They define gamification as either "the use of game design elements in non-game contexts" [245,249] or as an "umbrella term for the use of video game elements (rather than full-fledged games) to improve user experience and user engagement in nongame services and applications" [246]. Both definitions conceptualize gamification as the intentional use of game (design) elements in some kind of nongame artifact. The latter definition, however, explicitly includes the purpose of gamification as improving user experience and user engagement. Similarly, among the 43 papers conceptualizing gamification in line with the basic notion of Deterding et al [245,249], 33 (77%) did not explicitly specify the purpose of gamification. The other 23% (10/43) of papers all specify a variation of engagement, motivation, or both as the purpose of gamification [58-60]. There are also differences regarding the inspiration that gamification draws from games. Although most of the reviewed papers adopt the term game (design) elements from the definition by Deterding et al [245,249], others instead name game (design) techniques [61-64], game principles [65], playful elements [66], game components [67], or game design features [68]. Similarly, the views differed in the artifacts that gamification is applicable to. Instead of nongame contexts, some authors speak of nongame settings [43,69], nongame environments [59], or nongame mechanisms [70]. Others more generally refer to "real world processes or problems" [71], "nongameful or nongamelike situations, services, or tools" [62], or "applications that were not games to begin with" [72].

Another 15 articles provided a single definition that differs from the notion of Deterding et al [245] in a more substantial way. A group of 40% (6/15) of papers specify the nongame context of gamification as either health care [40,73-76] or education [18]. For example, Mendiola et al [75] described gamification as a "feature that offers points, badges, or movement through levels as a health objective is achieved or the more a patient is engaged." Park and Kim [18] state that "gamification in education applies game elements to an educational context." Furthermore, 47% (7/15) of articles neither narrowed the context down in a general way nor mentioned a specific context [63,77,78]. For instance, Brown et al [77] refer to gamification as "the application of game design elements to engage and motivate users." The remaining 13% (2/15) of papers describe gamification as a process to create serious games [79] or focus on gamification as the "overall proliferation of games in culture, society, and technology" [80]. Finally, three papers provide more than one definition of gamification. Among these 3 papers, 2 (67%) [81,82] contrast the aforementioned definition by Deterding et al [245] with an alternative definition that describes gamification as "a process of enhancing a service with affordances for gameful experiences in order to support user's overall value creation" [247]. Cheng et al [81] argue that the latter definition is more useful in an mHealth context, as it focuses on the goal of gamification instead of its method. Zakaria et al [54] provided various conceptualizations of gamification throughout the article without contrasting them.

Conceptions of Serious Games in Health Care

Regarding the content of serious games definitions, we first identified different views on how much serious games draw from games. Most studies simply state that serious games "are games" [83,84], whereas some expand on this by limiting serious games to, for example, "full-blown digital games" [85]. Brown et al [86] used this full game characteristic as the sole basis for their self-developed definition, as they define serious games as "the use of games in their entirety (as opposed to selected elements or individual features of a game)." On the contrary, some authors conceptualize serious games not necessarily as full games but rather as "game-like experiences" [87], "the use of game design elements" [58], or as something that "[uses] specific design principles derived from video games" [88]. Some authors even necessitate the presence of specific game elements in a serious game, such as a challenge [89,90], a responsive narrative [91,92], or interactive elements [93]. Four studies also mention that serious games must be situated in a nongame context [58,86,89,94].

Furthermore, we found that 90% (55/61) of papers that provided a definition of serious games relate to the purpose of a serious game. They do so in any of the following 3 ways. First, 15% (8/55) of studies neutrally state that a serious game needs to have any kind of purpose, also described as a characterizing goal [95]. Second, 40% (22/55) of studies state that a serious game needs to have a specific purpose. For example, Mack et al [96] refer to serious games as "games designed to fulfill a serious purpose by providing education from health professionals via a digital device." Other mentioned purposes include imparting in a user the skills, knowledge, or attitudes that are applicable in the real world [89], or to improve motivation when completing or addressing complex or bothersome tasks [88]. Third, the remaining 45% (25/55) of studies explicitly exclude specific purposes from being the main

purpose of a serious game. This nearly unanimously involved excluding entertainment [87,97] or enjoyment [98] as the main purpose of a serious game. Only Gamito et al [99] pass over entertainment and instead state that "serious games are games designed for other purposes than gaming."

There were also different views in the reviewed literature on the role of entertainment in serious games. As stated earlier, most studies exclude entertainment as the main purpose of a serious game [87,97]. Nevertheless, some authors still consolidate entertainment as a core part of a serious game by explicitly mentioning that serious games "use entertainment" [100] or "[bring] fun spring from video games" [101] or by referring to serious games as "entertaining games" [102,103]. In contrast, Vilardaga et al [97] refer to serious games as "games not for entertainment," possibly excluding entertainment as part of a serious game altogether. Furthermore, some authors [95,104,105] seem to consider serious games as an augmented form of (entertainment) games in the sense that they do "not simply [provide] entertainment" [87] or "merely entertain" [106] but rather are a "combination of serious aspects with fun" [101] or "intended to be both entertaining and educational" [103]. The studies by Andrade Ferreira et al [45] and Bindoff et al [107] most vividly represent this view, as they start with entertainment games as a basis, and then outline the differences between serious games and entertainment games (eg, an explicit purpose).

Finally, we identified several aspects that were unique to only one or two definitions. DeSmet et al [103] posit that serious games are "a form of organized play." Similarly, according to Shiyko et al [94], serious games use playful design strategies. The authors also stated that the term serious games is a synonym for a variety of terms, including gamification [58], applied games [108], games with a purpose [107], and transformational games [45]. In 2 studies, Tuti et al [51,109] posited that serious games have to be playable with a mobile phone. Robert et al [100] remarked that a serious game is a contest played against a computer and thus involves only 1 player.

Distinctions

Overall, we identified 59 articles that mention both the terms *gamification* and *serious games* in their full texts. We examined these articles to determine the relationship between the 2 terms used by the authors in each case. To start, 14% (8/59) of articles did not allow for an interpretation of the semantic link between the 2 terms (eg, the terms were used without reference to each other in completely different sections of the article). Within the remaining 86% (51/59) of articles, we identified four groups of articles with distinct approaches: (1) implicit distinction, (2) synonymous use of terms, (3) serious games as a type of gamified system, and (4) distinction based on the full game dimension. The 4 approaches are outlined in detail in the following sections.

The first group contained 29% (15/51) of articles that only implicitly differentiated between the terms *gamification* and *serious games*. The authors of these articles use the 2 terms in relation to each other in such a way that it becomes clear that they differentiate between the two terms without providing any explanations for why and how to do so. Typical examples of this approach include articles that use and differentiate the terms in enumeration. For example, Park and Kim [18] state "[b]efore gamification, educational games, game-based learning, and serious games were applied to the classroom. As gamification was defined, however, it has become preferred among instructors compared with the other techniques." Similarly, Martin et al [74] write "[I]n all, 2 different persuasive mechanisms—gamification and serious gaming—were proposed."

The second group contained 10% (5/51) of articles that used the terms gamification and serious games interchangeably. For example, Booth et al [58] state that "[s]erious gaming or gamification has been defined [...] as 'the use of game design elements in nongame contexts' for the purposes of engaging learners in solving complex problems." Although some authors use the terms synonymously without making it a subject of discussion, others show awareness of conceptual ambiguity. For example, Lumsden et al [56] report: "[There is a] lack of coherence in the field [...] partly due to poor definition of terms; for example, the gamelike tasks covered in this review could be described as 'serious games,' 'gamelike', 'gamified', 'games with a purpose', 'gamed-up', or simply 'computer based'."

The third group contains 43% (22/51) of papers that regard serious games as a type of gamified system. Many authors refer to gamification as the process of developing a game-based intervention. For example, Zhang et al [68] stated that "the gamification approaches used included the addition of gaming elements to existing tasks, transformation of a conventional task into a serious game." Furthermore, while some of these articles focus on serious games as the only possible outcome of gamification, many other authors seem to share the view that other types of game-based interventions can also be the outcome of gamification. For example, Fornasini et al [44] reported that "a recent trend has revealed virtual physical training through exergames and serious games, making gamification an effective tool to motivate." In this regard, serious games are also sometimes referred to as systems that integrate "gamification principles and gamification strategies" [110]. However, it mostly remains unclear what is meant by these terms and how they differ from other established terms such as game mechanics, game dynamics, and game aesthetics [258].

The last group contains 18% (9/51) of articles that distinguish between gamification and serious games based on whether the system in consideration is regarded as a full game. For example, de Vette et al [111] state that "gamification is defined as the use of elements from games in nongame contexts to improve user experience and engagement without making that system a full game as is the case with serious games." However, it largely remains unclear what the criteria are that determine whether a system is a full-fledged game or *only* gamified. Some authors try to counteract this by giving examples of gamification and serious games that are as vivid as possible. For example, Vermeir et al [60] state that "an interactive world in which players complete challenges designed to improve physical activity" is a typical example for a serious game, whereas a gamified system could be "a mobile health application that uses points and badges to encourage physical activity." Nevertheless, these authors also report that, in practice, the actual distinction

between the 2 concepts can be blurry and highly subjective because of a lack of clear criteria [60].

Game-Based Intervention Reporting Guidelines

General Remarks About the Guidelines

Through our review, we found that not only are the concepts of gamification and serious games construed in various ways but also their relationship is often fuzzy and unclear. To reduce this conceptual ambiguity, we compiled reporting guidelines consisting of 25 items that are considered essential when reporting studies that deal with game-based interventions and grouped them into 4 topics. In the following sections, we present these items, each accompanied by a description and, where applicable, exemplary statements. A consolidated version that can be used as a quick reference can be found in Multimedia Appendix 2 [245]. A more detailed version including detailed rationales for each individual item may be found in Multimedia Appendix 3 [245].

Game-Based Intervention Reporting Guidelines (GAMING) should be interpreted with some consideration. GAMING are intended to support the reporting of studies on game-based interventions. Although some of the items in GAMING may be beneficial to consider when developing a game-based intervention, our guidelines are not intended to guide the development of game-based interventions per se. GAMING are also intended only for studies that deal with game-based interventions. Accordingly, when we use the term concept in our guidelines, we refer to a type of game-based intervention (eg, gamification, serious games, or exergames). We further differentiated between those concepts that are the thematic focus of the study (henceforth termed *core concepts*) and those that only have auxiliary roles (henceforth termed *related concepts*). An example to clarify this: a study may focus on the development of a physical activity intervention based on the paradigm of serious games (core concept) but still compare its results to gamified physical activity interventions or exergames (related concepts). In the following sections, we briefly introduce the contents of our guidelines and the rationales for including them.

Topic 1: Conceptual Focus

The first topic contains items that ensure that the conceptual focus of a study is clear to the authors and readers alike (Table 2).

Table 2. Game-Based Intervention Reporting Guidelines—items in topic 1: conceptual focus.

Number and topic	Description	Exemplary or explanatory statements
1. Conceptual Focus		
	(a) Decide which concepts (ie, gamification or serious games) best reflect the interventions you want to investigate.	See Multimedia Appendix 4 [17,25,29,31,245,259-261] for existing theoretical distinctions between serious games and gamification.
	(b) Clearly state early in the paper which core concepts (ie, gamification or serious games) you focus on in your study and which criteria the decision for this core concept was based on.	"In this study, we focus on the concept of gamification, because we wanted to bring single game elements into the intervention instead of developing a full-fledged game."
	(c) Supply only metadata (eg, title and keywords) that corresponds to your core concepts.	For a study developing a physical activity intervention based on gamification: Gamification, gamified intervention.

Ultimately, it should be transparent for the reader, what kind of game-based interventions a study focuses on. A precondition for this is that the authors themselves understand what concepts best reflect the design of the game-based interventions they want to focus on (item 1a). In light of the plethora of existing concepts, as well as the different operationalizations of each, this is by no means a trivial task. Authors may find it beneficial to draw upon concept distinctions from extant literature to establish their conceptual focus. To decide between gamification and serious games, we compiled several distinctions from extant literature that authors may use as decision support (Multimedia Appendix 4). After deciding on their core concepts, the authors should communicate this decision early in their manuscript and include information on how the decision was reached (1b). This allows the reader to comprehend their decisions and grasp the study's conceptual focus. Furthermore, we suggest that authors should be mindful of the metadata they supply their study with (1c). When reviewing the extant literature, researchers often use metadata to initially assess the conceptual focus of a study and subsequently derive its relevance for their own purposes. Although we acknowledge that broadly diversified metadata increases the visibility of a study, it can also blur the conceptual focus of a study and increase noise for meta-studies or literature reviews.

Topic 2: Contribution

The items that we propose for the second topic (Table 3) are intended to clarify the knowledge contribution of a study to both authors and readers.

We believe that there are various research streams proximal to different concepts. What constitutes a research stream can be interpreted in various ways. Broadly, we understand a research stream as a set of studies with similar properties, such as their health care context (eg, all studies that describe the design of gamified physical activity interventions), or their applied methodology (eg, all studies that conduct literature reviews on serious games).

Table 3. Game-Based Intervention Reporting Guidelines—items in topic 2: contribution.

Number and topic	Description	Exemplary or explanatory statements
2. Contribution		
	(a) Decide which research stream within the focused concepts your study contributes to.	"Our work contributes to a better understanding of how gamification is being applied in real-world mHealth apps."
	(b) Report which research streams your study contributes to and which criteria the decision for research streams was based on.	"We contribute to a better understanding of the specific game element of leaderboards. We decided for leaderboards, as they are prominently used in mHealth apps to elicit social comparison."
	(c) Clarify your study's contributions to the chosen research streams, including the boundaries of your study.	"We provide rich insights into the psychological effects of leaderboards on patients when isolated from other game elements. [...] Our insights are about leaderboard specifically and are not necessarily transferable to other social comparison features."
	(d) Clarify your study's contributions to solving a problem or need in practice or society.	"The results of our study can support the design and implementation of successful physical activity mHealth apps in practice."
	(e) Report to which extent observed positive and negative outcomes can be attributed to your game-based interventions. If possible, narrow down the attribution of outcomes based on individual components of your game-based intervention (eg, game elements).	"Participants showed increased motivation, because they were able to compare themselves to others via the leaderboard function. However, this function also caused some participants to feel less competent, thus decreasing their motivation."

Each research stream comes with its own thematic emphasis and viewing angle. Thus, the transferability of knowledge between different research streams requires care and consciousness. For instance, insights from studies focusing on physical activity are often not transferable to studies focusing on health professionals' education [19]. Similarly, knowledge about badges is not necessarily transferable to leaderboards, although both approaches may be labeled as gamification. Thus, we think that authors should consciously decide which research streams they contribute to (2a). To facilitate transparency, the authors should communicate their decision to their readers (2b). By also including insights into the decision process (eg, explaining the criteria that shaped the decision for a research stream) in their manuscript, readers can gain a better understanding of a study's setting and on the generalizability and transferability of its results. After deciding on a research stream, the authors should clarify their study's contribution to the knowledge base of the research stream, including the boundaries of their study (2c). By doing so, readers can understand how the study contributes to a better understanding of core concepts. Furthermore, the ultimate goal of any intervention is usually to solve a problem in practice (eg, sedentary lifestyles). Accordingly, to foster an understanding of how specific game-based interventions can help to solve these problems, we recommend that the authors clarify their study's practical contribution (2d). Finally, for an understanding of game-based interventions, it is vital that the outcomes of an intervention are reported and attributed to the chosen type of game-based intervention in an as granular way as possible (2e). In this way, readers can easily contextualize knowledge and understand whether an outcome is, for example, evoked by a specific game element (eg, leaderboards) or connected to a specific affective outcome (eg, envy). We also encourage authors to embrace negative outcomes of their game-based interventions, as these, despite them usually not being the desired outcome of an intervention, can notably contribute to a better understanding of a concept [262].

Topic 3: Mindfulness About Related Concepts

Items in the third topic (Table 4) are concerned with ensuring that authors are mindful of the conceptual ambiguities surrounding game-based interventions and possible consequences for their study.

Table 4. Game-Based Intervention Reporting Guidelines—items in topic 3: mindfulness about related concepts.

Number and topic	Description	Exemplary or explanatory statements
3. Mindfulness about related concepts		
3.1. Introduction and use of related concepts		
	(a) Make efforts to identify possibly related concepts prominent in the context of your study.	Prominent related concepts for gamification in physical activity: Exergames, active video games, fitness games.
	(b) Mention only those related concepts that are substantive for your study.	'Substantive' in the sense that a research design necessitates the introduction of a concept. Example: A research design contrasting the effects of two game-based intervention concepts requires the introduction of both concepts.
	(c) Be mindful about nuanced terms in the domain of any introduced concept and use established vocabulary precisely.	Examples: Game design elements, game mechanics, gamification elements, ...
	(d) Avoid using related concepts interchangeably. If you use an umbrella term, specify which terms it comprises and clarify why you introduce it.	"To allow a better readability of the manuscript, we use the term activity games to describe gamified physical activity interventions, serious games for physical activity, as well as exergames."
3.2. Insights from extant literature		
	(a) Be mindful about conceptual ambiguities when drawing on the literature about game-based interventions.	—[a]
	(b) Do not presume easy transferability of insights from one concept to another.	Example: Drawing on serious games literature for a gamification-based intervention (or vice versa).
	(c) Specify precisely what you draw from the literature and why these insights are applicable to your study.	"Serious games and gamification share that they both center around game elements. Hence, to compile a list of possible game elements for our gamified intervention, we also drew upon serious games literature to widen our scope."

[a]Not available.

First, we think that studies benefit if authors identify related concepts that are in close proximity to the core concepts (3.1a). Considering any pair of concepts, the boundaries between these closely related concepts will likely be the vaguest, which also makes them the most important boundaries to define to avoid conceptual ambiguities. Authors may assess the "proximity" of 2 concepts based on various contextual factors of an intervention, such as targeted health behavior. For example, in a physical activity context, exergames are a prominent related concept to consider. Similarly, terms with high linguistic proximity are likely to be candidates for related concepts. As Tan et al [72] remarked in their study, the terms *serious games* and *serious gaming*, although ostensibly the same, can be construed as 2 different concepts that describe different classes of game-based interventions (ie, *serious games* describe games specifically designed for the serious purpose of health education, whereas *serious gaming* is the use of any game for said purpose). For each related concept, the authors should also consider the benefits of introducing it in their manuscript. Increasing the number of concepts introduced by nature also increases the possibility of conceptual ambiguity. In our review, we faced anecdotal mentions of gamification or serious games, which were often difficult to interpret regarding the authors' understanding of the individual concepts and the relationship between the two. Thus, we argue that concepts should only be introduced when they are substantive for understanding a study (3.1b). This also leads to less noise for meta-studies as well as less confusion for readers, as they have to keep track of only substantive concepts. Our guidelines also account for mindfulness regarding conceptual ambiguities in the vocabulary surrounding game-based interventions (3.1c). In game-based interventions, many terms can be quite nuanced or not necessarily correspond to its intuitive meaning. For instance, in games research, the term *game aesthetics* not only describes the artistic value of a game's visual interface but may also refer to the desirable emotional responses evoked in the player [258]. Accordingly, we suggest being precise while using vocabulary in the domain of any concept and not lightheartedly use or arbitrarily adapt possibly nuanced terms. As there are at least theoretical differences between most concepts, we also highly discourage the interchangeable use of concepts in a manuscript (3.1d). If several concepts are summarized under an umbrella term, this should be done in a transparent way to avoid confusing the reader.

Furthermore, we think that authors should be mindful of the applicability of insights from extant literature to their own research in 2 ways. First, if a researcher finds a study that is labeled as focusing on the same concept that they are trying to investigate, it does not necessarily follow that the researcher and the authors of said study have the same understanding of the concept in question. As our literature review illustrates for gamification and serious games, there are various understandings that are not always reconcilable. Hence, it is advisable to be mindful about conceptual ambiguities when reviewing the extant literature on game-based interventions (3.2a) and ideally challenging the label of each study against one's own understanding. Second, when transferring knowledge from one concept to another, the authors should carefully consider the

possible theoretical and empirical differences in the respective concepts. This aspect is supported by extant research that found several empirical differences with regard to the psychological and behavioral outcomes of gamification and serious games [26]. Such differences can easily impede the transferability of knowledge from one concept to another. Thus, the authors should not presume the easy transferability of knowledge from one concept to another (3.2b). Furthermore, after assessing transferability, authors should precisely specify what knowledge they draw from extant literature and why they think that said knowledge is transferable to their study (3.2c). This empowers the reader to put statements into the context of extant literature and allows a better understanding of how knowledge may be transferable, despite possible conceptual differences.

Topic 4: Individual Concept Definitions

The items in the fourth topic (Table 5) focus on aligning the understandings of authors and the reader regarding the introduced concepts.

Table 5. Game-Based Intervention Reporting Guidelines—items in topic 4: individual concept definitions.

Number and topic	Description	Exemplary or explanatory statements
4. Individual concept definitions		
4.1. Definition inspiration		
	(a) Familiarize yourself with definitions for a concept provided by extant literature.	Reviews can often provide a good overview of different views on a concept.
	(b) Decide whether a concept definition from extant literature is applicable for your research or if you need a self-developed definition.	Decision criteria: Deficits in extant literature? Incompatibility of own views with literature?
4.2. Definition of concepts		
	(a) Explicitly define each introduced concept independently in a principal clause. Ideally, justify your choice for a specific definition.	"We define gamification as the use of game design elements in non-game contexts, as this is the most widely applied definition of gamification across disciplines. We define serious games as games whose primary purpose is not entertainment. A game is [...]."
	(b) Explicitly distinguish each introduced related concept pairwise to at least to your core concepts; better even to all related concepts.	"Gamification differs from serious games in that [...]."
4.3. Definitions from extant literature: if definitions are taken from extant literature...		
	(a) Make efforts to identify the original source of a definition.	—[a]
	(b) Include an explicit reference to the source of a definition directly following the definition.	"We define gamification as the use of game-design elements in non-game contexts" [Deterding et al, 2011]; [245]
4.4. Self-developed definitions: if any definition for a concept is self-developed...		
	(a) Make sure to adhere to good definition design.	Be specific, avoid long sentences, do not repeat the term to be defined in the definition.
	(b) Clarify from which views your self-developed definition emerged.	"We include only specific game elements of point, badges and leaderboards in our definition of gamification, because [...]."
4.5. Multiple definitions for a single concept: if multiple definitions for a single concept are provided...		
	(a) State clearly, which definition(s) is (are) applied in the study, and why.	"Gamification can be either defined as [...][exemplary reference 1] or as [...][exemplary reference 2]."; "For the purposes of this study, we follow the view of [exemplary reference 2], because [...]."
	(b) Apply the chosen definition(s) consistently.	—

[a]Not available.

They cover individual concept definitions as well as the relationships between concepts. Having chosen that they want to introduce a certain concept, authors should first familiarize themselves with how extant literature defines it (4.1a). They should then decide whether they want to adopt a definition from the extant literature or need a self-developed definition (4.1b). The latter may especially be the case if authors find that there are no definitions for a concept in extant literature, deficits in existing definitions, or if they feel that their view of a concept is not sufficiently covered by any existing definition. Challenging existing definitions can bring vast benefits to the theoretical development of a research stream. In most cases, however, authors are better off when they build on existing definitions, as this may allow better placement within extant research and facilitate the building of cumulative knowledge. In particular, we urge not to *reinvent the wheel* when providing any self-developed definition. For most game-based intervention concepts, multiple prominent definitions exist that together cover a large range of views (eg, the studies by Deterding et al [245] or Huotari and Hamari [247] for gamification and the

studies by Michael and Chen [14] or Alvarez and Djaouti [255] for serious games). Hence, authors should always consider first if they want to align their views with those of the extant literature.

Irrespective of the source of a definition, it is vital that authors explicitly define every concept introduced in their study (4.2a) to ensure that readers and authors share a common understanding. A lack of an explicit definition for a concept requires the reader to subjectively appraise the authors' understanding of it based on context. Throughout our literature review, we found that about half of the mentions of gamification or serious games are not accompanied by an explicit definition and thus require such an appraisal, which is likely error-prone. Furthermore, concepts should not be initially defined based on their boundaries to other concepts but rather in an independent manner. For example, statements such as "exergames are games that differ from serious games in [...]" do not fully reveal an author's understanding of exergames to readers, but only that they somehow differ from serious games. Defining concepts should be done in a principal clause to ensure that a definition is actually recognized as such. Ideally, authors should also justify their choice for a particular concept definition, as this can provide valuable context to the reader (eg, the knowledge that a particular definition is the most used in a particular discipline). Once all concepts (ie, core concepts and related concepts that are substantive in the sense of 3.1b) are independently defined, the boundaries between concepts can be established (4.2b). Explicitly drawing the boundaries between concepts is the most reliable way to mitigate conceptual ambiguities. Ideally, the boundaries between all possible pairs of introduced concepts are drawn. In cases where this is infeasible (eg, due to a large number of introduced concepts), the boundaries adjacent to the core concepts of a study take priority. The study by Park and Kim [18] is a good example to note the benefits of properly introducing and delineating related concepts. They focus on developing a gamified learning environment, making gamification their core concept. Owing to their proximity to the education context, they identify the related concepts of game-based learning and educational gaming, define them individually, and subsequently sharpen their contribution by stating that gamification is their preferred intervention design for being a more accessible technique for teachers than the other concepts.

If authors decide to adopt definitions from the extant literature, it is important that they establish a clear and unambiguous connection to their respective sources. To do so, researchers should first aim to identify the original source of a definition (4.3a). Readers with domain knowledge are usually able to recognize prominent sources of definitions (especially in the case of seminal papers) and can then immediately attribute a study to a certain view of a concept. Not citing the original source of a definition (ie, indirect citation) obfuscates the original source of a definition and prevents this process. To ensure that a definition can be attributed to its source, the source should be explicitly referenced immediately after the principal clause that contains the definition (4.3b).

Formulating a *good* self-developed definition is a nontrivial task that requires great care. Authors that decide to do so find support in guidelines for a good definition design (4.4a). Although there is no universal way to formulate definitions, authors may consider general guidelines on definition design [263] or take guidance from definition development principles in other areas [264]. Within the context of our study, we particularly emphasize avoiding the use of any part of the term that is defined within the definition of the term itself. For example, when defining serious games as *games with a serious purpose*, the question immediately follows, how *games* and *serious purposes* are defined, essentially leaving the reader none the wiser concerning the meaning of serious games. Whenever authors decide on a self-developed definition of a concept (considering 4.1b), they should also allow the reader to comprehend this decision by clarifying from which view their self-developed definition emerged (4.4b). By fostering an understanding of the decision for a self-developed definition, the reader can also better grasp how a study contributes to an advanced, possibly novel understanding of a concept.

Finally, in some cases, authors may find it beneficial to name several definitions of a single concept, for example, to contrast different views [81,265] or to outline the theoretical development of a concept [112]. To avoid confusing the reader, they should clarify which definition they apply in their study (4.5a) and subsequently apply this definition consistently (4.5b). By further rationalizing the reasons for the choice of a specific definition or concept view, authors can increase the comprehension of the reader of why a certain definition was chosen.

Discussion

Current State of Conceptual Ambiguity in Serious Games and Gamification Literature

Overall, the findings of our study help to better understand the phenomena of gamification and serious games and how they are conceptualized in research published in JMIR journals. Our results paint a heterogeneous landscape of different conceptualizations and different ways of distinguishing the 2 concepts. In addition, there are also some major differences in the evolution of definitions over time. We identified some common pitfalls that may arise when researchers deal with gamification or serious games and provide guidance on how to avoid them. In the following sections, we discuss the principal findings of this study.

First, the results from our literature review provide insights into the vast differences in how researchers conceptualize gamification or serious games. Our results show that less than half (121/265, 45.7%) of concept mentions are accompanied by an explicit definition which leaves room for improvements in scientific rigor. Regarding the sources used for definitions, our results show clear differences between gamification and serious games. Although the majority of definitions used for gamification (43/61, 70%) can be traced back to the seminal paper by Deterding et al [245], no such single central article exists for the concept of serious games. Regarding the content of definitions, in the case of gamification, most studies follow the understanding of Deterding et al [245]. Deviations from this understanding are particularly evident in a change of the

nongame context, meaning that the context is either limited to a specific area of application (eg, health care and education) or omitted completely. From our perspective, both approaches can be problematic. First, we think that gamification and its potential to positively shape the cognition, affection, and behavior of people is not limited to a specific context. In fact, extant research has repeatedly stated that investigating the context specifics of gamification outcomes is one of the most interesting and challenging questions that researchers currently face [33,266]. Second, we think that omitting the *nongame context* from the definition of gamification can lead to great confusion around the concept, as this would technically allow including classic video games with pure entertainment purposes, which was actually intended to be avoided when conceptualizing gamification [245]. For serious games, the definition content was considerably more heterogeneous than that of gamification. However, one aspect stood out. Although some authors emphasize that serious games refer to video games in their entirety [85], others state that they only draw specific aspects from video games, such as design principles [88]. In the latter case, the definitions of serious games seemed considerably closer to the common definitions of gamification.

Second, by analyzing 59 studies that mentioned both gamification and serious games, our results reveal 4 approaches to the relationship between the 2 terms. From our point of view, 2 of these approaches, namely interchangeable use (5/59, 8%) and implicit differentiations (15/59, 25%), are not ideal for communicating one's own ideas of the concepts to the academic audience. As outlined in our reporting guidelines, we think that explicit definitions and distinctions play a key role in mitigating the conceptual ambiguity surrounding different types of game-based interventions. This is also echoed by extant literature on conceptual ambiguity problems, which states that such problems can arise when researchers implicitly assume that because they can make a logical distinction between concepts, this distinction will also exist in the minds of others [267]. The largest group of studies (22/59, 37%) labeled serious games as a type of gamified intervention or as the product resulting from the gamification process. This seems closely related to the classic process view of gamification [265], where gamification is construed as a process of transforming a purely utilitarian system into a system that combines utilitarian and hedonic functions by implementing game design elements [268]. From our point of view, this approach is incompatible with the understanding of the last group of studies (9/59, 15%) that, similar to Deterding et al [245], consolidate serious games as *full-fledged games*, whereas gamification involves systems that only partly consist of game design elements and in addition have nongame parts. A similar view is shared by Marczweski [259], who argues that contrary to gamification, serious games provide gameplay, which he describes as "[having] all the elements you would expect to see in a real game." Although conceptually sound, this distinction can also be problematic, as deciding whether users actually *play* a game or *use* a system is an often complex question that involves empirical, subjective, and social factors [56,245]. In fact, for medical interventions in practice, it seems infeasible to empirically investigate each individual user on whether they are using or playing their system. On a conceptual level, this issue may be approached by substituting actual user behavior with design intention.

Finally, when analyzing the thematic focus of our review literature, our results show an ever-increasing amount of both gamification and serious games studies in the *JMIR* research community since 2015, with seemingly no concept prevailing over the other. We found this particularly interesting because it contrasts previous observations that gamification is increasingly superseding serious games as a more popular concept [18]. In fact, we also think that the circumstance that no concept prevails over the other even strengthens the relevance of clearly distinguishing between different types of game-based health interventions and attributing research results accurately in the future.

Implications

Our study has several implications for future research. First, our study revealed a strong heterogeneity surrounding the conceptualizations of serious games and gamification. In particular, we demonstrate that just because 2 game-based interventions are labeled similarly, it does not necessarily follow that the designers share the same understanding of the underlying concept. For researchers, this implies that they need to exercise particular caution and scientific rigor when studying game-based interventions and reporting study results. In fact, the easy transferability of knowledge based on labels provided in the literature is often not given.

Second, our results strongly indicate that there are conceptual ambiguity problems surrounding serious games and gamification in the health care literature. To counteract this, we rigorously developed GAMING and evaluated them with the help of 9 outstanding experts from the area of game-based intervention research. To the best of our knowledge, this is the first study to formulate such reporting guidelines explicitly for game-based interventions. The individual guideline items are designed to mitigate common pitfalls that lead to conceptual ambiguities in the literature reviewed. Researchers can use GAMING as an inspiration for reporting the results of game-based health interventions while accounting for potential conceptual ambiguity. However, GAMING should not be interpreted as a prescription for the reporting of game-based health interventions in a strict or standardized format. The guideline items should be addressed in sufficient detail and with clarity somewhere in the manuscript, but the order and format for presenting the required information depends on author preferences, journal style, and the traditions of the research field.

For practitioners, we believe that our study can help raise awareness about the conceptual ambiguity surrounding game-based health interventions. Practitioners, similar to researchers, should exercise caution when interpreting research findings and transferring them into practice. Our review and guidelines may serve as a starting point to assess whether any study they want to draw on may be affected by conceptual ambiguity and subsequently support them in evaluating the transferability of knowledge for their own purposes.

Limitations and Future Research

The results of our study are limited in several ways, which also opens up avenues for future research. First, we limited our literature review to studies published in the *Journal of Medical Internet Research* and its sister journals. Although we are convinced that the *Journal of Medical Internet Research* is a suitable representative of the overall landscape of game-based health interventions, future research may also benefit from investigating the conceptual ambiguities surrounding gamification and serious games outside of the *JMIR* research community. We would be particularly interested in whether our findings regarding the different evolutions of gamification and serious games definitions (Figures 2 and 3) hold in a different literature corpus. Researchers may also find it beneficial to extend our approach to a full citation network analysis to investigate the flow of knowledge in serious games and gamification literature [269]. This may allow further insights into whether given conceptualizations are refinements or extensions of existing ones, which is a prime indicator for investigating conceptual ambiguities [28]. Second, we only included two concepts in our literature review: gamification and serious games. As we alluded to in our introduction, a plethora of concepts for health interventions draws on games as their inspirational source. However, gamification and serious games are currently prevailing concepts in game-based interventions [13,18]. Thus, we argue that investigating the conceptual conflict between this particular concept pair allows for the extraction of the highest proportion of knowledge. Nevertheless, future research may still benefit from including additional or different concepts within the scope of the investigation. As for our reporting guidelines, we believe that we converted the insights from our literature review to a sufficient level of abstraction so that they are transferable to other types of game-based interventions.

Conclusions

Games are an increasingly popular source of inspiration for the design of health care interventions. With the ever-increasing number of game-based interventions, concerns regarding conceptual ambiguity have arisen. In this study, we focus on the conceptual boundaries between two of the most prominent game-based intervention concepts: gamification and serious games. Our analysis of the literature in the *JMIR* research community unveils various understandings of the concepts themselves, as well as the boundaries between them. Thus, our results support the initial hypothesis of conceptual ambiguity between gamification and serious games in the health care literature. On the basis of these results, we proposed GAMING, consisting of 25 items designed to guide researchers in reporting their own game-based intervention studies in a way that mitigates conceptual ambiguity. We believe that our results can serve as a valuable supplement to existing research that has discussed conceptual differences between different types of game-based interventions and can help advance cumulative knowledge development in game-based health intervention research, without being impaired by conceptual ambiguity.

Acknowledgments

The authors would like to thank Michelle Berger, Lobna Hassan, Andreas Janson, Jeff Mullins, Stacie Petter, Lisa-Maria Putz, Sofia Schöbel, Stefan Stepanovic, and Tobias Wolf for providing valuable feedback on the reporting guidelines.

Authors' Contributions

The authors’ contributions according to the Contributor Roles Taxonomy are as follows: Conceptualization was done by SW, MSK, SR, and ST. Data curation was done by SW and MSK. Formal analysis was done by SW, MSK, and SR. Funding acquisition was done by ST and AS. Investigation was done by SW, MSK, SR, and ST. Methodology was developed by SW, MSK, SR, and ST. Project administration was done by SW. Resources were provided by AS. Supervision was done by ST and AS. Validation was done by SW and MSK. Visualization was done by SR. Writing the original draft was done by all authors. Writing, review, and editing was done by all authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of the 206 papers included in the literature review.
[XLSX File (Microsoft Excel File), 96 KB - jmir_v23i9e30390_app1.xlsx]

Multimedia Appendix 2
Consolidated, one-page version of Game-Based Intervention Reporting Guidelines, excluding rationales.
[DOCX File , 28 KB - jmir_v23i9e30390_app2.docx]

Multimedia Appendix 3
Consolidated, extended version of Game-Based Intervention Reporting Guidelines, including detailed rationales and exemplary statements.
[DOCX File , 32 KB - jmir_v23i9e30390_app3.docx]

Multimedia Appendix 4
Overview of existing distinctions between serious games and gamification in extant literature.
[DOCX File , 24 KB - jmir_v23i9e30390_app4.docx]

References

1. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]
2. Pramana G, Parmanto B, Lomas J, Lindhiem O, Kendall PC, Silk J. Using mobile health gamification to facilitate cognitive behavioral therapy skills practice in child anxiety treatment: open clinical trial. JMIR Serious Games 2018 May 10;6(2):e9 [FREE Full text] [doi: 10.2196/games.8902] [Medline: 29748165]
3. McGonigal J. Reality Is Broken: Why Games Make Us Better and How They Can Change the World. New York: Penguin Group; 2011:1-396.
4. Segura EM, Waern A, Segura LM, Recio DL. Playification: The PhySeEar case. In: Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play. 2016 Presented at: CHI PLAY '16: The Annual Symposium on Computer-Human Interaction in Play; October 16 - 19, 2016; Austin Texas USA p. 376-388. [doi: 10.1145/2967934.2968099]
5. Al-Azawi R, Al-Faliti F, Al-Blushi M. Educational gamification vs. game based learning: comparative study. Int J Innov Manage Technol 2016;7(4):131-136. [doi: 10.18178/ijimt.2016.7.4.659]
6. Prensky M. Digital game-based learning. Comput Entertain 2003 Oct;1(1):21. [doi: 10.1145/950566.950596]
7. Williams WM, Ayres CG. Can active video games improve physical activity in adolescents? A review of RCT. Int J Environ Res Public Health 2020 Jan 20;17(2):669 [FREE Full text] [doi: 10.3390/ijerph17020669] [Medline: 31968642]
8. Leibetseder A, Lux M. Gamifying fitness or fitnessifying games: a comparative study. In: Proceedings of the GamifIR Workshop. 2016 Presented at: Workshop on Gamification for Information Retrieval (GamifIR); December 2016; Pisa Italy p. 37-44 URL: http://www-itec.uni-klu.ac.at/bib/files/__GAMIFIR__Research_Fitnessification.pdf
9. Liu L, Stacey P, Tarafdar M, Kourentzes N. Simplifying fitness games for users with learning disabilities. In: Proceedings of the 50th Hawaii International Conference on System Sciences, HICSS 2017. 2017 Presented at: 50th Hawaii International Conference on System Sciences, HICSS 2017; January 4-7, 2017; Hilton Waikoloa Village, Hawaii, USA p. 1-10. [doi: 10.24251/HICSS.2017.463]
10. McBain T, Weston M, Crawshaw P, Haighton C, Spears I. Development of an exergame to deliver a sustained dose of high-intensity training: formative pilot randomized trial. JMIR Serious Games 2018 Mar 27;6(1):e4 [FREE Full text] [doi: 10.2196/games.7758] [Medline: 29588271]
11. Murphy B, Aleni A, Belaoucha B, Dyer J, Nolan H. Quantifying cognitive aging and performance with at-home gamified mobile EEG. In: Proceedings of the International Workshop on Pattern Recognition in Neuroimaging (PRNI). 2018 Presented at: International Workshop on Pattern Recognition in Neuroimaging (PRNI); June 12-14, 2018; Singapore p. 1-4. [doi: 10.1109/prni.2018.8423954]
12. Chen AM, Kiersma ME, Yehle KS, Plake KS. Impact of an aging simulation game on pharmacy students' empathy for older adults. Am J Pharm Educ 2015 Jun 25;79(5):65 [FREE Full text] [doi: 10.5688/ajpe79565] [Medline: 26396274]
13. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Dec;71:31-48. [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]
14. Michael D, Chen S. Serious Games: Games That Educate, Train and Inform. Boston, Massachusetts, United States: Thomson Course Technology; 2006:1-352.
15. Bui A, Veit D, Webster J. Gamification – a novel phenomenon or a new wrapping for existing concepts? In: Proceedings of the 36th International Conference on Information Systems. 2015 Presented at: 36th International Conference on Information Systems; December 13-16, 2015; Fort Worth, Texas, USA p. 1-21 URL: https://core.ac.uk/download/pdf/301367495.pdf
16. Orji R, Nacke L, Di Marco C. Towards personality-driven persuasive health games and gamified systems. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17: CHI Conference on Human Factors in Computing Systems; May 6 - 11, 2017; Denver Colorado USA p. 1015-1027. [doi: 10.1145/3025453.3025577]
17. Croon RD, Wildemeersch D, Wille J, Verbert K, Abeele V. Gamification and serious games in a healthcare informatics context. In: Proceedings of the IEEE International Conference on Healthcare Informatics (ICHI). 2018 Presented at: IEEE International Conference on Healthcare Informatics (ICHI); June 4-7, 2018; New York, NY, USA p. 53-63. [doi: 10.1109/ichi.2018.00014]
18. Park S, Kim S. A badge design framework for a gamified learning environment: cases analysis and literature review for badge design. JMIR Serious Games 2019 May 29;7(2):e14342 [FREE Full text] [doi: 10.2196/14342] [Medline: 31144664]
19. Stepanovic S, Mettler T. Gamification applied for health promotion: does it really foster long-term engagement? A scoping review. In: Proceedings of the 26th European Conference on Information Systems. 2018 Presented at: 26th European Conference on Information Systems; June 23-28, 2018; Portsmouth, UK p. 1-16 URL: https://serval.unil.ch/resource/serval:BIB_50A0EC5E3795.P002/REF
20. Wattanasoontorn V, Boada I, García R, Sbert M. Serious games for health. Entertain Comput 2013 Dec;4(4):231-247. [doi: 10.1016/j.entcom.2013.09.002]

21. Ricciardi F, De Paolis LT. A comprehensive review of serious games in health professions. Int J Comput Games Technol 2014:1-11. [doi: 10.1155/2014/787968]
22. Thiebes S, Lins S, Basten D. Gamifying information systems-a synthesis of gamification mechanics and dynamics. In: Proceedings of the 22nd European Conference on Information Systems. 2014 Presented at: 22nd European Conference on Information Systems; June 9-11, 2014; Tel Aviv, Israel URL: https://www.researchgate.net/publication/262805143_Gamifying_Information_Systems_-_A_Synthesis_of_Gamification_Mechanics_and_Dynamics
23. Werbach K. (Re)defining gamification: a process approach. In: Proceedings of the International Conference on Persuasive Technology. 2014 Presented at: International Conference on Persuasive Technology; May 21-23, 2014; Padua, Italy p. 266-272. [doi: 10.1007/978-3-319-07127-5_23]
24. Susi T, Johannesson M, Backlund P. Serious games – an overview. School of Humanities and Informatics, University of Skövde, Sweden. 2007. URL: https://www.diva-portal.org/smash/get/diva2:2416/FULLTEXT01.pdf [accessed 2021-07-29]
25. Liu D, Santhanam R, Webster J. Toward meaningful engagement: a framework for design and research of gamified information systems. MIS Q 2017 Apr 4;41(4):1011-1034. [doi: 10.25300/misq/2017/41.4.01]
26. Pfau J, Smeddinck J, Volkmar G, Wenig N, Malaka R. Do you think this is a game? Contrasting a serious game with a gamified application for health. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18: CHI Conference on Human Factors in Computing Systems; April 21 - 26, 2018; Montreal QC Canada URL: https://docplayer.net/212316150-Do-you-think-this-is-a-game-contrasting-a-serious-game-with-a-gamified-application-for-health.html [doi: 10.1145/3170427.3188651]
27. Shaffer JA, DeGeest D, Li A. Tackling the problem of construct proliferation. Organ Res Methods 2015 Aug 18;19(1):80-110. [doi: 10.1177/1094428115598239]
28. Le H, Schmidt FL, Harter JK, Lauver KJ. The problem of empirical redundancy of constructs in organizational research: an empirical investigation. Organ Behav Hum Decis Process 2010 Jul;112(2):112-125. [doi: 10.1016/j.obhdp.2010.02.003]
29. Yohannis A, Prabowo Y, Waworuntu A. Defining gamification: From lexical meaning and process viewpoint towards a gameful reality. In: Proceedings of the International Conference on Information Technology Systems and Innovation (ICITSI). 2014 Presented at: International Conference on Information Technology Systems and Innovation (ICITSI); November 24-27, 2014; Bandung, Indonesia p. 284-289. [doi: 10.1109/icitsi.2014.7048279]
30. Kim B. Understanding Gamification. Chicago: American Library Association; 2016:1-37.
31. Landers RN. Developing a theory of gamified learning. Simul Gaming 2015 Jan 15;45(6):752-768. [doi: 10.1177/1046878114563660]
32. Rousseau DM. A sticky, leveraging, and scalable strategy for high-quality connections between organizational practice and science. Acad Manage J 2007 Oct;50(5):1037-1042. [doi: 10.5465/amj.2007.27155539]
33. Koivisto J, Hamari J. The rise of motivational information systems: a review of gamification research. Int J Info Manage 2019 Apr;45:191-210. [doi: 10.1016/j.ijinfomgt.2018.10.013]
34. Kankanhalli A, Shin J, Oh H. Mobile-based interventions for dietary behavior change and health outcomes: scoping review. JMIR Mhealth Uhealth 2019 Jan 21;7(1):e11312 [FREE Full text] [doi: 10.2196/11312] [Medline: 30664461]
35. Schmidt-Kraepelin M, Warsinsky S, Thiebes S, Sunyaev A. The role of gamification in health behavior change: a review of theory-driven studies. In: Proceedings of the 53rd Hawaii International Conference on System Sciences. 2020 Presented at: 53rd Hawaii International Conference on System Sciences; Jan 07 - 10, 2020; Maui, Hawaii p. 1256-1265. [doi: 10.24251/hicss.2020.155]
36. Levy Y, Ellis TJ. A systems approach to conduct an effective literature review in support of information systems research. InformingSciJ 2006;9:181-212. [doi: 10.28945/479]
37. Journal citation report 2019-2020. Clarivate Analytics - Journal Impact Factor (ISI) 2019 - 2020. URL: https://www.researchgate.net/publication/334048283_Journal_citation_report_2019-2020_-_Clarivate_Analytics_-_Journal_Impact_Factor_ISI_2019_-_2020 [accessed 2021-07-29]
38. Search articles. JMIR Publications. URL: https://jmir.org/search/ [accessed 2021-08-10]
39. Bossenbroek R, Wols A, Weerdmeester J, Lichtwarck-Aschoff A, Granic I, van Rooij MM. Efficacy of a virtual reality biofeedback game (DEEP) to reduce anxiety and disruptive classroom behavior: single-case study. JMIR Ment Health 2020 Mar 24;7(3):e16066 [FREE Full text] [doi: 10.2196/16066] [Medline: 32207697]
40. Fishbein JN, Nisotel LE, MacDonald JJ, Amoyal PN, Jacobs JM, Flanagan C, et al. Mobile application to promote adherence to oral chemotherapy and symptom management: a protocol for design and development. JMIR Res Protoc 2017 Apr 20;6(4):e62 [FREE Full text] [doi: 10.2196/resprot.6198] [Medline: 28428158]
41. Health care. Merriam-Webster. URL: https://www.merriam-webster.com/dictionary/health%20care [accessed 2021-06-10]
42. Webster J, Watson RT. Analyzing the past to prepare for the future: writing a literature review. MIS Q 2002;26(2):xiii-xxiii [FREE Full text]
43. Edney S, Ryan JC, Olds T, Monroe C, Fraysse F, Vandelanotte C, et al. User engagement and attrition in an app-based physical activity intervention: secondary analysis of a randomized controlled trial. J Med Internet Res 2019 Nov 27;21(11):e14645 [FREE Full text] [doi: 10.2196/14645] [Medline: 31774402]

44. Fornasini S, Dianti M, Bacchiega A, Forti S, Conforti D. Using geocaching to promote active aging: qualitative study. J Med Internet Res 2020 Jun 11;22(6):e15339 [FREE Full text] [doi: 10.2196/15339] [Medline: 32525486]
45. Ferreira LD, Ferreira H, Cavaco S, Cameirão M, I Badia SB. User experience of interactive technologies for people with dementia: comparative observational study. JMIR Serious Games 2020 Aug 05;8(3):e17565 [FREE Full text] [doi: 10.2196/17565] [Medline: 32755894]
46. Tziraki C, Berenbaum R, Gross D, Abikhzer J, Ben-David BM. Designing serious computer games for people with moderate and advanced dementia: interdisciplinary theory-driven pilot study. JMIR Serious Games 2017 Jul 31;5(3):e16 [FREE Full text] [doi: 10.2196/games.6514] [Medline: 28760730]
47. Martinez W, Threatt AL, Rosenbloom ST, Wallston KA, Hickson GB, Elasy TA. A patient-facing diabetes dashboard embedded in a patient web portal: design sprint and usability testing. JMIR Hum Factors 2018 Sep 24;5(3):e26 [FREE Full text] [doi: 10.2196/humanfactors.9569] [Medline: 30249579]
48. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
49. Yokomitsu K, Irie T, Sekiguchi M, Shimizu A, Matsuoka H, Merry SN, et al. Gamified mobile computerized cognitive behavioral therapy for Japanese University students with depressive symptoms: protocol for a randomized controlled trial. JMIR Res Protoc 2020 Apr 07;9(4):e15164 [FREE Full text] [doi: 10.2196/15164] [Medline: 32254045]
50. Vervaeke J, Van Looy J, Hoorelbeke K, Baeken C, Koster EH. Gamified cognitive control training for remitted depressed individuals: user requirements analysis. JMIR Serious Games 2018 Apr 05;6(2):e6 [FREE Full text] [doi: 10.2196/games.8609] [Medline: 29622525]
51. Tuti T, Winters N, Edgcombe H, Muinga N, Wanyama C, English M, et al. Evaluation of adaptive feedback in a smartphone-based game on health care providers' learning gain: randomized controlled trial. J Med Internet Res 2020 Jul 06;22(7):e17100 [FREE Full text] [doi: 10.2196/17100] [Medline: 32628115]
52. IJgosse W, van Goor H, Rosman C, Luursema J. Construct validity of a serious game for laparoscopic skills training: validation study. JMIR Serious Games 2020 May 07;8(2):e17222 [FREE Full text] [doi: 10.2196/17222] [Medline: 32379051]
53. Haubruck P, Nickel F, Ober J, Walker T, Bergdolt C, Friedrich M, et al. Evaluation of app-based serious gaming as a training method in teaching chest tube insertion to medical students: randomized controlled trial. J Med Internet Res 2018 Dec 21;20(5):e195 [FREE Full text] [doi: 10.2196/jmir.9956] [Medline: 29784634]
54. Zakaria NS, Saripan MI, Subarimaniam N, Ismail A. Assessing Ethoshunt as a gamification-based mobile app in ethics education: pilot mixed-methods study. JMIR Serious Games 2020 Aug 10;8(3):e18247 [FREE Full text] [doi: 10.2196/18247] [Medline: 32663153]
55. Lister C, West JH, Cannon B, Sax T, Brodegard D. Just a fad? Gamification in health and fitness apps. JMIR Serious Games 2014;2(2):e9 [FREE Full text] [doi: 10.2196/games.3413] [Medline: 25654660]
56. Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR. Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. JMIR Serious Games 2016;4(2):e11 [FREE Full text] [doi: 10.2196/games.5888] [Medline: 27421244]
57. Hoffmann A, Christmann CA, Bleser G. Gamification in stress management apps: a critical app review. JMIR Serious Games 2017 Jun 07;5(2):e13 [FREE Full text] [doi: 10.2196/games.7216] [Medline: 28592397]
58. Booth R, Sinclair B, McMurray J, Strudwick G, Watson G, Ladak H, et al. Evaluating a serious gaming electronic medication administration record system among nursing students: protocol for a pragmatic randomized controlled trial. JMIR Res Protoc 2018 May 28;7(5):e138 [FREE Full text] [doi: 10.2196/resprot.9601] [Medline: 29807885]
59. Friehs MA, Dechant M, Vedress S, Frings C, Mandryk RL. Effective gamification of the stop-signal task: two controlled laboratory experiments. JMIR Serious Games 2020 Sep 08;8(3):e17810 [FREE Full text] [doi: 10.2196/17810] [Medline: 32897233]
60. Vermeir JF, White MJ, Johnson D, Crombez G, Van Ryckeghem DM. The effects of gamification on computerized cognitive training: systematic review and meta-analysis. JMIR Serious Games 2020 Aug 10;8(3):e18644 [FREE Full text] [doi: 10.2196/18644] [Medline: 32773374]
61. Hightow-Weidman L, Muessig K, Knudtson K, Srivatsa M, Lawrence E, LeGrand S, et al. A gamified smartphone app to support engagement in care and medication adherence for HIV-positive young men who have sex with men (AllyQuest): development and pilot study. JMIR Public Health Surveill 2018 Apr 30;4(2):e34 [FREE Full text] [doi: 10.2196/publichealth.8923] [Medline: 29712626]
62. Jessen S, Mirkovic J, Ruland CM. Creating gameful design in mHealth: a participatory co-design approach. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e11579 [FREE Full text] [doi: 10.2196/11579] [Medline: 30552080]
63. Leinonen A, Pyky R, Ahola R, Kangas M, Siirtola P, Luoto T, et al. Feasibility of gamified mobile service aimed at physical activation in young men: population-based randomized controlled study (MOPO). JMIR Mhealth Uhealth 2017 Oct 10;5(10):e146 [FREE Full text] [doi: 10.2196/mhealth.6675] [Medline: 29017991]
64. AlMarshedi A, Wills G, Ranchhod A. Guidelines for the gamification of self-management of chronic illnesses: multimethod study. JMIR Serious Games 2017 May 12;5(2):e12 [FREE Full text] [doi: 10.2196/games.7472] [Medline: 28500018]

65. Fleming T, Merry S, Stasiak K, Hopkins S, Patolo T, Ruru S, et al. The importance of user segmentation for designing digital therapy for adolescent mental health: findings from scoping processes. JMIR Ment Health 2019 May 08;6(5):e12656 [FREE Full text] [doi: 10.2196/12656] [Medline: 31066705]
66. Rohde A, Duensing A, Dawczynski C, Godemann J, Lorkowski S, Brombach C. An app to improve eating habits of adolescents and young adults (challenge to go): systematic development of a theory-based and target group-adapted mobile app intervention. JMIR Mhealth Uhealth 2019 Aug 12;7(8):e11575 [FREE Full text] [doi: 10.2196/11575] [Medline: 30903746]
67. White BK, Martin A, White JA, Burns SK, Maycock BR, Giglia RC, et al. Theory-based design and development of a socially connected, gamified mobile app for men about breastfeeding (Milk Man). JMIR Mhealth Uhealth 2016 Jun 27;4(2):e81 [FREE Full text] [doi: 10.2196/mhealth.5652] [Medline: 27349756]
68. Zhang M, Ying J, Song G, Fung DS, Smith H. Gamified cognitive bias modification interventions for psychiatric disorders: review. JMIR Ment Health 2018 Oct 25;5(4):e11640 [FREE Full text] [doi: 10.2196/11640] [Medline: 30361194]
69. Fortunato M, Harrison J, Oon AL, Small D, Hilbert V, Rareshide C, et al. Remotely monitored gamification and social incentives to improve glycemic control among adults with uncontrolled type 2 diabetes (iDiabetes): protocol for a randomized controlled trial. JMIR Res Protoc 2019 Nov 20;8(11):e14180 [FREE Full text] [doi: 10.2196/14180] [Medline: 31746765]
70. Pasquier P, Mérat S, Malgras B, Petit L, Queran X, Bay C, et al. A serious game for massive training and assessment of French soldiers involved in forward combat casualty care (3D-SC1): development and deployment. JMIR Serious Games 2016 May 18;4(1):e5 [FREE Full text] [doi: 10.2196/games.5340] [Medline: 27194369]
71. Gentry SV, Gauthier A, Ehrstrom BL, Wortley D, Lilienthal A, Car LT, et al. Serious gaming and gamification education in health professions: systematic review. J Med Internet Res 2019 Mar 28;21(3):e12994 [FREE Full text] [doi: 10.2196/12994] [Medline: 30920375]
72. Tan JW, Zary N. Diagnostic markers of user experience, play, and learning for digital serious games: a conceptual framework study. JMIR Serious Games 2019 Jul 16;7(3):e14620 [FREE Full text] [doi: 10.2196/14620] [Medline: 31313660]
73. Giunti G. 3MD for chronic conditions, a model for motivational mHealth design: embedded case study. JMIR Serious Games 2018 Aug 24;6(3):e11631 [FREE Full text] [doi: 10.2196/11631] [Medline: 30143476]
74. Martin A, Caon M, Adorni F, Andreoni G, Ascolese A, Atkinson S, et al. A mobile phone intervention to improve obesity-related health behaviors of adolescents across Europe: iterative co-design and feasibility study. JMIR Mhealth Uhealth 2020 Mar 02;8(3):e14118 [FREE Full text] [doi: 10.2196/14118] [Medline: 32130179]
75. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. JMIR Mhealth Uhealth 2015 May 13;3(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4283] [Medline: 25972309]
76. Nour M, Chen J, Allman-Farinelli M. Young adults' engagement with a self-monitoring app for vegetable intake and the impact of social media and gamification: feasibility study. JMIR Form Res 2019 May 10;3(2):e13324 [FREE Full text] [doi: 10.2196/13324] [Medline: 31094322]
77. Brown JM, Savaglio R, Watson G, Kaplansky A, LeSage A, Hughes J, et al. Optimizing child nutrition education with the foodbot factory mobile health app: formative evaluation and analysis. JMIR Form Res 2020 Apr 17;4(4):e15534 [FREE Full text] [doi: 10.2196/15534] [Medline: 32301743]
78. Burgess J, Watt K, Kimble RM, Cameron CM. Combining technology and research to prevent scald injuries (the cool runnings intervention): randomized controlled trial. J Med Internet Res 2018 Oct 10;20(10):e10361 [FREE Full text] [doi: 10.2196/10361] [Medline: 30305263]
79. Khalili-Mahani N, De Schutter B. Affective game planning for health applications: quantitative extension of gerontoludic design based on the appraisal theory of stress and coping. JMIR Serious Games 2019 Jun 06;7(2):e13303 [FREE Full text] [doi: 10.2196/13303] [Medline: 31172966]
80. Schmidt-Kraepelin M, Toussaint PA, Thiebes S, Hamari J, Sunyaev A. Archetypes of gamification: analysis of mhealth apps. JMIR Mhealth Uhealth 2020 Oct 19;8(10):e19280 [FREE Full text] [doi: 10.2196/19280] [Medline: 33074155]
81. Cheng VW, Davenport TA, Johnson D, Vella K, Mitchell J, Hickie IB. An app that incorporates gamification, mini-games, and social connection to improve men's mental health and well-being (MindMax): participatory design process. JMIR Ment Health 2018 Nov 19;5(4):e11068 [FREE Full text] [doi: 10.2196/11068] [Medline: 30455165]
82. Park S, Kim S. Patterns among 754 gamification cases: content analysis for gamification development. JMIR Serious Games 2018 Nov 06;6(4):e11336 [FREE Full text] [doi: 10.2196/11336] [Medline: 30401666]
83. Khalili-Mahani N, Assadi A, Li K, Mirgholami M, Rivard M, Benali H, et al. Reflective and reflexive stress responses of older adults to three gaming experiences in relation to their cognitive abilities: mixed methods crossover study. JMIR Ment Health 2020 Mar 26;7(3):e12388 [FREE Full text] [doi: 10.2196/12388] [Medline: 32213474]
84. LeGrand S, Muessig KE, Platt A, Soni K, Egger JR, Nwoko N, et al. Epic allies, a gamified mobile phone app to improve engagement in care, antiretroviral uptake, and adherence among young men who have sex with men and young transgender women who have sex with men: protocol for a randomized controlled trial. JMIR Res Protoc 2018 Apr 05;7(4):e94 [FREE Full text] [doi: 10.2196/resprot.8811] [Medline: 29622527]
85. Montagni I, Mabchour I, Tzourio C. Digital gamification to enhance vaccine knowledge and uptake: scoping review. JMIR Serious Games 2020 May 18;8(2):e16983 [FREE Full text] [doi: 10.2196/16983] [Medline: 32348271]

86. Brown M, O'Neill N, van Woerden H, Eslambolchilar P, Jones M, John A. Gamification and adherence to web-based mental health interventions: a systematic review. JMIR Ment Health 2016;3(3):e39 [FREE Full text] [doi: 10.2196/mental.5710] [Medline: 27558893]
87. Fordham J, Ball C. Framing mental health within digital games: an exploratory case study of hellblade. JMIR Ment Health 2019 Apr 18;6(4):e12432 [FREE Full text] [doi: 10.2196/12432] [Medline: 30998224]
88. Winter M, Pryss R, Probst T, Reichert M. Learning to read by learning to write: evaluation of a serious game to foster business process model comprehension. JMIR Serious Games 2020 Jan 09;8(1):e15374 [FREE Full text] [doi: 10.2196/15374] [Medline: 31917374]
89. Bossen D, Broekema A, Visser B, Brons A, Timmerman A, van Etten-Jamaludin F, et al. Effectiveness of serious games to increase physical activity in children with a chronic disease: systematic review with meta-analysis. J Med Internet Res 2020 Apr 01;22(4):e14549 [FREE Full text] [doi: 10.2196/14549] [Medline: 32234697]
90. Glover K, Bodzin A. Learner analysis to inform the design and development of a serious game for nongaming female emerging health care preprofessionals: qualitative sample study. JMIR Serious Games 2020 Feb 06;8(1):e16003 [FREE Full text] [doi: 10.2196/16003] [Medline: 32027312]
91. Abraham O, Thakur T, Brown R. Developing a theory-driven serious game to promote prescription opioid safety among adolescents: mixed methods study. JMIR Serious Games 2020 Jul 03;8(3):e18207 [FREE Full text] [doi: 10.2196/18207] [Medline: 32618571]
92. Adjedj J, Ducrocq G, Bouleti C, Reinhart L, Fabbro E, Elbez Y, et al. Medical student evaluation with a serious game compared to multiple choice questions assessment. JMIR Serious Games 2017 May 16;5(2):e11 [FREE Full text] [doi: 10.2196/games.7033] [Medline: 28512082]
93. Reichlin L, Mani N, McArthur K, Harris AM, Rajan N, Dacso CC. Assessing the acceptability and usability of an interactive serious game in aiding treatment decisions for patients with localized prostate cancer. J Med Internet Res 2011 Jan 12;13(1):e4 [FREE Full text] [doi: 10.2196/jmir.1519] [Medline: 21239374]
94. Shiyko M, Hallinan S, Seif EM, Subramanian S, Castaneda-Sceppa C. Effects of playing a serious computer game on body mass index and nutrition knowledge in women. JMIR Serious Games 2016 Jun 02;4(1):e8 [FREE Full text] [doi: 10.2196/games.4977] [Medline: 27255497]
95. Caserman P, Hoffmann K, Müller P, Schaub M, Straßburg K, Wiemeyer J, et al. Quality criteria for serious games: serious part, game part, and balance. JMIR Serious Games 2020 Jul 24;8(3):e19037 [FREE Full text] [doi: 10.2196/19037] [Medline: 32706669]
96. Mack I, Reiband N, Etges C, Eichhorn S, Schaeffeler N, Zurstiege G, et al. The Kids Obesity Prevention Program: Cluster randomized controlled trial to evaluate a serious game for the prevention and treatment of childhood obesity. J Med Internet Res 2020 Apr 24;22(4):e15725 [FREE Full text] [doi: 10.2196/15725] [Medline: 32329742]
97. Vilardaga R, Rizo J, Zeng E, Kientz JA, Ries R, Otis C, et al. User-centered design of learn to quit, a smoking cessation smartphone app for people with serious mental illness. JMIR Serious Games 2018 Jan 16;6(1):e2 [FREE Full text] [doi: 10.2196/games.8881] [Medline: 29339346]
98. Lindner P, Rozental A, Jurell A, Reuterskiöld L, Andersson G, Hamilton W, et al. Experiences of gamified and automated virtual reality exposure therapy for spider phobia: qualitative study. JMIR Serious Games 2020 Apr 29;8(2):e17807 [FREE Full text] [doi: 10.2196/17807] [Medline: 32347803]
99. Gamito P, Oliveira J, Lopes P, Brito R, Morais D, Silva D, et al. Executive functioning in alcoholics following an mHealth cognitive stimulation program: randomized controlled trial. J Med Internet Res 2014 Apr 17;16(4):e102. [doi: 10.2196/jmir.2923] [Medline: 24742381]
100. Robert P, Manera V, Derreumaux A, Montesino M, Leone E, Fabre R, et al. Efficacy of a web app for cognitive training (MeMo) regarding cognitive and behavioral performance in people with neurocognitive disorders: randomized controlled trial. J Med Internet Res 2020 Mar 11;22(3):e17167 [FREE Full text] [doi: 10.2196/17167] [Medline: 32159519]
101. Gautier EJ, Bot-Robin V, Libessart A, Doucède G, Cosson M, Rubod C. Design of a serious game for handling obstetrical emergencies. JMIR Serious Games 2016 Dec 21;4(2):e21 [FREE Full text] [doi: 10.2196/games.5526] [Medline: 28003175]
102. Ong JG, Lim-Ashworth NS, Ooi YP, Boon JS, Ang RP, Goh DH, et al. An interactive mobile app game to address aggression (RegnaTales): pilot quantitative study. JMIR Serious Games 2019 May 08;7(2):e13242 [FREE Full text] [doi: 10.2196/13242] [Medline: 31066682]
103. DeSmet A, Thompson D, Baranowski T, Palmeira A, Verloigne M, De Bourdeaudhuij I. Is participatory design associated with the effectiveness of serious digital games for healthy lifestyle promotion? A meta-analysis. J Med Internet Res 2016 Apr 29;18(4):e94 [FREE Full text] [doi: 10.2196/jmir.4444] [Medline: 27129447]
104. Vugts MA, Zedlitz AM, Joosen MC, Vrijhoef HJ. Serious gaming during multidisciplinary rehabilitation for patients with chronic pain or fatigue symptoms: mixed methods design of a realist process evaluation. J Med Internet Res 2020 Mar 09;22(3):e14766 [FREE Full text] [doi: 10.2196/14766] [Medline: 32149720]
105. Carvalho DB, Domingueti DB, Santos SM, Dias DR. A game to deal with alcohol abuse (Jib): development and game experience evaluation. JMIR Serious Games 2019 Oct 15;7(4):e11151 [FREE Full text] [doi: 10.2196/11151] [Medline: 31617849]

106. Jander A, Crutzen R, Mercken L, Candel M, de Vries H. Effects of a web-based computer-tailored game to reduce binge drinking among Dutch adolescents: a cluster randomized controlled trial. J Med Internet Res 2016 Feb 03;18(2):e29 [FREE Full text] [doi: 10.2196/jmir.4708] [Medline: 26842694]
107. Bindoff I, de Salas K, Peterson G, Ling T, Lewis I, Wells L, et al. Quittr: The design of a video game to support smoking cessation. JMIR Serious Games 2016 Dec 01;4(2):e19 [FREE Full text] [doi: 10.2196/games.6258] [Medline: 27908844]
108. Graafland M, Dankbaar M, Mert A, Lagro J, De Wit-Zuurendonk L, Schuit S, et al. How to systematically assess serious games applied to health care. JMIR Serious Games 2014;2(2):e11 [FREE Full text] [doi: 10.2196/games.3825] [Medline: 25654163]
109. Tuti T, Winters N, Muinga N, Wanyama C, English M, Paton C. Evaluation of adaptive feedback in a smartphone-based serious game on health care providers' knowledge gain in neonatal emergency care: protocol for a randomized controlled trial. JMIR Res Protoc 2019 Jul 26;8(7):e13034 [FREE Full text] [doi: 10.2196/13034] [Medline: 31350837]
110. Shepherd M, Fleming T, Lucassen M, Stasiak K, Lambie I, Merry SN. The design and relevance of a computerized gamified depression therapy program for indigenous māori adolescents. JMIR Serious Games 2015;3(1):e1 [FREE Full text] [doi: 10.2196/games.3804] [Medline: 25736225]
111. de Vette F, Tabak M, Dekker-van WM, Vollenbroek-Hutten M. Engaging elderly people in telemedicine through gamification. JMIR Serious Games 2015 Dec 18;3(2):e9 [FREE Full text] [doi: 10.2196/games.4561] [Medline: 26685287]
112. Brauner P, Ziefle M. Serious motion-based exercise games for older adults: evaluation of usability, performance, and pain mitigation. JMIR Serious Games 2020 Apr 01;8(2):e14182 [FREE Full text] [doi: 10.2196/14182] [Medline: 32234703]
113. Abraham O, LeMay S, Bittner S, Thakur T, Stafford H, Brown R. Investigating serious games that incorporate medication use for patients: systematic literature review. JMIR Serious Games 2020 Apr 29;8(2):e16096 [FREE Full text] [doi: 10.2196/16096] [Medline: 32347811]
114. Aksoy E. Comparing the effects on learning outcomes of tablet-based and virtual reality-based serious gaming modules for basic life support training: randomized trial. JMIR Serious Games 2019 May 01;7(2):e13442 [FREE Full text] [doi: 10.2196/13442] [Medline: 31042153]
115. Allam A, Kostova Z, Nakamoto K, Schulz PJ. The effect of social support features and gamification on a Web-based intervention for rheumatoid arthritis patients: randomized controlled trial. J Med Internet Res 2015 Jan 09;17(1):e14 [FREE Full text] [doi: 10.2196/jmir.3510] [Medline: 25574939]
116. AlMarshedi A, Wills G, Ranchhod A. Gamifying self-management of chronic illnesses: a mixed-methods study. JMIR Serious Games 2016 Sep 09;4(2):e14 [FREE Full text] [doi: 10.2196/games.5943] [Medline: 27612632]
117. Almeida J, Nunes F. The practical work of ensuring effective use of serious games in a rehabilitation clinic: a qualitative study. JMIR Rehabil Assist Technol 2020 Feb 21;7(1):e15428 [FREE Full text] [doi: 10.2196/15428] [Medline: 32130177]
118. Alyami H, Alawami M, Lyndon M, Alyami M, Coomarasamy CL, Henning M, et al. Impact of using a 3D visual metaphor serious game to teach history-taking content to medical students: longitudinal mixed methods pilot study. JMIR Serious Games 2019 Sep 26;7(3):e13748 [FREE Full text] [doi: 10.2196/13748] [Medline: 31573895]
119. Amantini SN, Montilha AA, Antonelli BC, Leite KT, Rios D, Cruvinel T, et al. Using augmented reality to motivate oral hygiene practice in children: protocol for the development of a serious game. JMIR Res Protoc 2020 Jan 17;9(1):e10987 [FREE Full text] [doi: 10.2196/10987] [Medline: 31951216]
120. Antoniou PE, Arfaras G, Pandria N, Athanasiou A, Ntakakis G, Babatsikos E, et al. Biosensor real-time affective analytics in virtual and mixed reality medical education serious games: cohort study. JMIR Serious Games 2020 Sep 02;8(3):e17823 [FREE Full text] [doi: 10.2196/17823] [Medline: 32876575]
121. Badawy SM, Cronin RM, Hankins J, Crosby L, DeBaun M, Thompson AA, et al. Patient-centered eHealth interventions for children, adolescents, and adults with sickle cell disease: systematic review. J Med Internet Res 2018 Jul 19;20(7):e10940 [FREE Full text] [doi: 10.2196/10940] [Medline: 30026178]
122. Bang C, Nam Y, Ko EJ, Lee W, Kim B, Choi Y, et al. A serious game-derived index for detecting children with heterogeneous developmental disabilities: randomized controlled trial. JMIR Serious Games 2019 Oct 24;7(4):e14924 [FREE Full text] [doi: 10.2196/14924] [Medline: 31651408]
123. Benda NC, Kellogg KM, Hoffman DJ, Fairbanks RJ, Auguste T. Lessons learned from an evaluation of serious gaming as an alternative to mannequin-based simulation technology: randomized controlled trial. JMIR Serious Games 2020 Sep 28;8(3):e21123 [FREE Full text] [doi: 10.2196/21123] [Medline: 32985993]
124. Besoain F, Perez-Navarro A, Jacques Aviñó C, Caylà JA, Barriga NA, de Olalla PG. Prevention of HIV and other sexually transmitted infections by geofencing and contextualized messages with a gamified app, UBESAFE: design and creation study. JMIR Mhealth Uhealth 2020 Mar 17;8(3):e14568 [FREE Full text] [doi: 10.2196/14568] [Medline: 32181752]
125. Billis A, Pandria N, Mouratoglou SA, Konstantinidis E, Bamidis P. Development of cognitive and physical exercise systems, clinical recordings, large-scale data analytics, and virtual coaching for heart failure patients: protocol for the BioTechCOACH-ForALL project. JMIR Res Protoc 2020 May 04;9(5):e17714 [FREE Full text] [doi: 10.2196/17714] [Medline: 32364512]
126. Blackburne T, Rodriguez A, Johnstone SJ. A serious game to increase healthy food consumption in overweight or obese adults: randomized controlled trial. JMIR Serious Games 2016 Jul 13;4(2):e10 [FREE Full text] [doi: 10.2196/games.5708] [Medline: 27417192]

127. Block G, Azar KM, Block TJ, Romanelli RJ, Carpenter H, Hopkins D, et al. A Fully Automated Diabetes Prevention Program, Alive-PD: Program design and randomized controlled trial protocol. JMIR Res Protoc 2015 Jan 21;4(1):e3 [FREE Full text] [doi: 10.2196/resprot.4046] [Medline: 25608692]
128. Boendermaker WJ, Gladwin TE, Peeters M, Prins PJ, Wiers RW. Training working memory in adolescents using serious game elements: pilot randomized controlled trial. JMIR Serious Games 2018 May 23;6(2):e10 [FREE Full text] [doi: 10.2196/games.8364] [Medline: 29792294]
129. Boendermaker WJ, Maceiras SS, Boffo M, Wiers RW. Attentional bias modification with serious game elements: evaluating the shots game. JMIR Serious Games 2016 Dec 06;4(2):e20 [FREE Full text] [doi: 10.2196/games.6464] [Medline: 27923780]
130. Bray L, Sharpe A, Gichuru P, Fortune PM, Blake L, Appleton V. The acceptability and impact of the xploro digital therapeutic platform to inform and prepare children for planned procedures in a hospital: before and after evaluation study. J Med Internet Res 2020 Aug 11;22(8):e17367 [FREE Full text] [doi: 10.2196/17367] [Medline: 32780025]
131. Brox E, Konstantinidis S, Evertsen G. User-centered design of serious games for older adults following 3 years of experience with exergames for seniors: a study design. JMIR Serious Games 2017 Jan 11;5(1):e2 [FREE Full text] [doi: 10.2196/games.6254] [Medline: 28077348]
132. Buffel C, van Aalst J, Bangels AM, Toelen J, Allegaert K, Verschueren S, et al. A web-based serious game for health to reduce perioperative anxiety and pain in children (CliniPup): pilot randomized controlled trial. JMIR Serious Games 2019 Jun 01;7(2):e12431 [FREE Full text] [doi: 10.2196/12431] [Medline: 31199324]
133. Buijs-Spanjers KR, Hegge HH, Cnossen F, Jaarsma DA, de Rooij SE. Reasons to engage in and learning experiences from different play strategies in a web-based serious game on delirium for medical students: mixed methods design. JMIR Serious Games 2020 Jul 29;8(3):e18479 [FREE Full text] [doi: 10.2196/18479] [Medline: 32723710]
134. Buijs-Spanjers KR, Hegge HH, Jansen CJ, Hoogendoorn E, de Rooij SE. A web-based serious game on delirium as an educational intervention for medical students: randomized controlled trial. JMIR Serious Games 2018 Oct 26;6(4):e17 [FREE Full text] [doi: 10.2196/games.9886] [Medline: 30368436]
135. Bul KC, Kato PM, Van der Oord S, Danckaerts M, Vreeke LJ, Willems A, et al. Behavioral outcome effects of serious gaming as an adjunct to treatment for children with attention-deficit/hyperactivity disorder: a randomized controlled trial. J Med Internet Res 2016 Feb 16;18(2):e26 [FREE Full text] [doi: 10.2196/jmir.5173] [Medline: 26883052]
136. Calderita LV, Manso LJ, Bustos P, Suárez-Mejías C, Fernández F, Bandera A. THERAPIST: Towards an autonomous socially interactive robot for motor and neurorehabilitation therapies for children. JMIR Rehabil Assist Technol 2014 Oct 07;1(1):e1 [FREE Full text] [doi: 10.2196/rehab.3151] [Medline: 28582242]
137. Carnicer-Pont D, Loureiro-Varela E, Manresa JM, Martinez M, Avecilla À, Montero-Pons L, et al. The Notijoves Project: Protocol for a randomized controlled trial about new communication technologies and gamification to promote partner notification of sexually transmitted infections among young people. JMIR Res Protoc 2019 May 26;8(6):e12896 [FREE Full text] [doi: 10.2196/12896] [Medline: 31199306]
138. Castro-Sánchez E, Sood A, Rawson TM, Firth J, Holmes AH. Forecasting implementation, adoption, and evaluation challenges for an electronic game-based antimicrobial stewardship intervention: co-design workshop with multidisciplinary stakeholders. J Med Internet Res 2019 Jun 04;21(6):e13365 [FREE Full text] [doi: 10.2196/13365] [Medline: 31165712]
139. Chang HY, Poh DY, Wong LL, Yap JY, Yap KY. Student preferences on gaming aspects for a serious game in pharmacy practice education: a cross-sectional study. JMIR Med Educ 2015 May 11;1(1):e2 [FREE Full text] [doi: 10.2196/mededu.3754] [Medline: 27731304]
140. Cheek C, Fleming T, Lucassen MF, Bridgman H, Stasiak K, Shepherd M, et al. Integrating health behavior theory and design elements in serious games. JMIR Ment Health 2015;2(2):e11 [FREE Full text] [doi: 10.2196/mental.4133] [Medline: 26543916]
141. Chen J, Ho E, Jiang Y, Whittaker R, Yang T, Bullen C. A mobile social network-based smoking cessation intervention for Chinese male smokers: protocol for a pilot randomized controlled trial. JMIR Res Protoc 2020 Sep 18;9(9):e18071 [FREE Full text] [doi: 10.2196/18071] [Medline: 32945261]
142. Chen J, Ho E, Jiang Y, Whittaker R, Yang T, Bullen C. Mobile social network-based smoking cessation intervention for Chinese male smokers: pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Oct 23;8(10):e17522 [FREE Full text] [doi: 10.2196/17522] [Medline: 33095184]
143. Cheng VW, Davenport T, Johnson D, Vella K, Hickie IB. Gamification in apps and technologies for improving mental health and well-being: systematic review. JMIR Ment Health 2019 Jun 26;6(6):e13717 [FREE Full text] [doi: 10.2196/13717] [Medline: 31244479]
144. Choi WT, Yu DK, Wong T, Lantta T, Yang M, Välimäki M. Habits and attitudes of video gaming and information technology use in people with schizophrenia: cross-sectional survey. J Med Internet Res 2020 Jul 22;22(7):e14865 [FREE Full text] [doi: 10.2196/14865] [Medline: 32459646]
145. Chon SH, Timmermann F, Dratsch T, Schuelper N, Plum P, Berlth F, et al. Serious games in surgical medical education: a virtual emergency department as a tool for teaching clinical reasoning to medical students. JMIR Serious Games 2019 Mar 05;7(1):e13028 [FREE Full text] [doi: 10.2196/13028] [Medline: 30835239]

146. Clausen JD, Nahen N, Horstmann H, Lasch F, Krutsch W, Krettek C, et al. Improving maximal strength in the initial postoperative phase after anterior cruciate ligament reconstruction surgery: randomized controlled trial of an app-based serious gaming approach. JMIR Serious Games 2020 Jan 24;8(1):e14282 [FREE Full text] [doi: 10.2196/14282] [Medline: 32012046]
147. Constantinescu G, Loewen I, King B, Brodt C, Hodgetts W, Rieger J. Designing a mobile health app for patients with dysphagia following head and neck cancer: a qualitative study. JMIR Rehabil Assist Technol 2017 Mar 24;4(1):e3 [FREE Full text] [doi: 10.2196/rehab.6319] [Medline: 28582245]
148. Creutzfeldt J, Hedman L, Felländer-Tsai L. Cardiopulmonary resuscitation training by avatars: a qualitative study of medical students' experiences using a multiplayer virtual world. JMIR Serious Games 2016 Dec 16;4(2):e22 [FREE Full text] [doi: 10.2196/games.6448] [Medline: 27986645]
149. Creutzfeldt J, Hedman L, Heinrichs L, Youngblood P, Felländer-Tsai L. Cardiopulmonary resuscitation training in high school using avatars in virtual worlds: an international feasibility study. J Med Internet Res 2013 Jan 14;15(1):e9 [FREE Full text] [doi: 10.2196/jmir.1715] [Medline: 23318253]
150. Creutzfeldt J, Hedman L, Medin C, Heinrichs WL, Felländer-Tsai L. Exploring virtual worlds for scenario-based repeated team training of cardiopulmonary resuscitation in medical students. J Med Internet Res 2010 Sep 03;12(3):e38 [FREE Full text] [doi: 10.2196/jmir.1426] [Medline: 20813717]
151. Curtis KE, Lahiri S, Brown KE. Targeting parents for childhood weight management: development of a theory-driven and user-centered healthy eating app. JMIR Mhealth Uhealth 2015 Jun 18;3(2):e69 [FREE Full text] [doi: 10.2196/mhealth.3857] [Medline: 26088692]
152. Dargél AA, Mosconi E, Masson M, Plaze M, Taieb F, Von Platen C, et al. Toi Même, a mobile health platform for measuring bipolar illness activity: protocol for a feasibility study. JMIR Res Protoc 2020 Aug 18;9(8):e18818 [FREE Full text] [doi: 10.2196/18818] [Medline: 32638703]
153. Diehl LA, Souza RM, Alves JB, Gordan PA, Esteves RZ, Jorge ML, et al. InsuOnline, a serious game to teach insulin therapy to primary care physicians: design of the game and a randomized controlled trial for educational validation. JMIR Res Protoc 2013 Jan 21;2(1):e5 [FREE Full text] [doi: 10.2196/resprot.2431] [Medline: 23612462]
154. Diehl LA, Souza RM, Gordan PA, Esteves RZ, Coelho IC. InsuOnline, an electronic game for medical education on insulin therapy: a randomized controlled trial with primary care physicians. J Med Internet Res 2017 Mar 09;19(3):e72 [FREE Full text] [doi: 10.2196/jmir.6944] [Medline: 28279950]
155. Edwards EA, Caton H, Lumsden J, Rivas C, Steed L, Pirunsarn Y, et al. Creating a theoretically grounded, gamified health app: lessons from developing the cigbreak smoking cessation mobile phone game. JMIR Serious Games 2018 Nov 29;6(4):e10252 [FREE Full text] [doi: 10.2196/10252] [Medline: 30497994]
156. Eleftheriou A, Bullock S, Graham CA, Ingham R. Using computer simulations for investigating a sex education intervention: an exploratory study. JMIR Serious Games 2017 May 03;5(2):e9 [FREE Full text] [doi: 10.2196/games.6598] [Medline: 28468747]
157. El-Hilly AA, Iqbal SS, Ahmed M, Sherwani Y, Muntasir M, Siddiqui S, et al. Game on? Smoking cessation through the gamification of mHealth: a longitudinal qualitative study. JMIR Serious Games 2016 Oct 24;4(2):e18 [FREE Full text] [doi: 10.2196/games.5678] [Medline: 27777216]
158. Espinosa-Curiel IE, Pozas-Bogarin EE, Martínez-Miranda J, Pérez-Espinosa H. Relationship between children's enjoyment, user experience satisfaction, and learning in a serious video game for nutrition education: empirical pilot study. JMIR Serious Games 2020 Sep 17;8(3):e21813 [FREE Full text] [doi: 10.2196/21813] [Medline: 32940609]
159. Fadda M, Galimberti E, Fiordelli M, Schulz PJ. Evaluation of a mobile phone-based intervention to increase parents' knowledge about the measles-mumps-rubella vaccination and their psychological empowerment: mixed-method approach. JMIR Mhealth Uhealth 2018 Mar 07;6(3):e59 [FREE Full text] [doi: 10.2196/mhealth.8263] [Medline: 29514772]
160. Fijačko N, Gosak L, Cilar L, Novšak A, Creber RM, Skok P, et al. The effects of gamification and oral self-care on oral hygiene in children: systematic search in app stores and evaluation of apps. JMIR Mhealth Uhealth 2020 Jul 08;8(7):e16365 [FREE Full text] [doi: 10.2196/16365] [Medline: 32673235]
161. Flogie A, Aberšek B, Aberšek MK, Lanyi CS, Pesek I. Development and evaluation of intelligent serious games for children with learning difficulties: observational study. JMIR Serious Games 2020 Apr 16;8(2):e13190 [FREE Full text] [doi: 10.2196/13190] [Medline: 32297864]
162. Floryan M, Chow PI, Schueller SM, Ritterband LM. The model of gamification principles for digital health interventions: evaluation of validity and potential utility. J Med Internet Res 2020 Jun 10;22(6):e16506 [FREE Full text] [doi: 10.2196/16506] [Medline: 32519965]
163. Giunti G, Fernández E, Zubiete E, Romero O. Supply and demand in mHealth apps for persons with multiple sclerosis: systematic search in app stores and scoping literature review. JMIR Mhealth Uhealth 2018 May 23;6(5):e10512 [FREE Full text] [doi: 10.2196/10512] [Medline: 29792295]
164. Giunti G, Mylonopoulou V, Romero O. More stamina, a gamified mHealth solution for persons with multiple sclerosis: research through design. JMIR Mhealth Uhealth 2018 Mar 02;6(3):e51 [FREE Full text] [doi: 10.2196/mhealth.9437] [Medline: 29500159]

165. Gonze B, Padovani R, Simoes M, Lauria V, Proença N, Sperandio E, et al. Use of a smartphone app to increase physical activity levels in insufficiently active adults: Feasibility Sequential Multiple Assignment Randomized Trial (SMART). JMIR Res Protoc 2020 Oct 23;9(10):e14322 [FREE Full text] [doi: 10.2196/14322] [Medline: 33094733]
166. Gorsic M, Clapp JD, Darzi A, Novak D. A brief measure of interpersonal interaction for 2-player serious games: questionnaire validation. JMIR Serious Games 2019 Jul 23;7(3):e12788 [FREE Full text] [doi: 10.2196/12788] [Medline: 31339107]
167. Hansen WB, Scheier LM. Specialized smartphone intervention apps: Review of 2014 to 2018 NIH Funded Grants. JMIR Mhealth Uhealth 2019 Jul 29;7(7):e14655 [FREE Full text] [doi: 10.2196/14655] [Medline: 31359866]
168. Hendrie GA, James-Martin G, Williams G, Brindal E, Whyte B, Crook A. The Development of VegEze: Smartphone app to increase vegetable consumption in Australian adults. JMIR Form Res 2019 Mar 27;3(1):e10731 [FREE Full text] [doi: 10.2196/10731] [Medline: 30916653]
169. Hernández-Reyes A, Cámara-Martos F, Recio G, Molina-Luque R, Romero-Saldaña M, Rojas R. Push notifications from a mobile app to improve the body composition of overweight or obese women: randomized controlled trial. JMIR Mhealth Uhealth 2020 Feb 12;8(2):e13747 [FREE Full text] [doi: 10.2196/13747] [Medline: 32049065]
170. Hilliard ME, Hahn A, Ridge AK, Eakin MN, Riekert KA. User preferences and design recommendations for an mHealth app to promote cystic fibrosis self-management. JMIR Mhealth Uhealth 2014 Oct 24;2(4):e44 [FREE Full text] [doi: 10.2196/mhealth.3599] [Medline: 25344616]
171. Hoffmann A, Faust-Christmann CA, Zolynski G, Bleser G. Toward gamified pain management apps: mobile application rating scale-based quality assessment of pain-mentor's first prototype through an expert study. JMIR Form Res 2020 May 26;4(5):e13170 [FREE Full text] [doi: 10.2196/13170] [Medline: 32452803]
172. Holzmann SL, Schäfer H, Plecher DA, Stecher L, Klinker G, Groh G, et al. Serious games for nutritional education: online survey on preferences, motives, and behaviors among young adults at university. JMIR Serious Games 2020 Jun 03;8(2):e16216 [FREE Full text] [doi: 10.2196/16216] [Medline: 32490847]
173. Hooglugt F, Ludden GD. A mobile app adopting an identity focus to promote physical activity (MoveDaily): iterative design study. JMIR Mhealth Uhealth 2020 Jun 15;8(6):e16720 [FREE Full text] [doi: 10.2196/16720] [Medline: 32538795]
174. Horvath KJ, Amico KR, Erickson D, Ecklund AM, Martinka A, DeWitt J, et al. Thrive with me: protocol for a randomized controlled trial to test a peer support intervention to improve antiretroviral therapy adherence among men who have sex with men. JMIR Res Protoc 2018 May 31;7(5):e10182 [FREE Full text] [doi: 10.2196/10182] [Medline: 29853437]
175. Hunter JF, Olah MS, Williams AL, Parks AC, Pressman SD. Effect of brief biofeedback via a smartphone app on stress recovery: randomized experimental study. JMIR Serious Games 2019 Nov 26;7(4):e15974 [FREE Full text] [doi: 10.2196/15974] [Medline: 31769761]
176. Idriss M, Tannous H, Istrate D, Perrochon A, Salle JY, Ho Ba Tho MC, et al. Rehabilitation-oriented serious game development and evaluation guidelines for musculoskeletal disorders. JMIR Serious Games 2017 Jul 04;5(3):e14 [FREE Full text] [doi: 10.2196/games.7284] [Medline: 28676468]
177. Ingadottir B, Blondal K, Thue D, Zoega S, Thylen I, Jaarsma T. Development, usability, and efficacy of a serious game to help patients learn about pain management after surgery: an evaluation study. JMIR Serious Games 2017 May 10;5(2):e10 [FREE Full text] [doi: 10.2196/games.6894] [Medline: 28490419]
178. Jaunay LB, Zerr P, Peguin L, Renouard L, Ivanoff AS, Picard H, et al. Development and evaluation of a new serious game for continuing medical education of general practitioners (Hygie): double-blinded randomized controlled trial. J Med Internet Res 2019 Nov 20;21(11):e12669 [FREE Full text] [doi: 10.2196/12669] [Medline: 31746775]
179. Jones M, Denisova A, Mitchell S, Owen T. Acceptability of a plasticity-focused serious game intervention for posttraumatic stress disorder: user requirements analysis. JMIR Serious Games 2019 Apr 16;7(2):e11909 [FREE Full text] [doi: 10.2196/11909] [Medline: 30990461]
180. Boulos MN, Gammon S, Dixon MC, MacRury SM, Fergusson MJ, Rodrigues FM, et al. Digital games for type 1 and type 2 diabetes: underpinning theory with three illustrative examples. JMIR Serious Games 2015 Mar 18;3(1):e3 [FREE Full text] [doi: 10.2196/games.3930] [Medline: 25791276]
181. Kelders SM, Sommers-Spijkerman M, Goldberg J. Investigating the direct impact of a gamified versus nongamified well-being intervention: an exploratory experiment. J Med Internet Res 2018 Jul 26;20(7):e247 [FREE Full text] [doi: 10.2196/jmir.9923] [Medline: 30049669]
182. Talantikite WK, Renou D, Magdelaine A, Harang W, Zmirou-Navier D, Deguen S. Development of serious games (Equit'Game) to address health and environmental inequalities: protocol for an app-delivered program to perform a territorial diagnosis. JMIR Res Protoc 2020 Jan 07;9(1):e11786 [FREE Full text] [doi: 10.2196/11786] [Medline: 31909726]
183. Klincova M, Harazim H, Schwarz D, Kosinova M, Smekalova O, Stourac P. What can be achieved with motivation-based teaching of medical students? A monocentric retrospective audit of retention among highly motivated graduates who underwent the learning-by-doing concept in anesthesiology and intensive care medicine. JMIR Serious Games 2019 Apr 09;7(2):e10155 [FREE Full text] [doi: 10.2196/10155] [Medline: 30964443]
184. Ko J, Lee JJ, Jang SW, Yun Y, Kang S, Shin DA, et al. An Epiduroscopy Simulator Based on a Serious Game for Spatial Cognitive Training (EpiduroSIM): User-centered design approach. JMIR Serious Games 2019 Aug 27;7(3):e12678 [FREE Full text] [doi: 10.2196/12678] [Medline: 31456580]

185. Kouwenhoven-Pasmooij TA, Robroek SJ, Ling SW, van Rosmalen J, van Rossum EF, Burdorf A, et al. A blended web-based gaming intervention on changes in physical activity for overweight and obese employees: influence and usage in an experimental pilot study. JMIR Serious Games 2017 Apr 03;5(2):e6 [FREE Full text] [doi: 10.2196/games.6421] [Medline: 28373157]
186. Kuipers DA, Terlouw G, Wartena BO, van 't Veer JT, Prins JT, Pierie JP. The role of transfer in designing games and simulations for health: systematic review. JMIR Serious Games 2017 Nov 24;5(4):e23 [FREE Full text] [doi: 10.2196/games.7880] [Medline: 29175812]
187. Lawitschka A, Buehrer S, Bauer D, Peters K, Silbernagl M, Zubarovskaya N, et al. A web-based mobile app (INTERACCT App) for adolescents undergoing cancer and hematopoietic stem cell transplantation aftercare to improve the quality of medical information for clinicians: observational study. JMIR Mhealth Uhealth 2020 Jun 30;8(6):e18781 [FREE Full text] [doi: 10.2196/18781] [Medline: 32602847]
188. Lázaro E, Amayra I, López-Paz JF, Martínez O, Pérez Alvarez M, Berrocoso S, et al. Using a virtual serious game (Deusto-e-motion1.0) to assess the theory of mind in primary school children: observational descriptive study. JMIR Serious Games 2020 Apr 02;8(2):e12971 [FREE Full text] [doi: 10.2196/12971] [Medline: 32238330]
189. Lee L, Chao Y, Huang C, Fang J, Wang S, Chuang C, et al. Cognitive style and mobile e-learning in emergent otorhinolaryngology-head and neck surgery disorders for millennial undergraduate medical students: randomized controlled trial. J Med Internet Res 2018 Feb 13;20(2):e56 [FREE Full text] [doi: 10.2196/jmir.8987] [Medline: 29439943]
190. LeGrand S, Knudtson K, Benkeser D, Muessig K, Mcgee A, Sullivan PS, et al. Testing the efficacy of a social networking gamification app to improve pre-exposure prophylaxis adherence (p3: prepared, protected, empowered): protocol for a randomized controlled trial. JMIR Res Protoc 2018 Dec 18;7(12):e10448 [FREE Full text] [doi: 10.2196/10448] [Medline: 30563818]
191. LeGrand S, Muessig KE, McNulty T, Soni K, Knudtson K, Lemann A, et al. Epic Allies: Development of a gaming app to improve antiretroviral therapy adherence among young hiv-positive men who have sex with men. JMIR Serious Games 2016 May 13;4(1):e6 [FREE Full text] [doi: 10.2196/games.5687] [Medline: 27178752]
192. Lin Y, Tudor-Sfetea C, Siddiqui S, Sherwani Y, Ahmed M, Eisingerich AB. Effective behavioral changes through a digital mhealth app: exploring the impact of hedonic well-being, psychological empowerment and inspiration. JMIR Mhealth Uhealth 2018 Jun 15;6(6):e10024 [FREE Full text] [doi: 10.2196/10024] [Medline: 29907557]
193. Lindqvist AK, Rutberg S. One step forward: development of a program promoting active school transportation. JMIR Res Protoc 2018 May 08;7(5):e123 [FREE Full text] [doi: 10.2196/resprot.9505] [Medline: 29739733]
194. Lucassen M, Samra R, Iacovides I, Fleming T, Shepherd M, Stasiak K, et al. How LGBT+ young people use the internet in relation to their mental health and envisage the use of e-therapy: exploratory study. JMIR Serious Games 2018 Dec 21;6(4):e11249 [FREE Full text] [doi: 10.2196/11249] [Medline: 30578194]
195. Lukhele BW, Musumari P, El-Saaidi C, Techasrivichien T, Suguimoto SP, Kihara MO, et al. Efficacy of mobile serious games in increasing hiv risk perception in swaziland: a randomized control trial (SGprev Trial) research protocol. JMIR Res Protoc 2016 Nov 22;5(4):e224 [FREE Full text] [doi: 10.2196/resprot.6543] [Medline: 27876685]
196. Lumsden J, Skinner A, Coyle D, Lawrence N, Munafo M. Attrition from web-based cognitive testing: a repeated measures comparison of gamification techniques. J Med Internet Res 2017 Nov 22;19(11):e395 [FREE Full text] [doi: 10.2196/jmir.8473] [Medline: 29167090]
197. Mairs RA, Bekker MJ, Patolo T, Hopkins SA, Cowley-Malcolm ET, Perese LM, et al. Gamifying parenting education using an app developed for pacific and other New Zealand families (Play Kindly): qualitative study. JMIR Serious Games 2020 Jun 10;8(2):e15647 [FREE Full text] [doi: 10.2196/15647] [Medline: 32519973]
198. Manning V, Piercy H, Garfield JB, Lubman DI. Personalized approach bias modification smartphone app ("SWIPE") to reduce alcohol use among people drinking at hazardous or harmful levels: protocol for an open-label feasibility study. JMIR Res Protoc 2020 Aug 14;9(8):e21278 [FREE Full text] [doi: 10.2196/21278] [Medline: 32795989]
199. Marin-Gomez F, Garcia-Moreno R, Mayos-Fernandez A, Flores-Mateo G, Granado-Font E, Barrera Uriarte M, et al. Exploring efficacy of a serious game (Tobbstop) for smoking cessation during pregnancy: randomized controlled trial. JMIR Serious Games 2019 Mar 27;7(1):e12835 [FREE Full text] [doi: 10.2196/12835] [Medline: 30916655]
200. Mejia CM, Acland D, Buzdugan R, Grimball R, Natoli L, McGrath MR, et al. An intervention using gamification to increase human immunodeficiency virus and sexually transmitted infection screening among young men who have sex with men in california: rationale and design of stick to it. JMIR Res Protoc 2017 Jul 17;6(7):e140 [FREE Full text] [doi: 10.2196/resprot.8064] [Medline: 28716771]
201. Mubin O, Alnajjar F, Jishtu N, Alsinglawi B, Al Mahmud A. Exoskeletons with virtual reality, augmented reality, and gamification for stroke patients' rehabilitation: systematic review. JMIR Rehabil Assist Technol 2019 Sep 08;6(2):e12010 [FREE Full text] [doi: 10.2196/12010] [Medline: 31586360]
202. Nicolaidou I, Antoniades A, Constantinou R, Marangos C, Kyriacou E, Bamidis P, et al. A virtual emergency telemedicine serious game in medical training: a quantitative, professional feedback-informed evaluation study. J Med Internet Res 2015 Jun 17;17(6):e150 [FREE Full text] [doi: 10.2196/jmir.3667] [Medline: 26084866]

203. Notebaert L, Grafton B, Clarke PJ, Rudaizky D, Chen NT, MacLeod C. Emotion-in-motion, a novel approach for the modification of attentional bias: an experimental proof-of-concept study. JMIR Serious Games 2018 Nov 28;6(4):e10993 [FREE Full text] [doi: 10.2196/10993] [Medline: 30487121]
204. Nurmi J, Knittle K, Ginchev T, Khattak F, Helf C, Zwickl P, et al. Engaging users in the behavior change process with digitalized motivational interviewing and gamification: development and feasibility testing of the precious app. JMIR Mhealth Uhealth 2020 Jan 30;8(1):e12884 [FREE Full text] [doi: 10.2196/12884] [Medline: 32003750]
205. Oakley-Girvan I, Lavista JM, Miller Y, Davis S, Acle C, Hancock J, et al. Evaluation of a Mobile Device Survey System for Behavioral Risk Factors (SHAPE): App development and usability study. JMIR Form Res 2019 Jan 11;3(1):e10246 [FREE Full text] [doi: 10.2196/10246] [Medline: 30684441]
206. Oliveri S, Mainetti R, Gorini A, Cutica I, Candiani G, Borghese NA, et al. Serious games for improving genetic literacy and genetic risk awareness in the general public: protocol for a randomized controlled trial. JMIR Res Protoc 2018 Dec 18;7(12):e189 [FREE Full text] [doi: 10.2196/resprot.9288] [Medline: 30563813]
207. Patchen L, Ellis L, Ma TX, Ott C, Chang KH, Araya B, et al. Engaging african American youth in the development of a serious mobile game for sexual health education: mixed methods study. JMIR Serious Games 2020 Jan 31;8(1):e16254 [FREE Full text] [doi: 10.2196/16254] [Medline: 32012041]
208. Payne HE, Moxley VB, MacDonald E. Health behavior theory in physical activity game apps: a content analysis. JMIR Serious Games 2015 Jul 13;3(2):e4 [FREE Full text] [doi: 10.2196/games.4187] [Medline: 26168926]
209. Perry Y, Werner-Seidler A, Calear A, Mackinnon A, King C, Scott J, et al. Preventing depression in final year secondary students: school-based randomized controlled trial. J Med Internet Res 2017 Nov 02;19(11):e369 [FREE Full text] [doi: 10.2196/jmir.8241] [Medline: 29097357]
210. Phungoen P, Promto S, Chanthawatthanarak S, Maneepong S, Apiratwarakul K, Kotruchin P, et al. Precourse preparation using a serious smartphone game on advanced life support knowledge and skills: randomized controlled trial. J Med Internet Res 2020 Mar 09;22(3):e16987 [FREE Full text] [doi: 10.2196/16987] [Medline: 32149711]
211. Rasche P, Schlomann A, Mertens A. Who is still playing Pokémon Go? A web-based survey. JMIR Serious Games 2017 Apr 05;5(2):e7 [FREE Full text] [doi: 10.2196/games.7197] [Medline: 28381393]
212. Rivera J, McPherson AC, Hamilton J, Birken C, Coons M, Peters M, et al. User-centered design of a mobile app for weight and health management in adolescents with complex health needs: qualitative study. JMIR Form Res 2018 Apr 04;2(1):e7 [FREE Full text] [doi: 10.2196/formative.8248] [Medline: 30684409]
213. Sabben G, Akelo V, Mudhune V, Ondeng'e K, Ndivo R, Stephenson R, et al. A smartphone game to prevent HIV among young Africans: protocol for a randomized pilot study of a mobile intervention. JMIR Res Protoc 2019 Mar 27;8(3):e11209 [FREE Full text] [doi: 10.2196/11209] [Medline: 30916661]
214. Sabben G, Mudhune V, Ondeng'e K, Odero I, Ndivo R, Akelo V, et al. A smartphone game to prevent HIV among young Africans (Tumaini): assessing intervention and study acceptability among adolescents and their parents in a randomized controlled trial. JMIR Mhealth Uhealth 2019 May 21;7(5):e13049 [FREE Full text] [doi: 10.2196/13049] [Medline: 31115348]
215. Sanders I, Short CE, Bogomolova S, Stanford T, Plotnikoff R, Vandelanotte C, et al. Characteristics of adopters of an online social networking physical activity mobile phone app: cluster analysis. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e12484 [FREE Full text] [doi: 10.2196/12484] [Medline: 31162130]
216. Santana C, Freitas A, Barreto G, De Avelar I, Mazaro-Costa R, Bueno G, et al. Serious game on a smartphone for adolescents undergoing hemodialysis: development and evaluation. JMIR Serious Games 2020 Sep 14;8(3):e17979 [FREE Full text] [doi: 10.2196/17979] [Medline: 32924950]
217. Santos J, Vairinhos M, Jesus LM. Treating children with speech sound disorders: development of a tangible artefact prototype. JMIR Serious Games 2019 Dec 05;7(4):e13861 [FREE Full text] [doi: 10.2196/13861] [Medline: 31804185]
218. Schakel L, Veldhuijzen DS, van Middendorp H, Prins C, Drittij AM, Vrieling F, et al. An internet-based psychological intervention with a serious game to improve vitality, psychological and physical condition, and immune function in healthy male adults: randomized controlled trial. J Med Internet Res 2020 Jul 24;22(7):e14861 [FREE Full text] [doi: 10.2196/14861] [Medline: 32706667]
219. Sezgin E, Lin S. Technology-based interventions, assessments, and solutions for safe driving training for adolescents: rapid review. JMIR Mhealth Uhealth 2019 Jan 24;7(1):e11942 [FREE Full text] [doi: 10.2196/11942] [Medline: 30679149]
220. Sharifzadeh N, Kharrazi H, Nazari E, Tabesh H, Khodabandeh M, Heidari S, et al. Health education serious games targeting health care providers, patients, and public health users: scoping review. JMIR Serious Games 2020 Mar 05;8(1):e13459 [FREE Full text] [doi: 10.2196/13459] [Medline: 32134391]
221. Simões M, Bernardes M, Barros F, Castelo-Branco M. Virtual travel training for autism spectrum disorder: proof-of-concept interventional study. JMIR Serious Games 2018 Mar 20;6(1):e5 [FREE Full text] [doi: 10.2196/games.8428] [Medline: 29559425]
222. Spook J, Paulussen T, Kok G, van Empelen P. Evaluation of a serious self-regulation game intervention for overweight-related behaviors ("Balance It"): a pilot study. J Med Internet Res 2016 Sep 26;18(9):e225 [FREE Full text] [doi: 10.2196/jmir.4964] [Medline: 27670222]

223. Steiner B, Elgert L, Saalfeld B, Wolf K. Gamification in rehabilitation of patients with musculoskeletal diseases of the shoulder: scoping review. JMIR Serious Games 2020 Aug 25;8(3):e19914 [FREE Full text] [doi: 10.2196/19914] [Medline: 32840488]
224. Stragier J, Vandewiele G, Coppens P, Ongenae F, Van den Broeck W, De Turck F, et al. Data mining in the development of mobile health apps: assessing in-app navigation through markov chain analysis. J Med Internet Res 2019 Jun 07;21(6):e11934 [FREE Full text] [doi: 10.2196/11934] [Medline: 31237838]
225. Suppan L, Abbas M, Stuby L, Cottet P, Larribau R, Golay E, et al. Effect of an e-learning module on personal protective equipment proficiency among prehospital personnel: web-based randomized controlled trial. J Med Internet Res 2020 Aug 21;22(8):e21265 [FREE Full text] [doi: 10.2196/21265] [Medline: 32747329]
226. Suppan M, Gartner B, Golay E, Stuby L, White M, Cottet P, et al. Teaching adequate prehospital use of personal protective equipment during the COVID-19 pandemic: development of a gamified e-learning module. JMIR Serious Games 2020 Jun 12;8(2):e20173 [FREE Full text] [doi: 10.2196/20173] [Medline: 32516115]
227. Tadas S, Coyle D. Barriers to and facilitators of technology in cardiac rehabilitation and self-management: systematic qualitative grounded theory review. J Med Internet Res 2020 Nov 11;22(11):e18025 [FREE Full text] [doi: 10.2196/18025] [Medline: 33174847]
228. Tark R, Metelitsa M, Akkermann K, Saks K, Mikkel S, Haljas K. Usability, acceptability, feasibility, and effectiveness of a gamified mobile health intervention (Triumf) for pediatric patients: qualitative study. JMIR Serious Games 2019 Sep 30;7(3):e13776 [FREE Full text] [doi: 10.2196/13776] [Medline: 31573904]
229. Thirkettle M, Lewis J, Langdridge D, Pike G. A mobile app delivering a gamified battery of cognitive tests designed for repeated play (OU brainwave): app design and cohort study. JMIR Serious Games 2018 Oct 30;6(4):e10519 [FREE Full text] [doi: 10.2196/10519] [Medline: 30377140]
230. Thomas TH, Sivakumar V, Babichenko D, Grieve VL, Klem ML. Mapping behavioral health serious game interventions for adults with chronic illness: scoping review. JMIR Serious Games 2020 Jul 30;8(3):e18687 [FREE Full text] [doi: 10.2196/18687] [Medline: 32729836]
231. Tong T, Chignell M, Tierney MC, Lee J. A serious game for clinical assessment of cognitive status: validation study. JMIR Serious Games 2016 May 27;4(1):e7 [FREE Full text] [doi: 10.2196/games.5006] [Medline: 27234145]
232. van de Weijer SC, Kuijf ML, de Vries NM, Bloem BR, Duits AA. Do-It-Yourself gamified cognitive training: viewpoint. JMIR Serious Games 2019 May 07;7(2):e12130 [FREE Full text] [doi: 10.2196/12130] [Medline: 31066713]
233. van Mierlo T, Hyatt D, Ching AT, Fournier R, Dembo RS. Behavioral economics, wearable devices, and cooperative games: results from a population-based intervention to increase physical activity. JMIR Serious Games 2016 Jan 28;4(1):e1 [FREE Full text] [doi: 10.2196/games.5358] [Medline: 26821955]
234. Verschueren S, Buffel C, Stichele G. Developing theory-driven, evidence-based serious games for health: framework based on research community insights. JMIR Serious Games 2019 May 02;7(2):e11565 [FREE Full text] [doi: 10.2196/11565] [Medline: 31045496]
235. Verschueren S, van Aalst J, Bangels A, Toelen J, Allegaert K, Buffel C, et al. Development of CliniPup, a serious game aimed at reducing perioperative anxiety and pain in children: mixed methods study. JMIR Serious Games 2019 Jun 01;7(2):e12429 [FREE Full text] [doi: 10.2196/12429] [Medline: 31199333]
236. Vugts MA, Joosen MC, Mert A, Zedlitz AM, Vrijhoef HJ. Effectiveness of serious gaming during the multidisciplinary rehabilitation of patients with complex chronic pain or fatigue: natural quasi-experiment. J Med Internet Res 2018 Aug 15;20(8):e250 [FREE Full text] [doi: 10.2196/jmir.9739] [Medline: 30111527]
237. White B, Giglia RC, White JA, Dhaliwal S, Burns SK, Scott JA. Gamifying breastfeeding for fathers: process evaluation of the milk man mobile app. JMIR Pediatr Parent 2019 Jun 20;2(1):e12157 [FREE Full text] [doi: 10.2196/12157] [Medline: 31518324]
238. Wiecek E, Torres-Robles A, Cutler RL, Benrimoj SI, Garcia-Cardenas V. Impact of a multicomponent digital therapeutic mobile app on medication adherence in patients with chronic conditions: retrospective analysis. J Med Internet Res 2020 Aug 12;22(8):e17834 [FREE Full text] [doi: 10.2196/17834] [Medline: 32784183]
239. Wilson H, Stoyanov SR, Gandabhai S, Baldwin A. The quality and accuracy of mobile apps to prevent driving after drinking alcohol. JMIR Mhealth Uhealth 2016 Aug 08;4(3):e98 [FREE Full text] [doi: 10.2196/mhealth.5961] [Medline: 27502956]
240. Wunsch K, Eckert T, Fiedler J, Cleven L, Niermann C, Reiterer H, et al. Effects of a collective family-based mobile health intervention called "Smartfamily" on promoting physical activity and healthy eating: protocol for a randomized controlled trial. JMIR Res Protoc 2020 Nov 11;9(11):e20534 [FREE Full text] [doi: 10.2196/20534] [Medline: 33174849]
241. Zhang M, Heng S, Song G, Fung DS, Smith HE. Co-designing a mobile gamified attention bias modification intervention for substance use disorders: participatory research study. JMIR Mhealth Uhealth 2019 Oct 03;7(10):e15871 [FREE Full text] [doi: 10.2196/15871] [Medline: 31584003]
242. Zhang M, Ying J, Ho RC. Gamified cognitive bias interventions for psychiatric disorders: protocol of a systematic review. JMIR Res Protoc 2018 Oct 16;7(10):e10154 [FREE Full text] [doi: 10.2196/10154] [Medline: 30327292]
243. Zhang M, Ying J, Song G, Fung D, Smith H. Attention and cognitive bias modification apps: review of the literature and of commercially available apps. JMIR Mhealth Uhealth 2018 May 24;6(5):e10034 [FREE Full text] [doi: 10.2196/10034] [Medline: 29793899]

244. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. PLoS Med 2010 Feb;7(2):e1000217 [FREE Full text] [doi: 10.1371/journal.pmed.1000217] [Medline: 20169112]
245. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference. 2011 Presented at: 15th International Academic MindTrek Conference; September 28 - 30, 2011; Tampere Finland p. 9-15. [doi: 10.1145/2181037.2181040]
246. Deterding S, Sicart M, Nacke L, O'Hara K, Dixon D. Gamification. using game-design elements in non-gaming contexts. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. 2011 Presented at: CHI '11: CHI Conference on Human Factors in Computing Systems; May 7-12, 2011; Vancouver BC Canada p. 2425-2428. [doi: 10.1145/1979742.1979575]
247. Huotari K, Hamari J. Defining gamification: a service marketing perspective. In: Proceedings of the 16th International Academic MindTrek Conference. 2012 Presented at: 16th International Academic MindTrek Conference; October 3 - 5, 2012; Tampere Finland p. 17-22. [doi: 10.1145/2393132.2393137]
248. King D, Greaves F, Exeter C, Darzi A. 'Gamification': influencing health behaviours with games. J R Soc Med 2013 Mar;106(3):76-78. [doi: 10.1177/0141076813480996] [Medline: 23481424]
249. Deterding S, Khaled R, Nacke L, Dixon D. Gamification: Toward a definition. CHI 2011 Gamification Workshop Proceedings. 2011. URL: https://www.researchgate.net/publication/273947177_Gamification_Toward_a_definition [accessed 2021-07-29]
250. Zichermann G, Cunningham C. Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps. Sebastopol, California, United States: O′Reilly Media, Inc; 2011:1-208.
251. Miller AS, Cafazzo JA, Seto E. A game plan: Gamification design principles in mHealth applications for chronic disease management. Health Informatics J 2014 Jul 1;22(2):184-193. [doi: 10.1177/1460458214537511] [Medline: 24986104]
252. Stieglitz S, Lattemann C, Robra-Bissantz S, Zarnekow R, Brockmann T. Gamification. Cham, Switzerland: Springer; 2017:1-164.
253. Abt C. Serious Games. Lanham, Maryland, United States: University Press of America; 1987.
254. Zyda M. From visual simulation to virtual reality to games. Computer 2005 Sep;38(9):25-32. [doi: 10.1109/mc.2005.297]
255. Alvarez J, Djaouti D. An introduction to serious game definitions and concepts. Serious Games Simul Risks Manage 2011;11(1):11-15 [FREE Full text]
256. Bergeron B. Developing Serious Games (Game Development Series). Massachusetts, United States: Charles River Media, Inc; 2006:1-480.
257. Marsh T. Serious games continuum: Between games for purpose and experiential environments for purpose. Entertain Comput 2011 Jan;2(2):61-68. [doi: 10.1016/j.entcom.2010.12.004]
258. Hunicke R, LeBlanc M, Zubek R. MDA: A formal approach to game design and game research. In: Proceedings of the Game Design and Tuning Workshop at the Game Developers Conference. 2004 Presented at: Game Design and Tuning Workshop at the Game Developers Conference; 2004; San Jose, CA p. 1-5 URL: https://aaai.org/Papers/Workshops/2004/WS-04-04/WS04-04-001.pdf
259. Marczweski A. What's the difference between Gamification and Serious Games? InformaTech. 2013. URL: https://www.gamasutra.com/blogs/AndrzejMarczewski/20130311/188218/Whats_the_difference_between_Gamification_and_Serious_Games.php [accessed 2021-04-20]
260. McCallum S. Gamification and serious games for personalized health. Stud Health Technol Inform 2012;177:85-96. [Medline: 22942036]
261. Santhanam R, Liu D, Shen WM. Research Note—Gamification of Technology-Mediated Training: Not all competitions are the same. Info Syst Res 2016 Jun;27(2):453-465. [doi: 10.1287/isre.2016.0630]
262. Schmidt-Kraepelin M, Thiebes S, Stepanovic S, Mettler T, Sunyaev A. Gamification in health behavior change support systems - a synthesis of unintended side effects. In: Proceedings of the 14th International Conference on Wirtschaftsinformatik. 2019 Presented at: 14th International Conference on Wirtschaftsinformatik; February 24-27, 2019; Siegen, Germany p. 1032-1046 URL: https://serval.unil.ch/resource/serval:BIB_4B8751E9C432.P001/REF
263. Hibberd F. What is scientific definition? J Mind Behav 2019;40(1):29-52 [FREE Full text]
264. Hartmann NB, Hüffer T, Thompson RC, Hassellöv M, Verschoor A, Daugaard AE, et al. Are we speaking the same language? Recommendations for a definition and categorization framework for plastic debris. Environ Sci Technol 2019 Feb 05;53(3):1039-1047. [doi: 10.1021/acs.est.8b05297] [Medline: 30608663]
265. Kari T, Piippo J, Frank L, Makkonen M, Moilanen P. To gamify or not to gamify? Gamification in exercise applications and its role in impacting exercise motivation. In: Proceedings of the 29th Bled eConference - Digital Economy. 2016 Presented at: 29th Bled eConference - Digital Economy; June 19-22, 2016; Bled, Slovenia URL: https://www.researchgate.net/publication/304899543_To_Gamify_or_Not_to_Gamify_Gamification_in_Exercise_Applications_and_Its_Role_in_Impacting_Exercise_Motivation
266. Nacke LE, Deterding S. The maturing of gamification research. Comput Hum Behav 2017 Jun;71:450-454. [doi: 10.1016/j.chb.2016.11.062]
267. Harter JK, Schmidt FL. Conceptual versus empirical distinctions among constructs: implications for discriminant validity. Ind Organ Psychol 2015 Jan 07;1(1):36-39. [doi: 10.1111/j.1754-9434.2007.00004.x]

268. Koivisto J, Hamari J. Demographic differences in perceived benefits from gamification. Comput Hum Behav 2014 Jun;35:179-188. [doi: 10.1016/j.chb.2014.03.007]
269. Schmidt-Kraepelin M, Thiebes S, Baumsteiger D, Sunyaev A. State of play: A citation network analysis of healthcare gamification studies. In: Proceedings of the 26th European Conference on Information Systems. 2018 Presented at: 26th European Conference on Information Systems; June 23-28, 2018; Portsmouth, UK. [doi: 10.5445/IR/1000092503]

Abbreviations

GAMING: Game-Based Intervention Reporting Guidelines
RQ: research question

Edited by G Eysenbach; submitted 12.05.21; peer-reviewed by S Goria, R Nuijten; comments to author 27.05.21; revised version received 11.06.21; accepted 17.06.21; published 10.09.21.

Please cite as:
Warsinsky S, Schmidt-Kraepelin M, Rank S, Thiebes S, Sunyaev A
Conceptual Ambiguity Surrounding Gamification and Serious Games in Health Care: Literature Review and Development of Game-Based Intervention Reporting Guidelines (GAMING)
J Med Internet Res 2021;23(9):e30390
URL: https://www.jmir.org/2021/9/e30390
doi:10.2196/30390
PMID:34505840

Original Paper

A Computer-Based Interactive Narrative and a Serious Game for Children With Asthma: Development and Content Validity Analysis

Made Ary Sarasmita[1,2*], MClinPharm; Luh Putu Febryana Larasanty[2], MSc; Li-Na Kuo[1,3], MSc; Kuei-Ju Cheng[1,3], PharmD; Hsiang-Yin Chen[1*], PharmD

[1]Department of Clinical Pharmacy, School of Pharmacy, Taipei Medical University, Taipei, Taiwan

[2]Program Study of Pharmacy, Faculty of Mathematics and Science, Udayana University, Badung, Indonesia

[3]Department of Pharmacy, Taipei Medical University, Wan Fang Hospital, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Hsiang-Yin Chen, PharmD
Department of Clinical Pharmacy
School of Pharmacy
Taipei Medical University
Health and Science Building, 7th Floor
250 Wuxing Street
Taipei, 110
Taiwan
Phone: 886 02 2736 1661 ext 6175
Fax: 886 02 2736 1661
Email: shawn@tmu.edu.tw

Abstract

Background: Nonadherence to medications, failure to prevent exposure to asthma triggers, lack of knowledge about using medications, and fixed mindsets contribute to poor asthma control in children. Digital learning could provide a new strategy for improving health-related outcomes in children with asthma.

Objective: The aim of this study is to develop and design a digital educational program, titled Module of Inhaler and Asthma Triggers for Children (MIRACLE), for Indonesian children with asthma. The program comprises an interactive narrative and a serious game. It was proposed to increase the understanding of asthma self-management, instruct on proper inhaler techniques, improve asthma control, and promote a growth mindset for children with asthma.

Methods: Two phases of research were conducted to develop the program. In the first phase, a literature search and two rounds of the Delphi technique were conducted to obtain agreement from an expert panel regarding elements of asthma self-management and the design of interactive narratives and a serious game. The expert panel item statements were evaluated using the content validity index (CVI). In the second phase, the SERES framework, Norma Engaging Multimedia Design, and Psychological Theory of Growth Mindset were applied to create a storyline, learn objectives, and game challenges.

Results: In the first phase, 40 experts were invited to participate in Delphi round 1. Forty responses were collected to generate 38 item statements that consisted of part 1, elements of asthma self-management (25 items), and part 2, design of an interactive narrative and a serious game (13 items); 38 experts were involved in Delphi round 2. In total, 24 statements in part 1 and 13 items in part 2 had item-CVI values >0.80. The average CVI was 0.9, which was considered acceptable. Four narrative plots and five game sessions were developed during the second phase. Challenges with the scenario, scoring, and feedback on asthma difficulties were designed to promote a growth mindset for learners.

Conclusions: We developed a culture-specific, computer-based asthma program containing an interactive narrative and a serious game to deliver asthma self-management and promote a growth mindset among Indonesian children.

(J Med Internet Res 2021;23(9):e28796) doi:10.2196/28796

KEYWORDS

asthma; computer; children; interactive narrative; serious game; digital education; mobile phone

Introduction

Background

Designing effective asthma education for children is a challenge for health care professionals, educators, and parents. Insufficient knowledge of asthma and nonadherence to medications in children are associated with high rates of morbidity and mortality [1]. Improved education and support to help pediatric patients manage their disease are urgently recommended to control the burden of asthma [2]. However, frequent occurrences of acute asthma attacks and negative outcomes imply that effective self-management of asthma is difficult to achieve in pediatric patients [3,4]. Children frequently have low engagement with asthma self-care and poor medication adherence [5]. Efforts to develop new asthma self-management education for children must consider their level of understanding and the development of a certain mindset.

Intertwining behavioral stages and a growth mindset in asthma education may help children overcome the challenges of disease exacerbation and engage with learning. Deploying principles from modern behavioral psychology, children with a growth mindset might improve their self-care abilities by identifying impediments and embracing challenges [6]. Children face greater difficulties than older patients in comprehending causal relationships between asthma triggers and their consequences [7]. Previous negative experiences of asthma exacerbations may escalate their fears and discourage them from engaging in asthma self-management activities. An animated web-based interactive brain function program [8] was built to develop a growth mindset and resiliency through interactive games and classroom activities. Children with asthma may be unable to manage their disease because of a lack of concepts [9], and asthma education with a growth mindset might empower children and their families to exercise greater control over their situation. Incorporating a similar approach to promote positive attitudes and behavioral changes may provide new insights into asthma education for children.

Game-based education that incorporates elements of child psychology may encourage children with asthma to acquire knowledge, skills, self-care engagement, and motivation. Serious games were reported to promote positive attitudes and acceptance among children with HIV or AIDS [10], cancer [11], and asthma [12]. A serious game is defined as an entertaining game provided through a digital platform with an enjoyable design, specific rules including challenge goals, and a scoring concept that provides feedback to the player [13]. The well-recognized SERES framework, developed by Verschueren et al [14], provides rigorous scientific and design foundations to inform the design of serious games for health-related outcomes. Using game-based education, children experience opportunities to solve problems through analytical thinking [15] and improve specific skills through training and adequate feedback [16]. Engagement is also essential for programs designed to promote healthy behaviors. Norma Engaging Multimedia Design (NEMD) provides a theory for constructing engaging materials and encouraging thoughts and feelings, and it has been used to develop interactive narratives for asthma education [17].

Appropriate cultural adaptation is an essential element that should be emphasized when creating asthma education for children [18]. A systematic review confirmed that educators should adopt culture-specific values, beliefs, and languages for educational materials to represent real-life role models and social support systems of target populations [19]. Different languages and cultures may act as barriers and decrease the effectiveness of educational programs [20,21]. Indonesia, an archipelagic country with diverse ethnicities, languages, and local beliefs, may require different strategies when delivering asthma education. Culture-specific asthma education proved superior to a generic program in terms of reducing exacerbations requiring hospitalization in children [18]. The development of a serious game related to asthma that adopts Indonesian cultural and language preferences may help fill this gap and strengthen self-management among children with asthma.

Objective

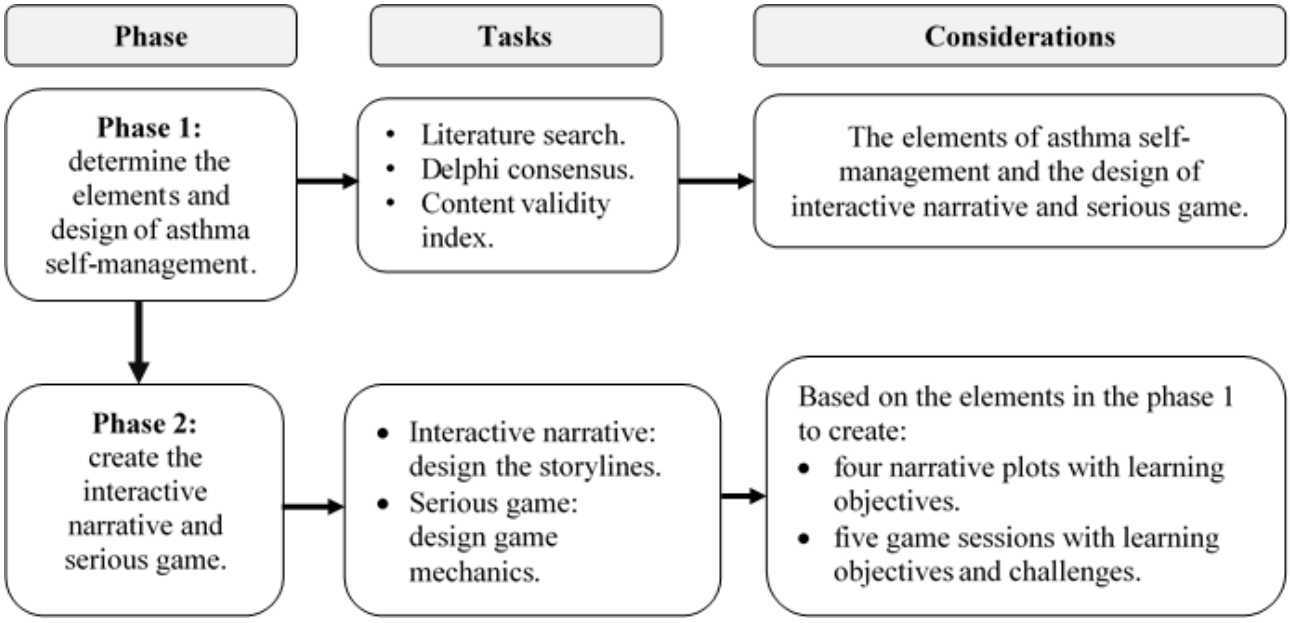
To address the main issue of delivering effective asthma education to children, we developed a digital asthma educational program titled Module of Inhaler and Asthma Triggers for Children (MIRACLE). It consisted of an interactive narrative, a serious game, and a written asthma action plan. The specific aims of this study were as follows: (1) to conduct the Delphi technique to obtain culture-specific agreements from a variety of relevant stakeholders on the elements and design of asthma education and (2) to design an interactive narrative and a serious game based on the NEMD theory and the SERES framework.

Methods

Study Design

There were two phases in this study, as shown in Figure 1. Phase 1 focused on determining elements of asthma self-management and designing an interactive narrative and serious game using previous literature, a Delphi consensus among experts, and calculation of the content validity index (CVI); phase 2 focused on designing an interactive narrative and a serious game using the NEMD theory and SERES framework. Ethical approval for this study was obtained from the Ethics Committee of the Faculty of Medicine, Udayana University, Indonesia (no. 2020.03.1.0865). Health care professionals who agreed to participate in this study received information on the study on the web and provided informed consent before the Delphi consensus was initiated.

Figure 1. Flowchart of the research phase.



Phase 1: Determining the Elements and Design of Asthma Self-management

Delphi Expert Panel

The expert panel consisted of physicians and clinical pharmacists invited to develop the elements and design an interactive narrative and serious game [22]. We attempted to invite diverse categories of health care providers from academia, hospital care, and primary care to accurately capture the content required for asthma information in various clinical settings in Indonesia. Physicians who served on the expert panel specialized in pediatrics and had practiced pediatric ambulatory care for patients with asthma for at least 1 year. Clinical pharmacists had at least a master's degree and had practiced in pediatric care or pediatric research for at least 2 years. After providing written informed consent, panelists were invited to participate in the panel every week via reminders sent by WhatsApp Messenger (Facebook) with a link to a Google Forms survey (Google). The researchers conducted a web-based consensus among health care providers to identify all categories of expected needs for Indonesian children with asthma via a two-round modified Delphi study. Participants answered questions and provided initial judgments until agreement was obtained using an anonymous, iterative technique to avoid the influence of different perspectives [23]. Individual responses were neither excluded nor shared with the other panelists. The Delphi study was conducted from May to July 2020.

In the first round of the Delphi process, researchers sent an open question to the experts using Google Forms: "According to your experience, what are the components of an interactive narrative and serious game that children can understand about asthma?" All suggestions collected in the first round were used to generate structural item statements. These statements were then categorized into two parts, including elements of asthma self-management (part 1) and the design of an interactive narrative and serious game (part 2). Generated statements were constructed with a four-point Likert scale (1=very irrelevant, 2=irrelevant, 3=relevant, and 4=very relevant) and sent to the experts again for Delphi round 2.

Content Validation

After the second round of Delphi, we used the CVI to quantify the content validity of the item statements. It relies on expert ratings for each item based on the relevance of asthma self-management and the design of the interactive narrative and serious game. Item-specific CVI (I-CVI) was also calculated for each item statement by counting the number of experts who rated the item as 3 (relevant) or 4 (very relevant) and dividing that number by the total number of experts [24]. The overall instrument-CVI was defined *as* "the average proportion of items rated as 3 or 4 (valid) across the experts," as the average CVI. It was calculated by summing the I-CVIs and dividing them by the number of items. We used a CVI cutoff point of at least 80% (a minimum I-CVI of 0.80) as acceptable [25] and an average CVI of ≥0.90 to have excellent content validity [26]. The experts were also prompted to comment on and justify their responses.

Phase 2: Creating an Interactive Narrative and Serious Game

General Design

After agreement was achieved, the elements and design identified in the two Delphi rounds were used to create an interactive narrative and serious game with the assistance of a multimedia expert. A serious game was created to support children's growth mindset after learning about asthma self-management through an interactive narrative. The main considerations included designing the storylines, characters, and learning objectives in the interactive narrative, and designing the game mechanics, learning objectives, and challenges in the serious game. Cartoon characters were created to reflect the cultural preferences of Indonesian children.

Target Audience

The appropriate age range for the target audience to receive this program was children aged 6-12 years based on their cognitive development and need for asthma education [27]. Researchers invited 1 pediatrician and 2 clinical pharmacists to provide feedback on the design of the interactive narrative and serious game. The proposed sequences of using this computer-based asthma educational intervention were that learners would engage with the interactive narrative and then play a serious game.

Design of the Interactive Narrative

Scientific Foundation

The design of the interactive narrative applied the NEMD theory and transformed the information we sought to convey into storylines with learning objectives. It was visualized using Articulate Storyline, version 2.0 (Articulate Global). This approach allows a designer to create a nonlinear storyline for which branches of the story differ, but each plot shares the same trunk of characters and learning objectives of asthma self-management. It allows the target audience to select actions and cartoon characters. The storyline adopted the NEMD theory, including simulation interactivity, construct interactivity, and immediacy.

Design Foundations

Sketches of cartoon characters were illustrated by a pharmacist using the Sketchbook Draw and Paint app, downloaded from the Google Android Play Store (Autodesk), and the sketches were imported into Articulate Storyline. The voices of pharmacists delivering asthma information and music were recorded and embedded in the software. The software generated HTML5-compatible output, which could be accessed on a PC.

Design of the Serious Game

Scientific Foundation

The scientific foundations of the SERES framework were applied to develop the serious game. The target audience, the theoretical basis from the literature, and the CVI were assessed using the iterative Delphi method. The learning objectives of the serious game were to achieve active learning of asthma self-management and increase confidence and the growth mindset of children with asthma.

Design Foundations

The design foundations included learning mechanics, game mechanics, and design requirements. Learning mechanics were defined according to the learning objectives [28], and these were changed into game mechanics, which incorporated the concept of a growth mindset. The concept of a growth mindset was incorporated into game sessions by providing challenges and rewarding efforts. Design requirements considered the platform and display appropriate to the children's mindset and their level of understanding. The user interface of the serious game was developed using the Construct 2D game programming software, version 2.0 (Scirra) and run on a Windows desktop platform. Several characteristics of using Construct 2.0, in this study, allowed researchers to publish the game on desktop computers (PC or Mac [Apple Inc]), Android mobile platforms, and websites via HTML5, relying on its event system and requiring no programming language or coding experience. The technical details of Construct 2.0 included a game design template, game objects, event system such as execution of commands, fun factors such as challenges and rewards, and the game genre. Game flow explained how the game was played from the beginning until a player either won or lost.

Results

Phase 1: Determining the Elements of Asthma Self-management

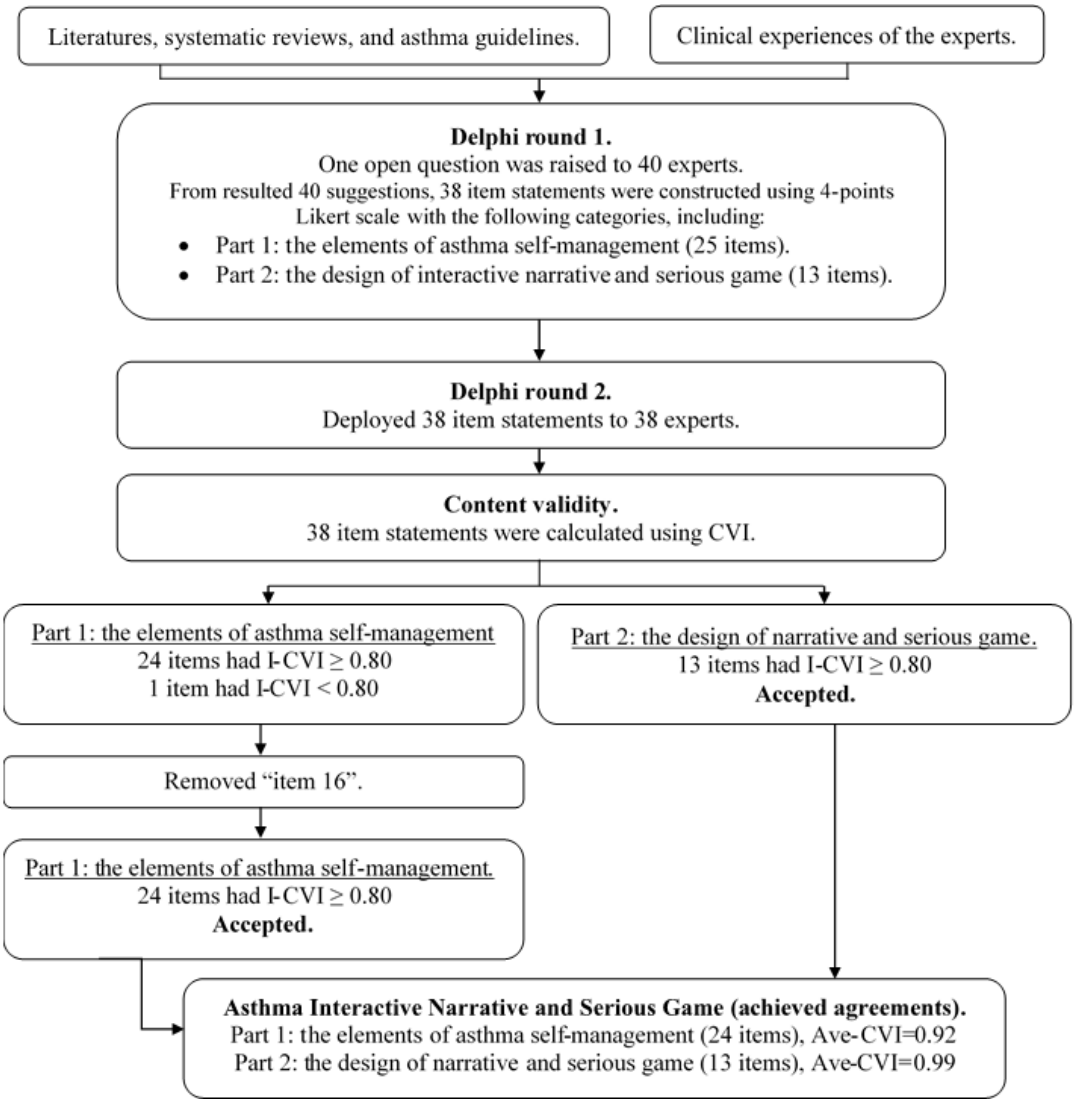
Of the 50 experts invited with various areas of expertise, including pediatricians and clinical pharmacists, 40 experts agreed to participate in Delphi round 1. The demographic characteristics of the 40 participants are given in Table 1. A total of 40 responses after raising an open question were collected in round 1 (Table S1 in Multimedia Appendix 1).

Table 1. Demographic characteristics of the Delphi expert panels.

Characteristic	Round 1 (N=40)	Round 2 (n=38)
Age (years), mean (SD)	33.4 (3.4)	33.3 (3.3)
Sex, n (%)		
Male	12 (31)	11 (29)
Female	28 (69)	27 (71)
Health care professional, n (%)		
Physician	15 (38)	14 (38)
Pharmacist	25 (63)	24 (63)
Specialty, n (%)		
Physician		
Pediatrics	2 (5)	1 (3)
Family medicine	7 (18)	7 (18)
Other	6 (15)	6 (16)
Pharmacist		
Master of Clinical Pharmacy	25 (63)	24 (63)
Practice field, n (%)		
Hospital	16 (40)	16 (42)
University	20 (50)	18 (47)
Primary care	4 (10)	4 (11)

Figure 2 demonstrates the Delphi process, including the experts involved in two Delphi rounds to obtain agreement on the elements of asthma self-management and design of the interactive narrative and serious game. Forty suggestions identified in round 1 were summarized into 38 item statements using a four-point Likert scale, including part 1, which consisted of elements of asthma self-management (25 items), and part 2, which consisted of the design (13 items). Two experts declined to participate in Delphi round 2; thus, the 38-item statements were deployed to 38 experts.

Figure 2. The Delphi process and content validity of the interactive narrative and serious game. Ave-CVI: average content validity index; CVI: content validity index; I-CVI: item content validity index.



Content validity was assessed after the 38 responses were collected, and consisted of the elements of asthma self-management and design of the interactive narrative and serious game (Table S2 in Multimedia Appendix 1). In total, 24 statements in part 1 and 13 item statements in part 2 had I-CVI values of ≥0.8. The experts considered one statement (item 16): (1) "explain the types of inhalers available in Indonesia" to be *not very relevant* or *not relevant*, as this statement had an I-CVI value of 0.76 (<0.80). This item had a low I-CVI value because the experts explained that there were many types of inhalers in Indonesia, which might contribute to confusing the children. According to the panel's suggestion, this item was removed, and 24 acceptable items on the elements of asthma self-management and 13 acceptable items on the design of the interactive narrative and serious game were retained. The agreement achieved among the experts resulted in average CVI values of 0.92 (for the 24 items of asthma self-management) and 0.99 (for the 13 items of the design), and achieved a universal average CVI of 0.5 (Table S3 in Multimedia Appendix 1). As item 16 was removed, we used a metered-dose inhaler with a spacer in the narrative and game sessions to introduce general information on inhaler use for children.

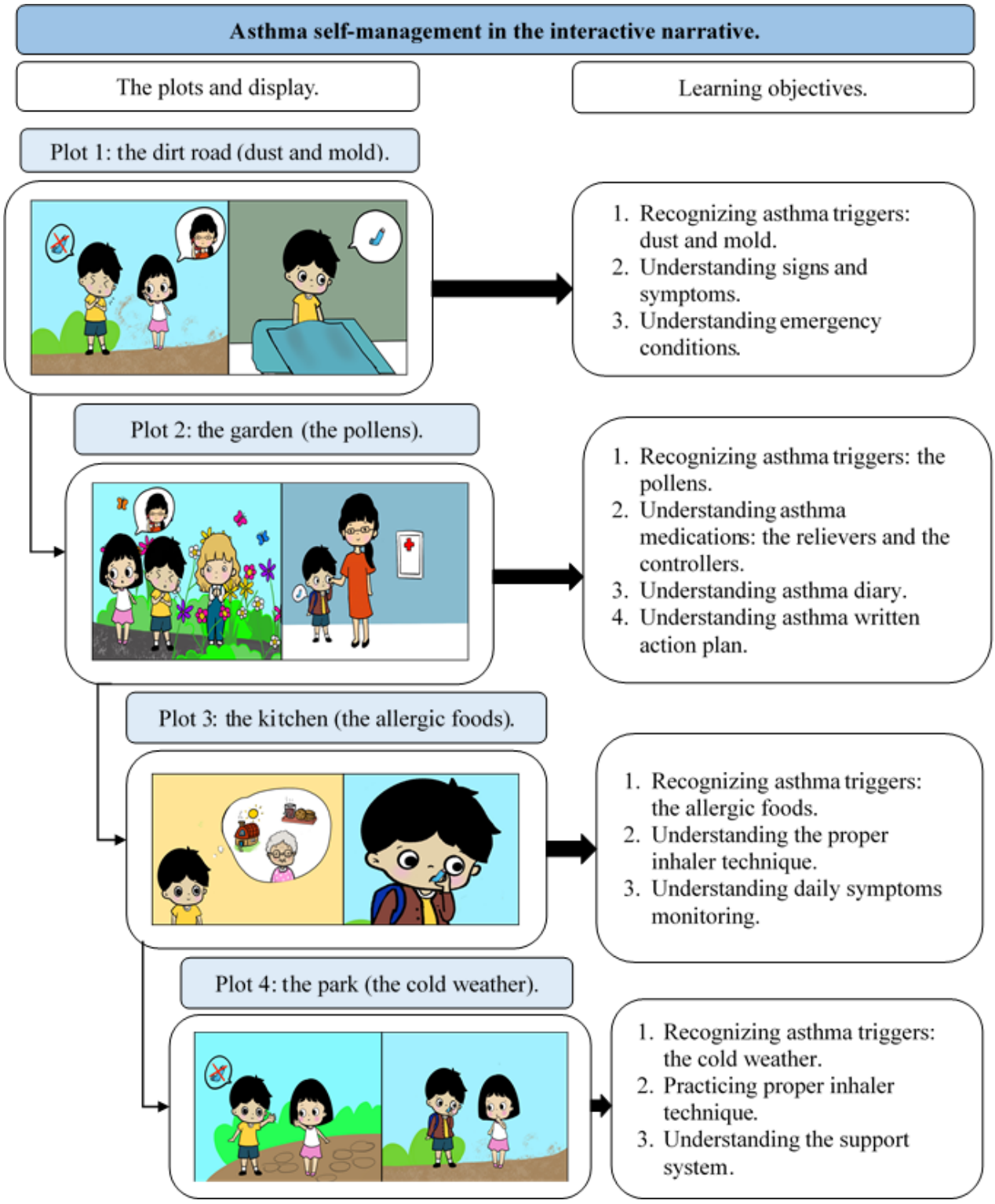
Phase 2: Create the Design of the Interactive Narrative and Serious Game

Design of the Interactive Narrative

Figure 3 illustrates the plots and learning outcomes of the four storylines. The target audience has to deal with asthma triggers and write an asthma action plan during their cartoon adventure, as structured in the storylines (Figure S3 in Multimedia Appendix 2). Asthma information, purposes, characters, and instructions are provided before the learners engage with the narrative. This combination of strategies contains elements of asthma self-management and supplements each other (Table S1 in Multimedia Appendix 3). Plots were created to represent asthma triggers, including a dirt road (dust and mold), a garden (pollen), a kitchen (food allergies), and a park (cold weather). Plot 1 was created to provide information on asthma triggers encountered in daily activities, asthma signs and symptoms, and emergency conditions and to help children become acquainted with their previous asthma experiences. Plot 2 was concerned with the importance of asthma medications,

medication adherence, and a written asthma action plan. Plot 3 was designed to deliver information on proper inhaler techniques and the importance of daily symptom observation. Plot 4 emphasized practicing proper inhaler use and strategies to obtain support from their immediate environment. Each plot contained a different ending to the story and achieved the represented goal by completing the plots without an asthma attack. The learners were asked to repeat the narrative if they failed to achieve the goal. Positive feedback from 3 experts was collected to supplement the design of the narrative (Table S2 in Multimedia Appendix 3).

Figure 3. The storyline and the learning objectives of the interactive narrative.



The visualization and expression of the characters were adjusted for each plot. On the basis of the NEMD theory, construct interactivity describes the availability of activities in the virtual world, as visualized as a short trip to the grandmother's house. It represents the unique culture of Indonesian children visiting their grandparents in a village during a religious holiday. The

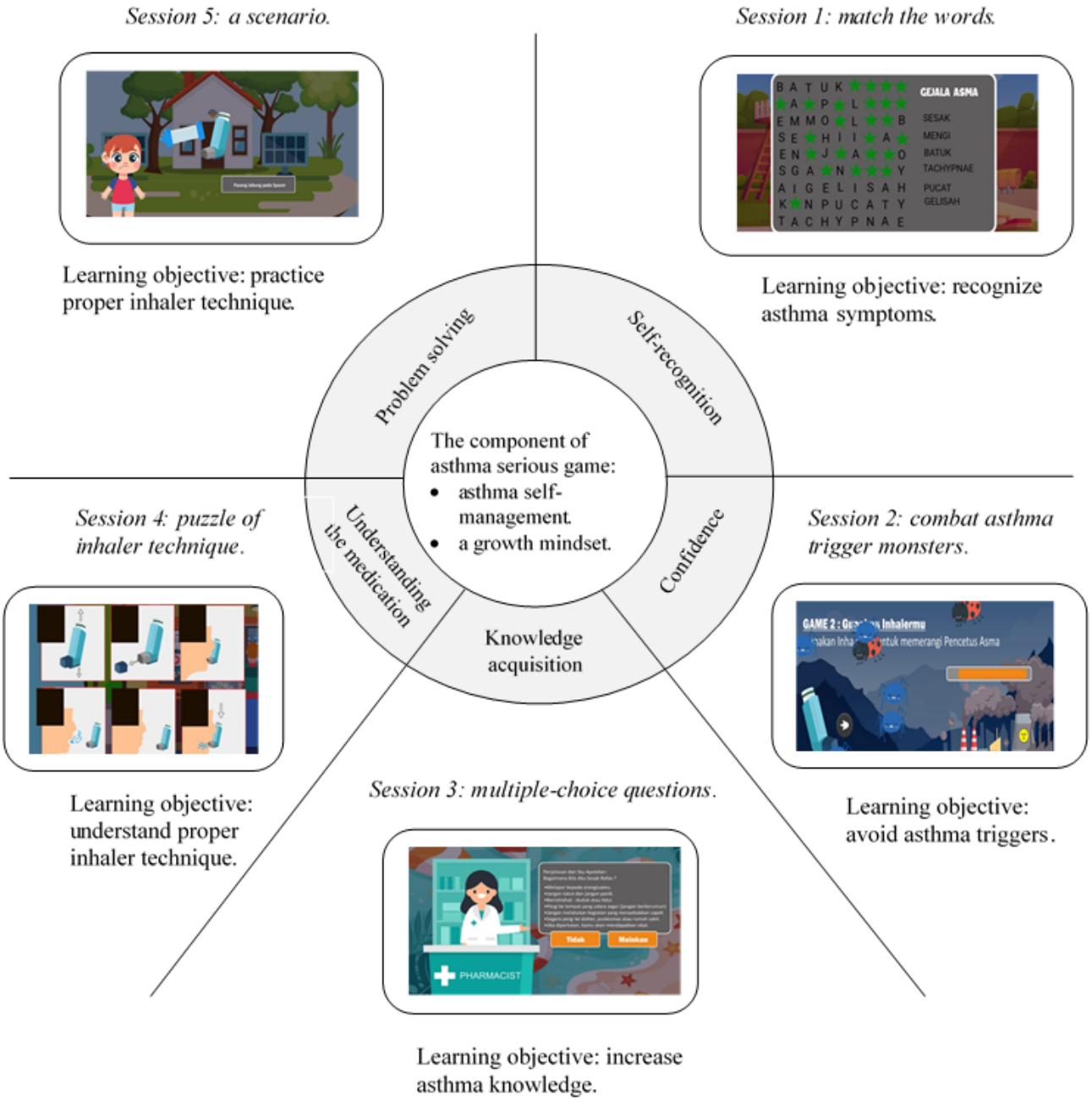
cartoon characters were designed as children and families with positive attitudes, and the cartoon family was visualized as gentle and caring to cover simulated interactivity. Learners are able to choose one cartoon character with asthma and engage with that character. Immediacy facilitates learners' choice and observation of all actions. In each plot, every choice matters, and learners can decide their own options and compromise with all of the resulting choices (Figure S3 in Multimedia Appendix 2). Music and daily conversations among cartoon characters were added to provide fun and active learning.

Design of the Serious Game

Scientific Foundations

Figure 4 demonstrates a visualization of the serious game with learning objectives, elements of asthma self-management, and the component of a growth mindset. The learning objectives were designed for each session of a serious game. Each session represents the preferences of children as the target audience, such as a cartoon protagonist, a colorful background, and musical tones. Session 1 covers recognizing asthma signs, symptoms, and emergency conditions. Session 2 encourages players to combat asthma trigger monsters. Session 3 was designed to recall knowledge by choosing the correct answer from multiple-choice questions. Session 4 introduces proper inhaler techniques, including how to correctly use, clean, and store the inhaler. Session 5 focuses on practicing steps of correctly using a metered-dose inhaler using a scenario.

Figure 4. Game sessions, growth mindset, and the learning objectives.



Design Foundations

Learning mechanics were integrated with the concept of a growth mindset. Before beginning the game, players are able to choose a character to represent themselves in game sessions. Learning mechanics include avoiding skipping content, embedding information through images, and providing feedback. A cartoon pharmacist provides brief information about asthma with an inserted link of images before entering each session to avoid skipping content. The development of a growth mindset is facilitated by ticking asthma difficulties on a checklist and achieving goals in daily life before entering the game sessions. Session 1 was designed to cope with asthma and increase self-recognition. Session 2 aimed to increase self-confidence and reduce fear of exposure to asthma triggers. Sessions 3 and 4 were constructed to improve the understanding of the disease and medications and to achieve goals. Session 5 provides a scenario to help a child use an inhaler with a spacer. After playing sessions 1 and 2, players were asked to write feedback and explain their experiences when experiencing an asthma attack.

The design requirements were transformed into components of the growth mindset, as shown in Figure 4. The visualization and screenshots were designed as a 2D game that facilitated the ability to complete the challenges (Figures S4-S6 in Multimedia Appendix 2). This game presents *lock* badges to reinforce the difficulties and reward efforts required to finish one session before playing the next session. Visual callouts appear to increase excitement in each featured session. The serious game also presents points and stars collected by the players on a score screen after a session is completed. This serious game can be applied directly (offline) on a computer. It is also available on the web as an app that can be downloaded for free and played on any Android PC, smartphone, or tablet (Figure S2 in Multimedia Appendix 2) [29].

The game design template is described in Table 2, and the game flow is illustrated in Figure 5. To facilitate problem solving for learners with a fixed mindset, game mechanics were explicitly applied in session 5. In session 5, players apply an inhaler with a spacer, which presents the learning mechanics of *selecting an inhaler with a spacer*, *properly applying the inhaler*, *understanding its use*, *and repeating the sequences*.

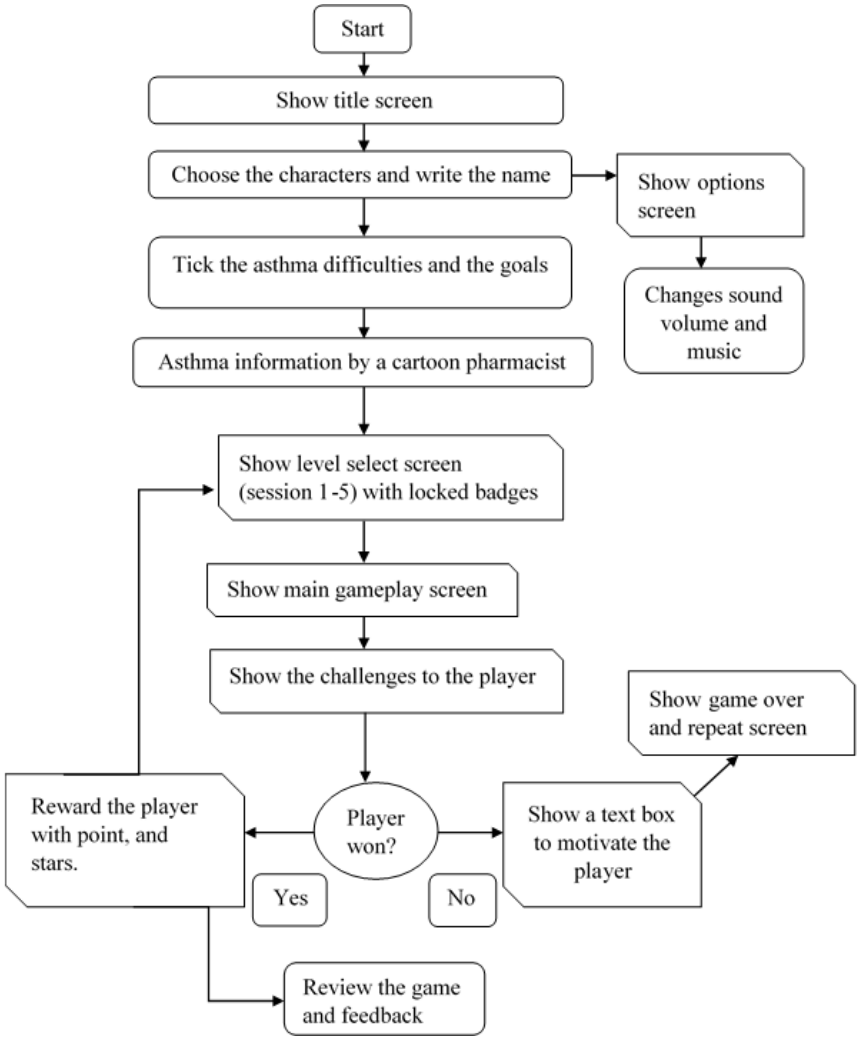
Table 2. Game design template.

Component	Description
Main purpose	• Improve asthma self-management and encourage children to act as a manager of their disease
Genre	• Casual game with puzzles, questions, and battle
Visual	• Cartoon 2D game
Game mechanics	
Session 1	• Matching words (making a line)
Session 2	• Battle (shooting asthma trigger monsters)
Session 3	• Questions and answers (select an option)
Session 4	• Puzzle of inhaler (dragging, picking up puzzle pieces, and placing them on the puzzle board)
Session 5	• A scenario (dragging an inhaler with a spacer)
Game behavior	
Session 1	• Drag and drop, anchor
Session 2	• Fade, bullet, destroy outside, eight directions
Session 3	• Drag and drop
Session 4	• Drag and drop
Session 5	• Drag and drop, pin
Project scope	
Number of characters	• The player's character and a cartoon pharmacist
Number of sessions	• Five sessions
Type of enemies	• Asthma trigger monsters (game session 2)
Type of weapons	• Inhaler (game session 2)
Number of puzzles	• Three puzzles (game session 4)
Rewards	• Point and stars every time a player finishes a session
Growth mindset	• Checklist of asthma difficulties • Checklist of goals in daily life • Writing about asthma experiences and feelings • A box with motivational text in each session
Winning condition	• Finish each session and unlock the next session
Mastering the program	• HTML5, Google Play Store
Image and audio files	• Image format: 32-bit PNG[a] file • Audio files: 16-bit PCM[b] wav files

[a]PNG: Portable Network Graphics.

[b]PCM: pulse code modulation.

Figure 5. Game flow.



Discussion

Principal Findings

This study describes the development and design of an Indonesian culture-specific, computer-based asthma education module for children, which comprises an interactive narrative and a serious game. The Delphi consensus was used to gather reliable and relevant perspectives and opinions systematically. Two rounds of a Delphi study and content validity led to low panel attrition and were considered appropriate to achieve agreement [30]. From Delphi rounds 1 and 2, experts emphasized the elements of asthma self-management and the design that should be included in building storylines and game challenges. The panel achieved agreement on 24 items of asthma self-management and 13 items on the design of preferences. The average CVI was acceptable for creating the interactive narrative and serious game. The scientific foundations and expert participatory action employed in this program were designed to overcome the limitations of existing asthma educational tools for children.

According to the 2020 Global Initiative for Asthma guidelines [31], asthma education needs to be delivered at every visit, which confirms that any educational asthma module for children should emphasize the importance of self-management ability. The development of the ability to manage asthma in daily life is highly recommended to enable children to successfully manage their disease. The guideline recommendations for asthma self-management are as follows: (1) identify asthma triggers that may predispose one to an asthma attack, (2) regularly determine an asthma action plan, (3) recognize asthma symptoms, (4) avoid asthma triggers to prevent asthma attacks, and (5) use proper inhaler techniques [27]. Children grow from being dependent to independent adolescents who must manage their asthma on their own; thus, knowledge, growth mindset, and skills of practicing correct inhaler techniques are crucial to asthma educational programs. The same five recommendations listed above were also highlighted during the initial phase of the Delphi technique and were used to develop the story lines and challenges.

The NEMD theory and SERES framework were applied during the second development phase to design the foundations of the module and visualize the interactive narrative and serious game. The interactive narrative presents lessons that are suitable for school-aged children, and the computer-based asthma program relies on the narrative to provide meaning to virtual engagement. Unlike a conventional paper-based story, this study supports active learning in the context of manipulating a plot based on a learner's decisions. It meets a learner's needs and expresses thoughts based on their personal asthma experiences. Four plots

coherently contained elements of asthma self-management and promoted positive attitudes. Although there was no punishment or reward after learning, achieving these goals is important in daily life.

The SERES framework was applied in several game development studies, which showed that it was effective in reducing learning barriers, such as perioperative anxiety and pain among children [32]. A gamified e-learning module also used the SERES framework to teach the adequate use of personal protective equipment during the COVID-19 pandemic [33]. This study proposes design foundations that encourage children to customize a cartoon character with positive attitudes, as visualized in the 2D journey to finish five sessions with a locked badge. It represents self-representation and is important for increasing their feelings and sense of relating to and engaging with the system. Each visualization in the plots was adjusted to the preferences of children and provided options to deal with each consequence. Take-home messages are offered to encourage children to adhere to asthma therapy and create a written asthma action plan. Our findings also highlighted the importance of achieving goals and providing feedback at the end of each game session and after each narrative.

The unique learning goals and design provided by the interactive narrative and serious game supplement each other, offering exciting opportunities to educate children in solving their daily problems. Several previous games [34,35] taught self-management skills through the recognition of asthma triggers and symptoms and the appropriate use of medications. In this study, the MIRACLE program explicitly covered comprehensive lessons on asthma education, including the pathophysiology, recognition of asthma triggers and symptoms, inhalation techniques, adherence to medications, and compliance. Children usually lack a long attention span and may have a low capacity to perceive terminology, limited coping, and limited social support [36]. A comic-based asthma educational program confirmed that children are highly interested in fun, comprehensive learning games, and comics accompanied by their parents [37]. Narratives in comic visualization also convey illness experiences and reduce feelings of fear [38]. A comprehensive learning tool with an enjoyable design increases user satisfaction, improves perception and acceptance, leads to acquisition of knowledge and skills, and promotes behavioral changes to improve health-related outcomes [39].

This study embedded the theory of cultivating a growth mindset in a computer-based interactive narrative and serious game. Compared with previous games that only included a simple adventure genre [40,41], this serious game encompasses a variety of fun gaming genres, such as adventure, quizzes, puzzles, and a scenario to promote a growth mindset. Asthmatic children face lifelong challenges, and a growth mindset can help them overcome difficulties of the disease and accomplish tasks [42]. Point rewards and badges as part of gamification are essential for attracting children's attention [43]. A previous study that applied a computer-based neonatal resuscitation training game for nursing students consisted of scenarios with an increasingly difficult neonatal resuscitation activity in each round [44]; it was designed to promote a growth mindset in terms of attitudes and performance. Provision of feedback and challenges in developing a growth mindset are meaningful and relevant to users. The combination of a growth mindset and active learning can enhance children's self-management skills and the courage to live with asthma.

Generalizability of the MIRACLE program should be a concern because of cultural and language-specific elements and design. Its applicability should be tested in other settings and cultural contexts before direct translation. As language barriers [45] and a fixed mindset seem to contribute to poorer health outcomes and knowledge in children, this study highlighted the importance of developing games using the target users' own language and promoting a growth mindset by presenting challenges and rewarding efforts. A culturally adapted translation and validation of the elements are needed to address this issue. The concepts and methodology, techniques for obtaining the elements, storylines, and game mechanics can potentially be applied to other clinical research in developing digital-based health education for children in primary school with chronic diseases.

Limitations

This study did not involve participation by children, and thus, actual children's feedback on their preferences is lacking. External validity remained unclear because no reliability testing was performed. Despite the fact that the participation rate of medical doctors in the expert panel was 50%, a limited number of pediatricians agreed to participate. However, this study consistently leveraged a wide range of perspectives on asthma self-management. The involvement of primary care doctors and pharmacists is important because of their crucial role in providing information on proper inhaler techniques to patients in clinical and community settings [46]. This computer-based narrative game was developed using an Indonesian cultural context, which in turn influenced the decisions of target audiences in learning asthma self-management and adherence to treatments. Cultural and religious beliefs, health literacy, and life experiences have become limitations in directly translating the MIRACLE program in other countries with different cultural preferences. The Delphi method and content validity may be warranted to adjust the cultural preferences of target audiences to add or delete a few elements of asthma education. Practicing inhaler techniques in this program solely through virtual learning may be inadequate. Supplementing periodic real-time demonstrations and practice by educators may be required to overcome this limitation.

Further Research

Reliability testing and randomized controlled trials are urgently needed to evaluate the effectiveness of this program and measure the magnitude and retention of changes in health-related outcomes among Indonesian children. The evaluation of satisfaction with the game should be provided after completing all game sessions. Feedback from parents and children could also be acquired after the program was delivered. We propose to evaluate asthma control and quality of life in children using a translated and validated version of the Asthma Control Questionnaire and Pediatric Asthma Quality of Life Questionnaires in Bahasa Indonesia with permission from the original author [47]. Correct inhaler use should be measured

using a checklist adapted from the Indonesian Standard of Community Pharmacy Practice [48]. Data before and after the intervention should be measured and evaluated to provide a preliminary overview of the effectiveness of this program among children with asthma. The feedback generated can be used to improve this computer-based program to better suit the needs of children with asthma and their parents.

Contributions of the Study

This study provides an innovative method for delivering effective asthma education in clinical practice. This may help reduce communication barriers when delivering asthma self-management and promoting medication adherence. Given the cost and resources, this program was proposed to enhance asthma self-management skills through an interactive and growth mindset–enhanced learning environment. Developing a growth mindset may contribute to increased self-confidence and problem solving to face asthma difficulties in daily life among children, thus leading them to be successful managers of their disease in the future.

Conclusions

A computer-based asthma self-management program composed of an interactive narrative and a serious game was developed in the Indonesian language to overcome the adversities of asthma education in children. The participatory approach, scientific framework, and culture-specific elements led to the successful development of an interactive narrative and serious game at the level of understanding and mindset of children.

Acknowledgments

The authors would like to give special thanks to the members of the panel of experts who made the content development of the MIRACLE program possible. They also thank pharmacist Kalih Sindhu Budari, who illustrated the cartoons and the storylines; Dr Rama Sutawan, MD, who provided advice on editing the visualization of the narrative; and Mr Fatchu Auliya, who constructed the game in appropriate software to accomplish the objectives. The authors would also like to thank Taipei Medical University (Taipei, Taiwan) and Udayana University (Bali, Indonesia) for supporting this research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Suggestions from the experts and content validity index.
[DOCX File , 27 KB - jmir_v23i9e28796_app1.docx]

Multimedia Appendix 2
Structure of the storylines and real screenshots.
[DOCX File , 11990 KB - jmir_v23i9e28796_app2.docx]

Multimedia Appendix 3
Elements of asthma self-management.
[DOCX File , 22 KB - jmir_v23i9e28796_app3.docx]

References

1. Knox BL, Luyet FM, Esernio-Jenssen D. Medical neglect as a contributor to poorly controlled asthma in childhood. J Child Adol Trauma 2019 Dec 10;13(3):327-334. [doi: 10.1007/s40653-019-00290-0] [Medline: 33088390]
2. Widayati A, Virginia DM, Setiawan CH, Fenty F, Donowati MW, Christasani PD, et al. Pharmacists' views on the development of asthma pharmaceutical care model in Indonesia: a needs analysis study. Res Social Adm Pharm 2018 Dec;14(12):1172-1179. [doi: 10.1016/j.sapharm.2018.01.008] [Medline: 29395903]
3. Farzandipour M, Nabovati E, Sharif R, Arani MH, Anvari S. Patient self-management of asthma using mobile health applications: a systematic review of the functionalities and effects. Appl Clin Inform 2017 Dec;8(4):1068-1081 [FREE Full text] [doi: 10.4338/ACI-2017-07-R-0116] [Medline: 29241254]
4. Bellin MH, Newsome A, Lewis-Land C, Kub J, Mudd SS, Margolis R, et al. Improving care of inner-city children with poorly controlled asthma: what mothers want you to know. J Pediat Health Care 2018 Jul;32(4):387-398. [doi: 10.1016/j.pedhc.2017.12.009] [Medline: 29540280]
5. McPherson AC. A randomized, controlled trial of an interactive educational computer package for children with asthma. Pediatrics 2006 Apr 01;117(4):1046-1054. [doi: 10.1542/peds.2005-0666] [Medline: 16585298]
6. Dweck CS, Yeager DS. Mindsets: a view from two eras. Perspect Psychol Sci 2019 Feb 01;14(3):481-496. [doi: 10.1177/1745691618804166] [Medline: 30707853]
7. Pradel FG, Hartzema AG, Bush PJ. Asthma self-management: the perspective of children. Patient Edu Counsel 2001 Dec;45(3):199-209. [doi: 10.1016/s0738-3991(01)00123-9] [Medline: 11722856]

8. Donohoe C, Topping K, Hannah E. The impact of an online intervention (Brainology) on the mindset and resiliency of secondary school pupils: a preliminary mixed methods study. Edu Psychol 2012 Apr 03;32(5):641-655 [FREE Full text] [doi: 10.1080/01443410.2012.675646]
9. Burkhart PV, Rayens MK. Self-concept and health locus of control: factors related to children's adherence to recommended asthma regimen. Pediatr Nurs 2005;31(5):404-409. [Medline: 16295156]
10. Whiteley L, Mena L, Craker LK, Healy MG, Brown LK. Creating a theoretically grounded gaming app to increase adherence to pre-exposure prophylaxis: lessons from the development of the viral combat mobile phone game. JMIR Serious Games 2019 Mar 27;7(1):e11861. [doi: 10.2196/11861] [Medline: 30916652]
11. Kato PM, Cole SW, Bradlyn AS, Pollock BH. A video game improves behavioral outcomes in adolescents and young adults with cancer: a randomized trial. Pediatrics 2008 Aug;122(2):305-317. [doi: 10.1542/peds.2007-3134] [Medline: 18676516]
12. Howell KJ. “Quest for the Code”: A study of a computer-based education program for children with asthma. Thesis and Dissertations - Syracuse University. 2005. URL: https://www.proquest.com/docview/305383018 [accessed 2020-10-12]
13. Bergeron BP. Stud Health Technol Inform 2008;132:26-30. [Medline: 18391250]
14. Verschueren S, Buffel C, Stichele G. Developing theory-driven, evidence-based serious games for health: framework based on research community insights. JMIR Serious Games 2019 May 02;7(2):e11565. [doi: 10.2196/11565] [Medline: 31045496]
15. Gerwin RL, Kaliebe K, Daigle M. The interplay between digital media use and development. Child Adolesc Psychiatr Clin N Am 2018 Apr;27(2):345-355. [doi: 10.1016/j.chc.2017.11.002] [Medline: 29502755]
16. Charlier N, Zupancic N, Fieuws S, Denhaerynck K, Zaman B, Moons P. Serious games for improving knowledge and self-management in young people with chronic conditions: a systematic review and meta-analysis. J Am Med Inform Assoc 2016 Jan;23(1):230-239. [doi: 10.1093/jamia/ocv100] [Medline: 26186934]
17. Wyatt TH, Li X, Huang Y, Farmer R, Reed D, Burkhart PV. Developing an interactive story for children with asthma. Nurs Clin North Am 2013 Jun;48(2):271-285. [doi: 10.1016/j.cnur.2013.01.006] [Medline: 23659813]
18. Mitchell SJ, Bilderback AL, Okelo SO. Racial disparities in asthma morbidity among pediatric patients seeking asthma specialist care. Acad Pediat 2016 Jan;16(1):64-67. [doi: 10.1016/j.acap.2015.06.010] [Medline: 26329015]
19. McCallum GB, Morris PS, Brown N, Chang AB. Culture-specific programs for children and adults from minority groups who have asthma. Cochrane Data Syst Rev 2017 Aug 22;8(8):CD006580. [doi: 10.1007/bf03085652] [Medline: 28828760]
20. Al Sayah F, Majumdar SR, Williams B, Robertson S, Johnson JA. Health literacy and health outcomes in diabetes: a systematic review. J Gen Intern Med 2012 Oct 13;28(3):444-452. [doi: 10.1007/s11606-012-2241-z] [Medline: 23065575]
21. Murray L, Elmer S, Elkhair J. Perceived barriers to managing medications and solutions to barriers suggested by bhutanese former refugees and service providers. J Transcult Nurs 2018 Apr 05;29(6):570-577. [doi: 10.1177/1043659618768702] [Medline: 29618276]
22. McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. Int J Clin Pharm 2016 Feb 5:655-662. [doi: 10.1007/s11096-016-0257-x] [Medline: 26846316]
23. Hsu C, Sandford B. The Delphi Technique: Making sense of consensus. Pract Assess Res Eval. 2007. URL: https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1177&context=pare [accessed 2021-08-24]
24. Abrar EA, Yusuf S, Sjattar EL, Rachmawaty R. Development and evaluation educational videos of diabetic foot care in traditional languages to enhance knowledge of patients diagnosed with diabetes and risk for diabetic foot ulcers. Primary Care Diabetes 2020 Apr;14(2):104-110. [doi: 10.1016/j.pcd.2019.06.005] [Medline: 31311727]
25. Almanasreh E, Moles R, Chen TF. Evaluation of methods used for estimating content validity. Res Soc Administ Pharm 2019 Feb;15(2):214-221. [doi: 10.1016/j.sapharm.2018.03.066] [Medline: 29606610]
26. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health 2007 Aug;30(4):459-467. [doi: 10.1002/nur.20199] [Medline: 17654487]
27. Noenoeng RC, Kartasasmita C, Supriyatno B, Darmawan BS. Pedoman Nasional Asma Anak. Jakarta: Badan Penerbit Ikatan Dokter Anak Indonesia; 2017.
28. Arnab S, Lim T, Carvalho M, Bellotti F, de FS, Louchart S. Mapping learning and game mechanics for serious games analysis. Br J Edu Technol 2015;46:391-411 [FREE Full text] [doi: 10.1111/bjet.12113]
29. Miracle game edukasi asma. APKPure. URL: https://m.apkpure.com/miracle-game-edukasi-asma/com.madearysasmita.MiracleGameEdukasiAsma [accessed 2020-12-10]
30. Aronson BD, Janke KK, Traynor AP. Investigating Student Pharmacist Perceptions of Professional Engagement Using a Modified Delphi Process. AJPE 2012 Sep 10;76(7):125. [doi: 10.5688/ajpe767125] [Medline: 23049097]
31. Louis-Philippe BH, Levy M, Cruz A, Decker R. Global initiative for asthma: Asthma management and prevention for adults and children older than 5 years 2020. Global Initiative for Asthma. URL: https://ginasthma.org/ [accessed 2020-03-03]
32. Buffel C, van Aalst J, Bangels A, Toelen J, Allegaert K, Verschueren S, et al. A web-based serious game for health to reduce perioperative anxiety and pain in children (CliniPup): pilot randomized controlled trial. JMIR Serious Games 2019 Jun 1;7(2):e12431. [doi: 10.2196/12431] [Medline: 31199324]
33. Suppan M, Gartner B, Golay E, Stuby L, White M, Cottet P, et al. Teaching adequate prehospital use of personal protective equipment during the COVID-19 pandemic: development of a gamified e-learning module. JMIR Serious Games 2020 Jun 12;8(2):e20173. [doi: 10.2196/20173] [Medline: 32516115]

34. Huss K, Winkelstein M, Nanda J, Naumann P, Sloand E, Huss R. Computer game for inner-city children does not improve asthma outcomes. J Pediat Health Care 2003 Mar;17(2):72-78. [doi: 10.1016/s0891-5245(02)88328-6] [Medline: 12665729]
35. Kaufman D, Sauvé L, Renaud L. Enhancing learning through an online secondary school educational game. J Edu Comput Res 2011 Aug 10;44(4):409-428 [FREE Full text] [doi: 10.2190/ec.44.4.c]
36. Trollvik A, Ringsberg KC, Silén C. Children's experiences of a participation approach to asthma education. J Clin Nurs 2013 Jan 12;22(7-8):996-1004. [doi: 10.1111/jocn.12069] [Medline: 23311685]
37. Mickel CF, Shanovich KK, Evans MD, Jackson DJ. Evaluation of a school-based asthma education protocol. J School Nurs 2016 Jul 22;33(3):189-197. [doi: 10.1177/1059840516659912] [Medline: 27450449]
38. Lee T, Sheu S, Chang H, Hung Y, Tseng L, Chou S, et al. Developing a web-based comic for newly diagnosed women with breast cancer: an action research approach. J Med Internet Res 2019 Feb 04;21(2):e10716. [doi: 10.2196/10716] [Medline: 30714947]
39. Drummond D, Monnier D, Tesnière A, Hadchouel A. A systematic review of serious games in asthma education. Pediatr Allergy Immunol 2017 Jan 27;28(3):257-265. [doi: 10.1111/pai.12690] [Medline: 27992659]
40. Yawn B, Algatt-Bergstrom P, Yawn R, Wollan P, Greco M, Gleason M. An in-school cd-rom asthma education program. J Sch Health 2000 Apr;70(4):153-159. [doi: 10.1111/j.1746-1561.2000.tb06462.x] [Medline: 10790839]
41. Shames RS, Sharek P, Mayer M, Robinson TN, Hoyte EG, Gonzalez-Hensley F, et al. Effectiveness of a multicomponent self-management program in at-risk, school-aged children with asthma. Ann Allergy Asthma Immunol 2004 Jun;92(6):611-618. [doi: 10.1016/s1081-1206(10)61426-3] [Medline: 15237762]
42. Haimovitz K, Dweck CS. The origins of children's growth and fixed mindsets: new research and a new proposal. Child Dev 2017 Sep 14;88(6):1849-1859. [doi: 10.1111/cdev.12955] [Medline: 28905371]
43. AlMarshedi A, Wills G, Ranchhod A. Guidelines for the gamification of self-management of chronic illnesses: multimethod study. JMIR Serious Games 2017 May 12;5(2):e12 [FREE Full text] [doi: 10.2196/games.7472] [Medline: 28500018]
44. Cutumisu M, Ghoman SK, Lu C, Patel SD, Garcia-Hidalgo C, Fray C, et al. Health care providers' performance, mindset, and attitudes toward a neonatal resuscitation computer-based simulator: empirical study. JMIR Serious Games 2020 Dec 21;8(4):e21855. [doi: 10.2196/21855] [Medline: 33346741]
45. Poureslami I, Nimmon L, Doyle-Waters M, Rootman I, Schulzer M, Kuramoto L, et al. Effectiveness of educational interventions on asthma self-management in Punjabi and Chinese asthma patients: a randomized controlled trial. J Asthma 2012 Jun 21;49(5):542-551. [doi: 10.3109/02770903.2012.682125] [Medline: 22715910]
46. Setiawan C, Widayati A, Virginia DM, Armour C, Saini B. The role of pharmacists in the pharmaceutical care of asthma patients in Yogyakarta, Indonesia: the patients' views. J Asthma 2019 Jun 17;57(9):1017-1028. [doi: 10.1080/02770903.2019.1622715] [Medline: 31204546]
47. Juniper EF, Gruffydd-Jones K, Ward S, Svensson K. Asthma Control Questionnaire in children: validation, measurement properties, interpretation. Eur Resp J 2010 Jun 07;36(6):1410-1416. [doi: 10.1183/09031936.00117509] [Medline: 20530041]
48. Indonesia KK. Petunjuk teknis standar pelayanan kefarmasian di apotek. Direktorat Jenderal Kefarmasian dan Alat Kesehatan (Farmalkes). Jakarta: Ministry of Health, the Republic of Indonesia; 2017. URL: https://farmalkes.kemkes.go.id/2020/01/petunjuk-teknis-standar-pelayanan-kefarmasian-di-apotek/ [accessed 2020-10-12]

Abbreviations

CVI: content validity index
I-CVI: item-specific content validity index
MIRACLE: Module of Inhaler and Asthma Triggers for Children
NEMD: Norma Engaging Multimedia Design

Edited by R Kukafka; submitted 17.03.21; peer-reviewed by R Sakulbumrungsil, A Teles; comments to author 24.04.21; revised version received 22.05.21; accepted 05.07.21; published 13.09.21.

Please cite as:
Sarasmita MA, Larasanty LPF, Kuo LN, Cheng KJ, Chen HY
A Computer-Based Interactive Narrative and a Serious Game for Children With Asthma: Development and Content Validity Analysis
J Med Internet Res 2021;23(9):e28796
URL: https://www.jmir.org/2021/9/e28796
doi:10.2196/28796
PMID:34515641

Original Paper

Acceptability Evaluation of the Use of Virtual Reality Games in Smoking-Prevention Education for High School Students: Prospective Observational Study

Jong-Long Guo[1], PhD; Hsiao-Pei Hsu[2], MSN; Tzu-Ming Lai[3], MEd; Mei-Ling Lin[2], MSN; Chih-Ming Chung[4], PhD; Chiu-Mieh Huang[5], PhD

[1]Department of Health Promotion and Health Education, College of Education, National Taiwan Normal University, Taipei, Taiwan

[2]Department of Nursing, College of Nursing, National Yang Ming Chiao Tung University, Taipei, Taiwan

[3]Wanfang High School, Taipei, Taiwan

[4]Department of Applied Information Technology, Hsing Wu University, New Taipei, Taiwan

[5]Institute of Clinical Nursing, College of Nursing, National Yang Ming Chiao Tung University, Taipei, Taiwan

Corresponding Author:
Chiu-Mieh Huang, PhD
Institute of Clinical Nursing
College of Nursing
National Yang Ming Chiao Tung University
No.155, Sec.2, Linong Street
Taipei, 112
Taiwan
Phone: 886 2 28267362
Fax: 886 2 28202487
Email: cmhuang2021@nycu.edu.tw

Abstract

Background: Alternative forms of cigarettes, such as electronic cigarettes (e-cigarettes), are becoming increasingly common among adolescents. Many high schools now provide smoking-prevention education in an attempt to minimize the potential negative health effects and illness burdens e-cigarettes may induce in adolescents. However, it is often difficult to motivate young students to engage with traditional education regarding the harmful effects of tobacco; thus, the development of alternative approaches may be required.

Objective: In this study, we aimed to conduct an acceptability evaluation of educational virtual reality games designed to support smoking-prevention measures. We based the acceptability evaluation on the following two experience types: game-playing and content-learning experiences. The paths by which these experience types affect the intention to abstain from smoking were also examined.

Methods: We applied a prospective observational study design. We developed educational games based on three-dimensional virtual reality technology, in which participants operated joysticks to complete challenge tasks. To increase the possibility of the games fostering motivation to abstain from smoking, the ARCS motivational model (comprising attention, relevance, confidence, and satisfaction) was used as a framework during the games' design. We measured the participants' game-playing experiences by inquiring about the strength of the ARCS elements; content-learning experiences were measured using overall knowledge improvement and the perceived persuasiveness of the content. A total of 130 students participated in the program. Study hypotheses for this evaluation were derived from a literature review. We used partial least squares structural equation modeling to examine the proposed hypotheses.

Results: Based on the responses of the students to questionnaire items concerning attention, relevance, confidence, and satisfaction in the context of the games, most students agreed or strongly agreed that the educational games were motivational, and that their game-playing experiences were positive. Regarding content-learning experiences, there was a significant improvement in knowledge (t_{129}=25.67, P<.001), and most students perceived themselves as being persuaded to abstain from smoking. Attention, relevance, and satisfaction significantly influenced perceived persuasiveness (t=3.19, P<.001; t=4.28, P<.001; and t=3.49, P<.001, respectively); however, confidence did not (t=0.42, P=.67). Perceived persuasiveness, relevance, and satisfaction significantly influenced the intention to abstain from smoking (t=3.57, P<.001). In addition to directly affecting the intention to abstain from

smoking, indirect effects were observed from both relevance and satisfaction to intention via perceived persuasiveness (t=2.87, P=.004 and t=2.11, P=.04, respectively). However, intention was not significantly influenced by knowledge improvement.

Conclusions: Our findings revealed that the educational games were positively accepted by the participating students. This indicates that the integration of the ARCS framework and persuasive strategies is applicable for smoking-prevention education. We recommend that the games be included as teaching materials for smoking-prevention education.

(J Med Internet Res 2021;23(9):e28037) doi:10.2196/28037

KEYWORDS

behavioral intention; ARCS motivation model; persuasiveness; smoking prevention; educational games

Introduction

Smoking causes approximately 7 million deaths worldwide annually and is commonly associated with noncommunicable diseases and disability [1]. Thus, smoking causes great harm to human health, and it can have particularly pronounced adverse effects among adolescents. Adolescence is a developmental stage, and engaging in smoking during this period can impair brain development [2]. In addition, adolescents can be prone to early nicotine dependence symptoms, which can increase the likelihood that they will be addicted to smoking in adulthood [3].

According to the World Health Organization's Framework Convention on Tobacco Control, it is important to promote and strengthen education and public awareness regarding the harms associated with tobacco [4]. In recent years, the prevalence of electronic cigarette (e-cigarette) use on campuses has rapidly increased; this is occurring despite the fact that there are many unanswered questions regarding the safety of e-cigarettes, such as whether they represent a less-harmful substitute for traditional cigarettes [5]. Further, their effectiveness as a means of facilitating smoking cessation and their overall impact on health, especially among adolescents, are also unclear [5]. Implementing e-cigarette–prevention measures in schools could help to minimize the potential negative health effects and illness burdens e-cigarettes may induce in adolescents. One of the most important school-based e-cigarette–prevention efforts is educating young students regarding the toxic exposure, potential risks, and health effects associated with e-cigarettes [6]. However, it is often difficult to motivate young students to engage with traditional education regarding the harms associated with tobacco; thus, the development of alternative approaches, such as educational games, may be more effective in this regard.

The use of games for health promotion and illness prevention is becoming increasingly common, and previous studies have revealed that gaming-based approaches can promote health behaviors among adolescents [7]. In the context of education, game-based learning (GBL) can be a practical and effective approach. Recently, immersive virtual reality (VR) has been introduced as a medium for learning and teaching, and has had a considerable influence in this regard [8]. Immersive VR allows users to perform actions in a virtual environment and provides opportunities for contextual practice [8]. Notably, conducting GBL through immersive VR enables situational learning, with learners using devices, such as helmets and joysticks, to observe and interact with virtual educational scenes [9,10]. Its characteristic of providing an interactive learning environment means this approach could increase learners' engagement with the content. This is notable because researchers have suggested that a lack of engagement may undermine learning effectiveness [11]. Further, studies have found that it is important to present engaging content as early as possible, and educational games quickly engage young students [12]. Enhancing engagement may increase the likelihood of favorable learning outcomes in which learners develop sufficient motivation to perform the target behavior [13]. This indicates that educational games may be an effective alternative option or supplementary material for smoking-prevention courses in schools.

In Taiwan, health education is a compulsory subject for high school students, and smoking prevention is one of the mandated learning contents stipulated in the national basic K-12 education curriculum [14]. The present researchers developed VR games to function as supplementary material for smoking-prevention education. School teachers may integrate these games into health education in a manner that suits their students and circumstances. Despite the potential benefits of educational games, the acceptability of educational VR games remains an underresearched area. Consequently, this study concerns an acceptability evaluation of educational VR games for smoking prevention, rather than an evaluation of the effectiveness of including VR games as part of a smoking-prevention program. Enhancing the understanding of the acceptability of VR games in this regard could help improve the design of such games and learners' engagement, thereby increasing the effectiveness of the education.

In this study, we based the acceptability evaluation on the following two experience types: game-playing and content-learning experiences. We derived measures of game-playing experiences from Keller's ARCS (attention, relevance, confidence, and satisfaction) motivation model [15]. To achieve effective learning, motivating learners is an important first step; however, scholars have suggested that motivation is unpredictable and changeable [16]. Thus, a systematic method of identifying motivation is essential for enhancing learning effectiveness. The ARCS motivation model provides a framework for understanding motivation. According to the ARCS model, learners who experience attention, relevance, confidence, and satisfaction through a learning process are likely to be motivated in regard to the educational content in question [17,18]. Therefore, in this study, we applied the ARCS model to assess students' game-playing experiences. We expected students to be motivated by playing the games, and sought to determine their game-playing experiences through ARCS assessment.

Content-learning experiences are defined as the knowledge obtained from educational material and the perceived persuasiveness of educational material. Behavioral knowledge is essential for behavior changes; without proper knowledge and guidance, it can be difficult to adopt a specific behavior [19]. However, knowledge alone is unlikely to generate behavior change [20], and a more integrated approach is needed. Thus, learners' perceptions of the persuasiveness of content are also important. The perceived persuasiveness of a system is an estimation of its ability to motivate behavior change [21].

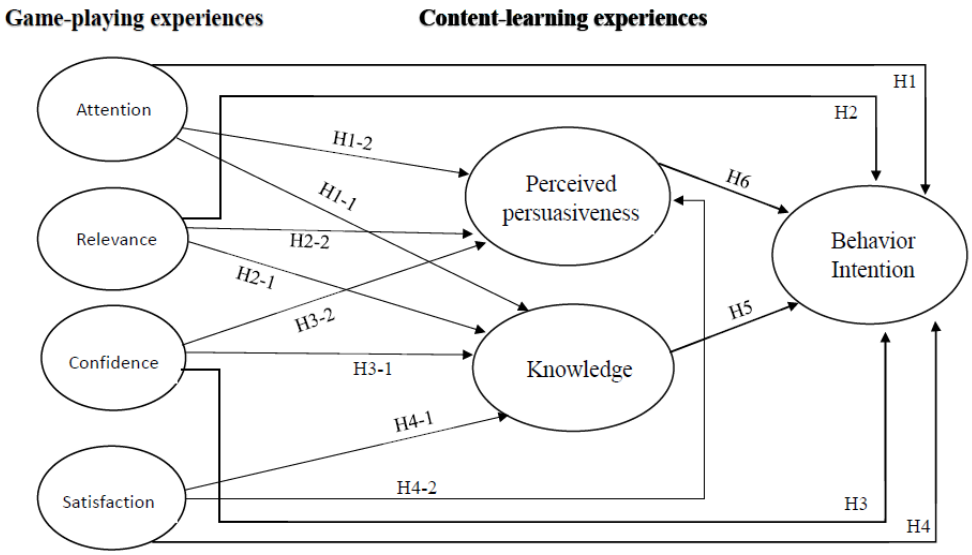
Persuasive system design includes behavior change strategies that support learners in attaining the target behavior. The following four features have been suggested for persuasive system design: task support (ie, self-monitoring), dialogue support (ie, positive reinforcement), credibility (ie, conveying trustworthiness), and social support (ie, provision of social information) [22]. Scholars have suggested several behavioral intervention strategies, such as providing information and instructions to increase learners' understanding and abilities to achieve a goal, anticipation of barriers, monitoring and feedback, social comparison, and rewards [23].

The educational VR games were designed to convey messages relating to smoking prevention. In this context, effective content learning should induce, among learners, enhanced knowledge of smoking prevention and a higher likelihood (through perceived persuasion) of adopting smoking-prevention behaviors.

Educational games for health are designed to motivate individuals to improve their health through implementing behavior changes. In this study, the target change was developing an intention to abstain from smoking. Behavioral intentions have a strong causal impact on behavior [24], and we expected that strengthening students' intentions to abstain from smoking would decrease their risk of adopting smoking behaviors in the future. Students' engagement with the games should produce both game-playing and content-learning experiences, and these two types of experiences should influence students' intentions to abstain from smoking. Based on this, we proposed the following hypotheses: hypothesis 1 (H1), attention influences behavioral intention; hypothesis 2 (H2), relevance influences behavioral intention; hypothesis 3 (H3), confidence influences behavioral intention; hypothesis 4 (H4), satisfaction influences behavioral intention; hypothesis 5 (H5), knowledge influences behavioral intention; and hypothesis 6 (H6), perceived persuasiveness influences behavioral intention.

Our educational VR games were designed to not only be fun, but also improve health. They represent digital persuasion agents that affect learners' engagement and, hopefully, motivate learners to change their behaviors. Notably, effective learning occurs through successful persuasion, which is a product of positive changes in cognition and attitude [25]. Previous studies have suggested that engagement with digital technology (in this study, "game-playing experiences") impacts the quality of users' experiences with the technology [26]. Effective game-playing experiences are necessary to enhance learners' engagement with the learning process. We propose that game-playing experiences cause learners to engage with learning content, which consequently creates a content-learning experience for the learners. Thus, we developed the following hypotheses regarding the influence of game-playing experiences on content-learning experiences: hypothesis 1-1 (H1-1), attention influences knowledge; hypothesis 1-2 (H1-2), attention influences perceived persuasiveness; hypothesis 2-1 (H2-1), relevance influences knowledge; hypothesis 2-2 (H2-2), relevance influences perceived persuasiveness; hypothesis 3-1 (H3-1), confidence influences knowledge; hypothesis 3-2 (H3-2), confidence influences perceived persuasiveness; hypothesis 4-1 (H4-1), satisfaction influences knowledge; and hypothesis 4-2 (H4-2), satisfaction influences perceived persuasiveness. All the hypotheses of the proposed model are summarized in Figure 1.

Figure 1. The hypotheses of the proposed model.



Methods

Participants and Recruitment

This study featured a prospective observational study design. The participating students (1) were high school students, (2) did not have any major chronic diseases (eg, asthma, heart disease, and diabetes), (3) were not pregnant, (4) were willing to comply with verbal instructions, and (5) were able to provide written informed consent.

We conducted this study among students from a high school in Taipei City, Taiwan. The research team made an appointment with the school personnel who were responsible for implementing smoking-prevention measures on the school campus. During the meetings, we distributed study information and discussed with the school personnel potential barriers in regard to the implementation of the materials. Once a consensus was reached in regard to the manner by which the materials would be implemented, the school personnel announced the research project to students and distributed an information sheet. Students who were willing to participate were informed by the school personnel that the research sought to measure students' responses toward educational games. Participating students received a small gift as compensation for their contribution to smoking-prevention research.

Design of the Educational VR Games

We designed the educational VR games to function as supplementary material for smoking-prevention education. The theoretical concepts underlying the educational games' designs are depicted in Multimedia Appendix 1. We based the games' designs on the ARCS model of motivation (Multimedia Appendix 1); this model was developed by Dr John Keller [15], who designed the model to identify the important contributors to students' motivation. In addition, we presented the motivating concepts in the educational games in a manner that allowed them to correspond with specific persuasive strategies.

The educational VR games included a whack-a-mole game, a wire loop game, a square baseball game, and a Taiko drum game. Participating students wore a head-mounted display (HMD), which enabled them to have immersive experiences with 3D images. To complete the challenge tasks, the participants interacted with the virtual environments through the use of joysticks; they also received in-game guidance from an avatar.

Measurements

We used questionnaires to collect research data; these data included the students' background information, their perceived acceptability of the VR games, and their intention to abstain from smoking. The questionnaires were reviewed by four professionals, including two senior health-education teachers, one professor of health education, and one professor of nursing. We used the content validity index (CVI) to evaluate whether the items were appropriate for measuring the study concepts; only items with CVI values >0.80 were retained [27]. One item pertaining to confidence, "I found the games difficult," was deleted because of low factor loading (<0.6). This question was the only reverse-scored item in the questionnaire, which may have been the reason for the low factor loading.

Background Information and Intention to Abstain From Smoking

We collected background information for each participant, including their sex, age, whether they lived with both parents, their personal smoking experience, and their parents' and peers' smoking status. We performed the acceptability evaluation by measuring their game-playing and content-learning experiences. We expected that these experiences would strengthen the students' intention to abstain from smoking. Thus, the primary outcome of this study was students' intention to abstain from smoking, which was measured using two items; these were scored using a Likert-type scale ranging from 1 ("strongly disagree") to 5 ("strongly agree"), with higher scores indicating higher levels of intention. The study variables, corresponding items, and convergent validity values are presented in Table 1.

Table 1. Study variables, corresponding items, and convergent validity.

Experiences and variables	Corresponding items	Factor loading	CR[a]	AVE[b]
Game-playing experiences[c]				
Attention			0.93	0.77
Atten 1	The games' virtual reality scenes were eye-catching.	0.86		
Atten 2	The educational games provoked my curiosity.	0.87		
Atten 3	The interactional elements of the games held my attention.	0.90		
Atten 4	The educational games were fun to play.	0.88		
Relevance			0.87	0.62
Relev 1	The content presented in the games was relevant to my needs.	0.84		
Relev 2	The content presented in the games was relevant to my interests.	0.66		
Relev 3	The content presented in the games was useful to me.	0.76		
Relev 4	I was familiar with the content presented in the games.	0.87		
Confidence			0.80	0.57
Confi 1	I was confident I could complete the tasks in the games.	0.71		
Confi 2	My successes in the games were a result of my own efforts.	0.70		
Confi 3	I could effectively operate the joysticks to interact with the virtual reality scenes.	0.86		
Satisfaction			0.86	0.61
Satis 1	I could apply the content conveyed in the games in real life.	0.80		
Satis 2	I benefitted by playing the games.	0.76		
Satis 3	I felt a sense of accomplishment after finishing the tasks in the games.	0.82		
Satis 4	I was satisfied with my performance in the games.	0.73		
Content-learning experiences[d]				
Knowledge improvement	Fifteen true/false type questions related to smoking prevention were used to measure knowledge. Improvement was defined as postscore minus prescore.	1.00	1.00	1.00
Perceived persuasiveness			0.91	0.70
Persu 1	The games enhanced my awareness of smoking prevention.	0.82		
Persu 2	The games enhanced my concern regarding smoking prevention.	0.81		
Persu 3	The games enhanced my understanding of smoking-prevention methods.	0.86		
Persu 4	The games enhanced my self-efficacy in regard to practicing smoking-prevention skills.	0.86		
Intention to abstain from smoking			0.95	0.91
Inten 1	The content of the games has made me consider abstaining from smoking.	0.95		
Inten 2	In the future, I will abstain from smoking by using the content of the games as a reference.	0.96		

[a]CR: composite reliability. The CR criterion was >0.7 [26].

[b]AVE: average variance extracted. The AVE criterion was >0.5 [26].

[c]Motivation concepts were used to measure the game-playing experiences.

[d]Knowledge improvement and perceived persuasiveness were used to measure the content-learning experiences.

Acceptability Evaluation: Game-Playing Experiences

We designed the educational games to improve students' motivation. Therefore, we used motivation concepts (ARCS) to measure the game-playing experiences. Specifically, this study intended to examine the educational VR games' effectiveness as a teaching medium in terms of the games' ability to focus learners' attention, stimulate learning confidence, and effectively improve learning satisfaction. We based the measurement items on existing conceptualizations of the ARCS motivation model [13]. Attention was defined as follows: "Students' attention is directed to the stimuli, and their interest is maintained over time." Relevance was defined as follows: "Students perceive the contents as being related to their real-life experiences." Confidence was defined as follows: "Confident students attribute their successes in the games to their own ability and efforts." Satisfaction was defined as follows: "Students perceived the education as inducing internal feelings of satisfaction." We measured each aspect using four items, all of which were scored using a Likert-type scale ranging from 1 ("strongly disagree") to 5 ("strongly agree"), with higher scores indicating higher levels of motivation. Details of the ARCS measures are shown in Table 1.

Acceptability Evaluation: Content-Learning Experiences

In addition to motivating students through game-playing experiences, we expected that the educational VR games would, through content-learning experiences, influence the students' perspectives of smoking. The strength of the content-learning experiences was measured by examining knowledge improvement and perceived persuasiveness.

To measure knowledge improvement, we presented 15 true/false-type questions related to smoking prevention to the participants. Through administering pretests and posttests, we determined changes in the participants' scores between the pre-experiment and postexperiment periods.

Perceived persuasiveness was defined as students' awareness of and concern regarding smoking prevention, as well as their understanding of and self-efficacy regarding practicing preventive skills. We used four items to measure perceived persuasiveness (Table 1), with higher scores indicating higher levels of perceived persuasiveness.

Data Collection Procedure

This study received approval from the Research Ethics Review Committee of En Chu Kong Hospital (ECKIRB1090401). All of the participating students expressed their approval of the study and provided written informed consent. First, research staff gave the participants brief instructions regarding the educational VR games. Then, before playing, students completed a questionnaire that measured their knowledge of smoking prevention. The research staff then gave the students guidance regarding how to use the HMDs and joysticks, which enabled them to have immersive interactive experiences with 3D images. Students who agreed to participate in the research received the educational VR games to be played outside of class. The students were allowed to play the games freely; there was no fixed order regarding the games played. It takes about 35 to 45 minutes to complete the educational tasks. Health education is a compulsory course in Taiwan. Substance use prevention, including smoking prevention, is one of the required topics of the course. However, these students did not receive a similar course during the research. They learned about substance use prevention during a different semester. Research staff were present while the students completed the games, but did not provide any oral instructions. Once the students completed the educational games, the research staff provided them with a structured questionnaire that evaluated the acceptability of the games and their intention to abstain from smoking.

Data Analysis

We performed statistical analysis using SPSS 23.0 (IBM Corp). The participants' characteristics were described using numbers and percentages. Means and standard deviations were used to represent the variables associated with the games' acceptability. To examine the proposed hypotheses, we conducted partial least squares structural equation modeling (PLS-SEM) using the SmartPLS v3.0 program (SmartPLS GmbH); this approach was chosen because PLS-SEM is suitable for creating exploratory models and for research involving small sample sizes [28]. Using parameters with a significance level of .05, a statistical power of 80%, and an R^2 value of at least 0.25, the suggested sample size was 75 for the maximum number of six arrows pointing to a latent variable in the model [29]. Thus, a sample size of 130 in this study was sufficient.

PLS-SEM does not presume that the data are normally distributed. The process of bootstrapping involves repeated random sampling with replacement from the original sample to create a bootstrap sample. Using bootstrapping assumes that the sample distribution is a reasonable representation of the intended population distribution. The minimum number of bootstrap samples is 5000 [28]. The bootstrap sample enables the estimated path coefficients in PLS-SEM to be tested for their significance. Significant paths showing the hypothesized direction empirically supported the proposed relationship.

We conducted PLS-SEM in the following two steps: (1) assessment of the reliability and validity of the measurement model, which was used to test the relationships between each latent variable and its indicators; and (2) assessment of the structural model, through which estimates were provided for the path coefficients, which represented the hypothesized relationships among the latent variables in the proposed model [30,31].

Measurement Model

We examined the convergent and discriminant validity in order to verify that the latent variables (the game-playing and content-learning experiences and intention to abstain from smoking) were valid and reliable.

Convergent validity can be determined by considering composite reliability (CR) and average variance extracted (AVE) values. To meet the convergent validity and reliability requirements for the model, CR should be greater than 0.7 [32] and AVE should be greater than 0.5 [28]. We used CR to evaluate the construct measures' internal consistency reliability. Using CR to measure internal consistency reliability meant that the PLS-SEM could accommodate different indicator reliabilities

and that the underestimation associated with Cronbach α was avoided [28]. Meanwhile, AVE indicates the latent variables' accounted variance from the study measures, and higher AVE values indicate higher accounted variance.

To establish discriminant validity, we examined the Fornell-Larcker criterion [28] and the heterotrait-monotrait (HTMT) [32]. To satisfy the Fornell-Larcker criterion, the square root of the AVE for each variable should exceed the correlation of the latent variables [28]. Meanwhile, for the HTMT, values of <1.00 indicate discriminant validity [32].

Structural Model

PLS-SEM indicates explained variances for the latent variables through R^2 (coefficient of determination), which describes levels of predictive accuracy. Estimates (β coefficient) were provided for the path coefficients to indicate the strength of the relationship between the latent variables [31].

To test the significance of the relationship between the latent variables, we estimated *t* values and reported their corresponding *P* values [31]. We used the standardized root mean square residual (SRMR) to test the model's fit (<0.08 indicates a good fit [32]).

Results

Participants' Characteristics

The participants (N=130) comprised 57 (43.8%) male and 73 (56.2%) female students. The mean age of the sample was 16.64 years, and most participants (113/130, 86.9%) were under 17 years of age. Over half (102/130, 78.5%) lived with both parents. Approximately 20.8% (27/130) of the students had experience of smoking; this included any tobacco product or e-cigarette. Over 30% had family members (45/130, 34.6%) and close friends (41/130, 31.5%) who were smokers (Table 2).

Table 2. Participants' characteristics (N=130).

Characteristic	Value
Sex, n (%)	
Male	57 (43.8)
Female	73 (56.2)
Age (years), mean (SD)	16.64 (0.73)
Age group (years), n (%)	
≤17	113 (86.9)
≥18	17 (13.1)
Household status, n (%)	
Live with both parents	102 (78.5)
Other	28 (21.5)
Have you ever smoked? (any tobacco product or e-cigarette), n (%)	
Yes	27 (20.8)
No	103 (79.2)
Do any of your family members smoke?, n (%)	
Yes	45 (34.6)
No	85 (65.4)
Do any of your close friends smoke?, n (%)	
Yes	41 (31.5)
No	89 (68.5)

Game-Playing and Content-Learning Experiences

We applied the following two measures to examine the acceptability of the games: game-playing and content-learning experiences. Table 3 shows the distribution of ARCS elements (representing game-playing experiences), knowledge improvement and perceived persuasiveness (representing content-learning experiences), and intention to abstain from smoking (outcome). The item average scores for these variables (defined as the total score divided by the number of items) revealed that the participants provided positive responses. The ARCS-related responses indicated that most students agreed or strongly agreed that the educational games were motivational. The item average scores for attention, relevance, confidence, and satisfaction were all approximately 4 (measured using a Likert-type scale ranging from 1 to 5; scores for the four variables ranged from 3.95 to 4.19). There was a significant improvement in knowledge, with a mean difference of 2.97 between pretest and posttest (t_{129}=25.67, *P*<.001). Most students indicated that they were persuaded, with an item average score of 4.28 points (determined using a Likert-type scale ranging from 1 to 5).

Table 3. Descriptive results for the acceptability evaluation and intention to abstain from smoking.

Game-playing experiences	Item average scores	Sum of mean scores	SD
ARCS			
Attention (four items)	4.19	16.76	2.75
Relevance (four items)	3.95	15.78	2.70
Confidence (three items)	4.03	12.08	1.84
Satisfaction (four items)	4.10	16.38	2.52
Content-learning experiences			
Knowledge improvement (postscore−prescore)	N/A[a]	2.97[b]	1.39
Perceived persuasiveness (four items)	4.25	17.00	2.48
Outcome measure			
Behavioral intention (two items)	4.18	8.36	1.85

[a]N/A: not applicable.

[b]Mean score for the difference between pretest and posttest.

Measurement Model Assessment

We established the convergent validity of the latent variables through consideration of CR and AVE. All variables in the model displayed acceptable internal consistency, as evidenced by the fact that their CR scores were greater than 0.7 (ranging from 0.80 to 0.96), and all AVE values were higher than 0.5 (ranging from 0.57 to 0.92). Thus, it was concluded that the model had satisfactory convergent validity (Table 1). The values of these indices satisfied the recommended criteria and suggested that the proposed measurement model was reliable and valid.

As shown in Tables 4 and 5, we established the discriminative validity using the Fornell-Larcker criterion and HTMT. The Fornell-Larcker criterion was satisfied, as the square root of AVE for each variable exceeded the correlation of the latent variables. Almost all of the HTMT values were <1.00 [32], with one of the HTMT values being 1.06, indicating that this criterion was acceptable.

Table 4. Results for the Fornell-Larcker criterion.

Variable	Attention	Relevance	Confidence	Satisfaction	Knowledge	Perceived persuasiveness	Behavior intention
Attention							
Latent variable correlation	0.88[a,b]	0.61	0.59	0.65	0.05	0.66	0.42
Relevance							
Latent variable correlation	0.61	0.79	0.60	0.62	−0.03	0.70	0.49
Confidence							
Latent variable correlation	0.59	0.60	0.76	0.76	0.05	0.62	0.41
Satisfaction							
Latent variable correlation	0.65	0.62	0.76	0.78	0.01	0.70	0.51
Knowledge							
Latent variable correlation	0.05	−0.03	0.05	0.01	1.00	−0.01	0.03
Perceived persuasiveness							
Latent variable correlation	0.66	0.70	0.62	0.70	−0.01	0.84	0.60
Behavior intention							
Latent variable correlation	0.42	0.49	0.41	0.51	0.03	0.60	0.95

[a]In SmartPLS output, average variance extracted is calculated and written on the diagonal of the table.

[b]There are no *P* values to provide.

Table 5. Results for the heterotrait-monotrait ratio of correlations.

Variable	Attention	Relevance	Confidence	Satisfaction	Knowledge	Perceived persuasiveness	Behavior intention
Attention							
Value	N/A[a,b,c]	0.70	0.77	0.76	0.06	0.75	0.46
Relevance							
Value	0.70	N/A	0.82	0.74	0.07	0.83	0.56
Confidence							
Value	0.77	0.82	N/A	1.06	0.06	0.83	0.52
Satisfaction							
Value	0.76	0.74	1.06	N/A	0.03	0.84	0.59
Knowledge							
Value	0.06	0.07	0.06	0.03	N/A	0.04	0.03
Perceived persuasiveness							
Value	0.75	0.83	0.83	0.84	0.04	N/A	0.67
Behavior intention							
Value	0.46	0.56	0.52	0.59	0.03	0.67	N/A

[a]N/A: not applicable.

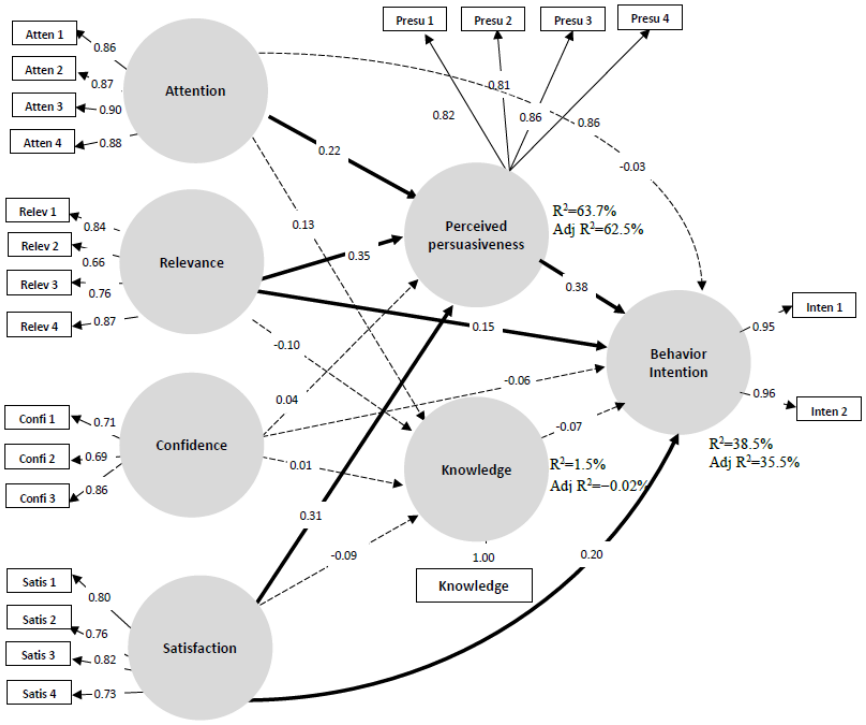
[b]The heterotrait-monotrait ratio is the geometric mean of the heterotrait-heteromethod correlations (ie, the correlations of indicators across constructs) divided by the average of the monotrait-heteromethod correlations (ie, the correlations of indicators within the same construct).

[c]There are no *P* values to provide.

Structural Model Assessment

The model exhibited adequate correspondence with the data (SRMR=0.075). To assess the structural model, we used the R^2 values and the path coefficients as the essential measures. As shown in Figure 2, the four latent variables of ARCS (attention, relevance, confidence, and satisfaction) accounted for 63.7% of the variance in perceived persuasiveness. However, these four latent variables accounted for only 1.1% of the variance in knowledge improvement. Overall, the four latent variables of ARCS, perceived persuasiveness, and knowledge improvement accounted for 38.2% of the variance in the intention to abstain from smoking.

Figure 2. Model for acceptability evaluation of educational virtual reality games.



Regarding the path analysis, Table 6 lists the path coefficients and *P* values for each hypothesis. The results of the hypothesis test showed that six of the 14 hypotheses were supported. These six supported hypotheses postulated that the three latent variables of attention, relevance, and satisfaction significantly influenced perceived persuasiveness (H1-2: path coefficient=0.23, *t*=3.19, *P*<.001; H2-2: path coefficient=0.35, *t*=4.28, *P*<.001; H4-2: path coefficient=0.31, *t*=3.49, *P*<.001). Only confidence did not influence perceived persuasiveness (H3-2: path coefficient=0.04, *t*=0.42, *P*=.67). In addition, relevance, satisfaction, and perceived persuasiveness significantly influenced the intention to abstain from smoking (H2: path coefficient=0.26, *t*=2.39, *P*=.02; H4: path coefficient=0.34, *t*=3.07, *P*=.002; H6: path coefficient=0.45, *t*=3.57, *P*<.001). In contrast, intention was not significantly influenced by knowledge improvement (H5: path coefficient=−0.06, *t*=1.31, *P*=.19).

In addition to the direct effect on the intention to abstain from smoking, an indirect effect was observed from relevance to intention via perceived persuasiveness (indirect effect, *t*=2.87, *P*=.004). Further, another indirect effect was observed from satisfaction to intention via perceived persuasiveness (indirect effect, *t*=2.11, *P*=.04). Finally, although not directly affecting the intention to abstain from smoking, attention showed an indirect effect on intention via perceived persuasiveness (indirect effect, *t*=2.16, *P*=.03).

Table 6. Results of hypothesis tests.

Hypothesis	Hypothesized paths	Original coefficient	Bootstrapped coefficient	*t*	*P* value	Path support
H1	Attention → behavioral intention	0.06	0.06	0.51	.61	No
H2[a]	Relevance → behavioral intention	0.27	0.26	2.39	.02	Yes
H3	Confidence → behavioral intention	−0.05	−0.04	0.42	.68	No
H4[a]	Satisfaction → behavioral intention	0.34	0.34	3.07	.002	Yes
H5	Knowledge → behavioral intention	−0.07	−0.06	1.31	.19	No
H6[a]	Perceived persuasiveness → behavioral intention	0.44	0.45	3.57	<.001	Yes
H1-1	Attention → knowledge	0.13	0.12	0.78	.44	No
H1-2[a]	Attention → perceived persuasiveness	0.22	0.23	3.19	<.001	Yes
H2-1	Relevance → knowledge	−0.11	−0.10	0.92	.36	No
H2-2[a]	Relevance → perceived persuasiveness	0.35	0.35	4.28	<.001	Yes
H3-1	Confidence → knowledge	0.01	0.01	0.06	.95	No
H3-2	Confidence → perceived persuasiveness	0.04	0.04	0.42	.67	No
H4-1	Satisfaction → knowledge	−0.09	−0.09	0.50	.62	No
H4-2[a]	Satisfaction → perceived persuasiveness	0.31	0.31	3.49	<.001	Yes

[a]Paths with significant coefficients.

Discussion

Acceptability Evaluation

In this study, we aimed to evaluate the acceptability of smoking-prevention–focused educational VR games among high school students. We measured acceptability based on the following two aspects: game-playing and content-learning experiences. As motivation is essential for effective learning [33], we adopted the ARCS motivation model as a framework for examining the students' game-playing experiences. Our findings indicated that the educational games invoked a noticeable degree of learning motivation. Based on the responses of the students to the ARCS-related items, most students agreed or strongly agreed that the educational games were motivational. Many researchers have applied the ARCS model as a framework for instructional design, and the efficacy of the ARCS framework for enhancing students' motivation has been validated [34,35]. Further, a previous study supported the effectiveness of applying the ARCS framework in the context of learning activities by successfully using it in a randomized controlled trial [36]; the applicability of the ARCS framework for new digital technology, such as mobile augmented reality, has also been proven [37]. Our findings are consistent with those of previous studies and revealed that the educational games could function as supplementary material that engages students and motivates them to learn. The content-learning experiences included in the games comprised knowledge improvement and perceived persuasiveness. We found that the students' knowledge had significantly improved as a result of playing the VR games. Further, similar to the item average scores for ARCS, most students agreed or strongly agreed that the games had persuaded them to improve their attitudes regarding smoking prevention.

We used persuasive strategies in the design of the human-computer interactions in the games. Several persuasive strategies have been suggested across previous studies [38,39];

however, our findings indicate that strategies concerning competition, self-awareness, feedback, social comparison, and reinforcement are effective in the context of smoking prevention among high school students. Our findings also revealed that the games' content-learning experiences were influenced by the game-playing experiences. Researchers have emphasized that, to achieve favorable learning outcomes, effective engagement should be pursued over mere passive engagement [40]; theoretically, the students who played our games were motivated to learn through ARCS experiences [41]. The ARCS framework comprises not only instructional approaches to motivate students, but also preliminary elements of effective content learning. We expected that motivated students would show better results in terms of content learning (ie, knowledge improvement and perceived persuasiveness). The study findings consequently showed that attention, relevance, and satisfaction had notable effects on content learning, but confidence did not. Specifically, for the educational games, students who reported high levels of attention, relevance, and satisfaction perceived themselves as being persuaded. Persuasion is a process that concerns changing or reinforcing attitudes or behaviors [24]. Our educational games were designed to discourage students from smoking by reinforcing nonsmoking attitudes and behaviors. In other words, we tried to influence students' behaviors through messages conveyed in the games. After playing the games, the students reported increased awareness of and concern regarding smoking prevention. In addition, their understanding and self-efficacy in relation to practicing preventive skills were enhanced.

Our findings revealed that only some of the four aspects of motivation (ARCS) contribute to perceived persuasiveness, namely, attention, relevance, and satisfaction. Similar to our findings, attention and relevance have previously been found to be associated with overall persuasiveness [21]. Thus, the present and existing findings imply that, in order to effectively persuade learners, the inclusion of attention and relevance in game design should be prioritized. Further, we found that students who were satisfied with the educational games tended to be persuaded, indicating that satisfaction should also be included in future game design. In contrast, confidence was not an influencing factor; a possible reason for this is that students felt the game tasks were not very challenging and were easy to complete. Some students suggested speeding up the games or increasing the difficulty or number of levels.

The ARCS experiences did not have significant effects on knowledge improvement. In other words, students' levels of knowledge were not influenced by the game-playing experiences. A possible reason the games nevertheless induced significant knowledge improvement is that the students received written materials while they were waiting in line to play the educational games. Our findings indicate the necessity of integrating other learning strategies into educational games in order to obtain knowledge improvement. As the games function as supplementary material for smoking-prevention education, game playing alone may not be sufficient to influence knowledge. Further, there are conflicting findings in the literature regarding the association among knowledge, positive attitudes, and intent to take action [42,43]. The structural model revealed that knowledge improvement does not influence the intention to abstain from smoking. Consequently, in addition to knowledge of the hazards of smoking, we recommend, for future game development, the inclusion of an emphasis on changing attitudes regarding smoking [44]. However, the gap between knowledge and behavior intention requires further study. Knowledge improvement was viewed as a mediator in the proposed model. Considering knowledge improvement as parallel to behavioral intention rather than a mediator is an idea for further studies.

Once a student's learning motivation is aroused, how does the level of motivation affect subsequent attitude changes or the achievement of learning objects? We examined this mechanism by analyzing the path from ARCS elements to intention to abstain from smoking. The results revealed that the four aspects of motivation (ARCS) had different influencing paths to intention. In addition to directly influencing students' intention to abstain from smoking, relevance and satisfaction had indirect effects on intention via perceived persuasiveness. Although the level of attention did not directly influence intention, it had an indirect effect on intention via perceived persuasiveness. These findings indicate that perceived persuasiveness has mediating effects on motivation's influence on intention. To discourage students from smoking, along with fostering their motivation to learn, perceived persuasiveness should be taken into consideration.

Implementation

Future efforts to design educational games should incorporate theoretical frameworks that can advance the understanding of the influencing mechanisms for motivation. In this research, we proposed the integration of the ARCS framework and persuasive strategies, and the PLS-SEM results consequently revealed that game-playing experiences that are based on the ARCS framework have direct effects on persuasiveness. Future educational game–focused research could adopt this integrative approach. Second, teaching staff who provide smoking-prevention education in schools could seek to make these VR games as accessible for students as possible, such as through holding activities in fair stalls on school days or during orientation for first-year students. Interesting games can be used to increase students' awareness of smoke-free campus policies and strengthen students' commitment to abstain from smoking. Finally, knowledge alone is unlikely to generate behavioral intention; thus, designers of smoking-prevention activities should seek to expand their focus beyond the improvement of knowledge. In addition, a possible problem with true/false questions is that respondents can easily guess the correct answer [45]. Multiple choice questions should be adopted in the future to measure knowledge.

Limitations

There are several limitations in this study. First, our investigation of the influencing paths for behavior intention was performed in the context of smoking prevention (an unhealthy behavior). While we were certain of the applicability of the ARCS framework and persuasive strategies to educational games that focused on such behaviors, the findings may not be fully generalizable to contexts concerning healthy behaviors. Thus,

they should be applied with caution. Second, we could not determine the separate effects of each educational game on behavioral intention, as the participants' game-playing experiences were obtained as a whole, rather than in terms of each type of game. Third, we could not validate the intervention effects of the educational games, as no comparison group was involved. Finally, one of the HTMT values was larger than 1.00. This implies that two latent constructs may be conceptually similar. Despite these limitations, this study, which was intended as an acceptability evaluation, elaborated on the mechanisms of game-playing and content-learning experiences in the context of behavior intention. We recommend that this experiment design be adopted in future endeavors.

Conclusion

In this study, we evaluated the acceptability of educational VR games for use in smoking-prevention education for high school students. Our findings revealed that the educational games were positively accepted by the participating students, and that the integration of the ARCS framework and persuasive strategies is an applicable approach for smoking-prevention education. The ARCS framework was significantly associated with perceived persuasiveness, and perceived persuasiveness has direct effects on the intention to abstain from smoking. We recommend that the games be included as teaching material in smoking-prevention education.

Acknowledgments

We thank all the study participants for their time and commitment. This research was funded by the K-12 Education Administration, Ministry of Education, Taiwan, ROC. This article was subsidized by the National Taiwan Normal University, Taiwan, ROC.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Theoretical concepts used in the design of the educational games.
[DOCX File , 3531 KB - jmir_v23i9e28037_app1.docx]

References

1. WHO report on the global tobacco epidemic 2015. World Health Organization. 2015. URL: https://apps.who.int/tobacco/global_report/2015/en/index.html [accessed 2021-09-02]
2. Tobore TO. On the potential harmful effects of E-Cigarettes (EC) on the developing brain: The relationship between vaping-induced oxidative stress and adolescent/young adults social maladjustment. J Adolesc 2019 Oct;76:202-209. [doi: 10.1016/j.adolescence.2019.09.004] [Medline: 31574388]
3. Dierker L, Hedeker D, Rose J, Selya A, Mermelstein R. Early emerging nicotine dependence symptoms in adolescence predict daily smoking in young adulthood. Drug Alcohol Depend 2015 Jun 01;151:267-271 [FREE Full text] [doi: 10.1016/j.drugalcdep.2015.03.009] [Medline: 25840749]
4. Guidelines for implementation of Article 12 of the WHO Framework Convention on Tobacco Control (Education, communication, training and public awareness). World Health Organization. URL: https://www.who.int/fctc/guidelines/Decision.pdf?ua=1 [accessed 2021-09-02]
5. Leavens ELS, Lechner WV, Stevens EM, Miller MB, Meier E, Brett EI, et al. Electronic cigarette and combustible cigarette use following a campus-wide ban: Prevalence of use and harm perceptions. J Am Coll Health 2020;68(4):332-335. [doi: 10.1080/07448481.2018.1551803] [Medline: 30681933]
6. Gaiha SM, Duemler A, Silverwood L, Razo A, Halpern-Felsher B, Walley SC. School-based e-cigarette education in Alabama: Impact on knowledge of e-cigarettes, perceptions and intent to try. Addict Behav 2021 Jan;112:106519. [doi: 10.1016/j.addbeh.2020.106519] [Medline: 32890911]
7. Boyle EA, Hainey T, Connolly TM, Gray G, Earp J, Ott M, et al. An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. Computers & Education 2016 Mar;94:178-192. [doi: 10.1016/j.compedu.2015.11.003]
8. Arlati S, Colombo V, Ferrigno G, Sacchetti R, Sacco M. Virtual reality-based wheelchair simulators: A scoping review. Assist Technol 2020 Nov 01;32(6):294-305. [doi: 10.1080/10400435.2018.1553079] [Medline: 30615571]
9. Kamińska MS, Miller A, Rotter I, Szylińska A, Grochans E. The effectiveness of virtual reality training in reducing the risk of falls among elderly people. Clin Interv Aging 2018;13:2329-2338 [FREE Full text] [doi: 10.2147/CIA.S183502] [Medline: 30532523]
10. Prasertsakul T, Kaimuk P, Chinjenpradit W, Limroongreungrat W, Charoensuk W. The effect of virtual reality-based balance training on motor learning and postural control in healthy adults: a randomized preliminary study. Biomed Eng Online 2018 Sep 18;17(1):124 [FREE Full text] [doi: 10.1186/s12938-018-0550-0] [Medline: 30227884]
11. Sun H, Li S, Zhu Y, Hsiao B. The effect of user's perceived presence and promotion focus on usability for interacting in virtual environments. Appl Ergon 2015 Sep;50:126-132. [doi: 10.1016/j.apergo.2015.03.006] [Medline: 25959326]

12. Lamb RL, Annetta L, Firestone J, Etopio E. A meta-analysis with examination of moderators of student cognition, affect, and learning outcomes while using serious educational games, serious games, and simulations. Computers in Human Behavior 2018 Mar;80:158-167. [doi: 10.1016/j.chb.2017.10.040]
13. Lu OHT, Huang JCH, Huang AYQ, Yang SJH. Applying learning analytics for improving students engagement and learning outcomes in an MOOCs enabled collaborative programming course. Interactive Learning Environments 2017 Jan 16;25(2):220-234. [doi: 10.1080/10494820.2016.1278391]
14. K-12 Education Administration, Ministry of Education. URL: https://www.k12ea.gov.tw/Tw [accessed 2021-09-01]
15. Keller JM. Development and use of the ARCS model of instructional design. Journal of Instructional Development 1987 Sep;10(3):2-10. [doi: 10.1007/bf02905780]
16. Daugherty KK. ARCS motivation model application in a pharmacy elective. Curr Pharm Teach Learn 2019 Dec;11(12):1274-1280. [doi: 10.1016/j.cptl.2019.09.009] [Medline: 31836153]
17. Soltani AS, Bigdeli S, Kouhpayezade J. Designing a Reliable and Valid Instrument to Determine the Effective Motivational Factors on Medical Students Learning Based on Keller's ARCS Model. Educ Strategy Med Sci 2014;6(4):241-246 [FREE Full text]
18. Li K, Moore DR. Motivating Students in Massive Open Online Courses (MOOCs) Using the Attention, Relevance, Confidence, Satisfaction (ARCS) Model. J Form Des Learn 2018 Jul 19;2(2):102-113. [doi: 10.1007/s41686-018-0021-9]
19. Amaritakomol A, Kanjanavanit R, Suwankruhasn N, Topaiboon P, Leemasawat K, Chanchai R, et al. Enhancing Knowledge and Self-Care Behavior of Heart Failure Patients by Interactive Educational Board Game. Games Health J 2019 Jun;8(3):177-186. [doi: 10.1089/g4h.2018.0043] [Medline: 30383438]
20. Rosendahl KI, Galanti MR, Gilljam H, Ahlbom A. Knowledge about tobacco and subsequent use of cigarettes and smokeless tobacco among Swedish adolescents. J Adolesc Health 2005 Sep;37(3):224-228. [doi: 10.1016/j.jadohealth.2004.08.021] [Medline: 16109342]
21. Orji R, Reilly D, Oyibo K, Orji FA. Deconstructing persuasiveness of strategies in behaviour change systems using the ARCS model of motivation. Behaviour & Information Technology 2018 Sep 17;38(4):319-335. [doi: 10.1080/0144929x.2018.1520302]
22. Oinas-Kukkonen H, Harjumaa M. Persuasive Systems Design: Key Issues, Process Model, and System Features. CAIS 2009;24. [doi: 10.17705/1cais.02428]
23. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 05;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
24. Ferebee S. Successful Persuasive Technology for Behavior Reduction: Mapping to Fogg's Gray Behavior Grid. In: Ploug T, Hasle P, Oinas-Kukkonen H, editors. Persuasive Technology. PERSUASIVE 2010. Lecture Notes in Computer Science, vol 6137. Berlin, Heidelberg: Springer; 2010:70-81.
25. Hess TJ, McNab AL, Basoglu KA. Reliability Generalization of Perceived Ease of Use, Perceived Usefulness, and Behavioral Intentions. MISQ 2014 Jan 1;38(1):1-28. [doi: 10.25300/MISQ/2014/38.1.01]
26. O'Brien HL, Toms EG. What is user engagement? A conceptual framework for defining user engagement with technology. J. Am. Soc. Inf. Sci 2008 Apr;59(6):938-955. [doi: 10.1002/asi.20801]
27. Davis LL. Instrument review: Getting the most from a panel of experts. Applied Nursing Research 1992 Nov;5(4):194-197. [doi: 10.1016/S0897-1897(05)80008-4]
28. Hair JF, Ringle CM, Sarstedt M. PLS-SEM: Indeed a Silver Bullet. Journal of Marketing Theory and Practice 2014 Dec 08;19(2):139-152. [doi: 10.2753/MTP1069-6679190202]
29. Wong KKK. Partial Least Squares Structural Equation Modeling (PLS-SEM) Techniques Using SmartPLS. Marketing Bulletin 2013;24:1-32 [FREE Full text]
30. Hair JF, Sarstedt M, Ringle CM. Rethinking some of the rethinking of partial least squares. European Journal of Marketing 2019 Apr 08;53(4):566-584. [doi: 10.1108/ejm-10-2018-0665]
31. F Hair Jr J, Sarstedt M, Hopkins L, G Kuppelwieser V. Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research. European Business Review 2014;26(2):106-121. [doi: 10.1108/EBR-10-2013-0128]
32. Henseler J, Hubona G, Ray PA. Using PLS path modeling in new technology research: updated guidelines. Industrial Management & Data Systems 2016;116(1):2-20. [doi: 10.1108/IMDS-09-2015-0382]
33. Waheed M, Kaur K, Ain N, Hussain N. Perceived learning outcomes from Moodle. Information Development 2016 Jul 09;32(4):1001-1013. [doi: 10.1177/0266666915581719]
34. Turel YK, Ozer Sanal S. The effects of an ARCS based e-book on student's achievement, motivation and anxiety. Computers & Education 2018 Dec;127:130-140. [doi: 10.1016/j.compedu.2018.08.006]
35. Angelo TA. Assessing Motivation to Improve Learning: Practical Applications of Keller's MVP Model and ARCS-V Design Process. Teaching and Learning 2017 Dec 13;2017(152):99-108. [doi: 10.1002/tl.20272]
36. Kim H, Suh EE. The Effects of an Interactive Nursing Skills Mobile Application on Nursing Students' Knowledge, Self-efficacy, and Skills Performance: A Randomized Controlled Trial. Asian Nurs Res (Korean Soc Nurs Sci) 2018 Mar;12(1):17-25 [FREE Full text] [doi: 10.1016/j.anr.2018.01.001] [Medline: 29463484]

37. Chang Y, Hu K, Chiang C, Lugmayr A. Applying Mobile Augmented Reality (AR) to Teach Interior Design Students in Layout Plans: Evaluation of Learning Effectiveness Based on the ARCS Model of Learning Motivation Theory. Sensors (Basel) 2019 Dec 23;20(1) [FREE Full text] [doi: 10.3390/s20010105] [Medline: 31878045]
38. Orji R, Vassileva J, Mandryk RL. Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. User Model User-Adap Inter 2014 Jul 14;24(5):453-498. [doi: 10.1007/s11257-014-9149-8]
39. Chow Y, Susilo W, Phillips J, Baek J, Vlahu-Gjorgievska E. Video Games and Virtual Reality as Persuasive Technologies for Health Care: An Overview. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications 2017;8:18-35. [doi: 10.22667/JOWUA.2017.09.30.018]
40. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
41. Kurt PY, Keçik I. The effects of ARCS motivational model on student motivation to learn English. European Journal of Foreign Language Teaching 2017;2(1):22-44. [doi: 10.5281/zenodo.254605]
42. Park J, Lim MK, Yun EH, Oh J, Jeong BY, Cheon Y, et al. Influences of Tobacco-Related Knowledge on Awareness and Behavior towards Smoking. J Korean Med Sci 2018 Nov 19;33(47):e302 [FREE Full text] [doi: 10.3346/jkms.2018.33.e302] [Medline: 30450026]
43. Ganley BJ, Rosario DI. The smoking attitudes, knowledge, intent, and behaviors of adolescents and young adults: Implications for nursing practice. JNEP 2013 Jan 01;3(1):40. [doi: 10.5430/jnep.v3n1p40]
44. Lee HS, Addicott M, Martin LE, Harris KJ, Goggin K, Richter KP, et al. Implicit Attitudes and Smoking Behavior in a Smoking Cessation Induction Trial. Nicotine Tob Res 2017 Dec 13;20(1):58-66 [FREE Full text] [doi: 10.1093/ntr/ntw259] [Medline: 27679606]
45. de Kok LC, Oosting D, Spruit M. The Influence of Knowledge and Attitude on Intention to Adopt Cybersecure Behaviour. ISIJ 2020;46(3):251-266. [doi: 10.11610/isij.4618]

Abbreviations

ARCS: attention, relevance, confidence, and satisfaction
AVE: average variance extracted
CR: composite reliability
CVI: content validity index
e-cigarette: electronic cigarette
GBL: game-based learning
HMD: head-mounted display
HTMT: heterotrait-monotrait
PLS-SEM: partial least squares structural equation modeling
SRMR: standardized root mean square residual
VR: virtual reality

Edited by R Kukafka; submitted 24.02.21; peer-reviewed by JY Liao, CY Lin; comments to author 16.03.21; revised version received 30.04.21; accepted 14.06.21; published 28.09.21.

Please cite as:
Guo JL, Hsu HP, Lai TM, Lin ML, Chung CM, Huang CM
Acceptability Evaluation of the Use of Virtual Reality Games in Smoking-Prevention Education for High School Students: Prospective Observational Study
J Med Internet Res 2021;23(9):e28037
URL: https://www.jmir.org/2021/9/e28037
doi:10.2196/28037
PMID:34581679

Review

Predicting Patient Deterioration: A Review of Tools in the Digital Hospital Setting

Kay D Mann[1], MMathStat, PhD; Norm M Good[1], MSc; Farhad Fatehi[2,3], MD, PhD, FIAHSI; Sankalp Khanna[1], MInfTech, PhD, FAIDH; Victoria Campbell[4,5,6], MBBS, FRACP, FCICM; Roger Conway[4], RN, PGDip; Clair Sullivan[2,7], MBBS, MD, FRACP, FAIDH; Andrew Staib[5,8,9], MBBS, PhD, FACEM, FAIDH; Christopher Joyce[8,9], MB, ChB, PhD, FANZCA, FCICM; David Cook[8,9,10], MBBS, PhD, FANZCA, FCICM

[1]The Australian e-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation, Brisbane, Australia

[2]Centre for Health Services Research, Faculty of Medicine, The University of Queensland, Brisbane, Australia

[3]School of Psychological Sciences, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia

[4]Sunshine Coast University Hospital, Sunshine Coast Hospital and Health Service, Birtinya, Australia

[5]Clinical Excellence Queensland, Queensland Health, Queensland, Australia

[6]School of Medicine, Griffith University, Nathan Campas, Australia

[7]Metro North Hospital and Health Service, Brisbane, Australia

[8]Princess Alexandra Hospital, Metro South Hospital and Health Service, Brisbane, Australia

[9]Faculty of Medicine, The University of Queensland, Brisbane, Australia

[10]School of Computer Science, Faculty of Science, Queensland University of Technology, Brisbane, Australia

Corresponding Author:
Norm M Good, MSc
The Australian e-Health Research Centre
Commonwealth Scientific and Industrial Research Organisation
Level 7, Surgical Treatment and Rehabilitation Service
Herston
Brisbane, 4029
Australia
Phone: 61 732533640
Email: norm.good@csiro.au

Abstract

Background: Early warning tools identify patients at risk of deterioration in hospitals. Electronic medical records in hospitals offer real-time data and the opportunity to automate early warning tools and provide real-time, dynamic risk estimates.

Objective: This review describes published studies on the development, validation, and implementation of tools for predicting patient deterioration in general wards in hospitals.

Methods: An electronic database search of peer reviewed journal papers from 2008-2020 identified studies reporting the use of tools and algorithms for predicting patient deterioration, defined by unplanned transfer to the intensive care unit, cardiac arrest, or death. Studies conducted solely in intensive care units, emergency departments, or single diagnosis patient groups were excluded.

Results: A total of 46 publications were eligible for inclusion. These publications were heterogeneous in design, setting, and outcome measures. Most studies were retrospective studies using cohort data to develop, validate, or statistically evaluate prediction tools. The tools consisted of early warning, screening, or scoring systems based on physiologic data, as well as more complex algorithms developed to better represent real-time data, deal with complexities of longitudinal data, and warn of deterioration risk earlier. Only a few studies detailed the results of the implementation of deterioration warning tools.

Conclusions: Despite relative progress in the development of algorithms to predict patient deterioration, the literature has not shown that the deployment or implementation of such algorithms is reproducibly associated with improvements in patient outcomes. Further work is needed to realize the potential of automated predictions and update dynamic risk estimates as part of an operational early warning system for inpatient deterioration.

(J Med Internet Res 2021;23(9):e28209) doi:10.2196/28209

KEYWORDS

patient deterioration; early warning scores; digital tools; vital signs; electronic medical record

Introduction

Background

In recent years, proactive clinical processes have been developed to target timely and appropriate care for deteriorating or high-risk patients. Emergency responses such as *Rapid Response Systems* have been implemented with the aim to intervene and avoid preventable death, cardiac arrest, or transfer to an intensive care unit (ICU) in adult [1-3] and pediatric [4,5] patients. These systems have evolved to consist of a recognition (afferent) limb, commonly known as early warning scores (EWSs), and a response (efferent) limb (escalation and intervention). The responders rely on an accurate recognition limb, which in turn relies on a combination of empirical rules, statistical models, and clinical judgment to recognize deterioration. Initial EWSs were limited to vital signs, as these were the only routinely collected physiological data available for analysis in real time. The EWSs are currently available as paper or digital tools and vary significantly in their modeling, design, and escalation guidance [3,6-10].

The growth of rich, detailed, and dynamic clinical digital documentation in electronic medical records (EMRs) raises the possibility of using a broader range of clinical data, including pathology and diagnoses. An EMR collects the detailed phenotype of the patient in real time. Data collected, such as previous history, comorbidities, and demographic descriptors, are static during an admission. Observations of dynamic processes such vital signs, clinical measurements, imaging, and laboratory results that document biological and pathological processes are also recorded and continuously updated. Further evolving diagnoses, events, and interventions (eg, operations and drugs that are administered) capture the changing status of a patient. Finally, a rich source of dynamic information lies in the metadata: the timing, frequency, and location of actions and observations that occur to the patient and patient movements in the system. These data enable refined predictive models and more effective, patient-specific treatments. To create clinical decision support, these data must be analyzed and risk interpreted, and then critically, the clinical decision support communicating this real-time risk needs to be engineered back into the routine clinical workflows of the clinicians caring for the patient.

There is a diversity of models for predicting patient deterioration. Some risk estimates are static systems that identify high-risk patients at the time of diagnosis and allow triage of patients to a higher intensity care destination. Other approaches use vital sign observations to maintain an up-to-date risk evaluation. Typically, these dynamic systems either identify an extreme singular derangement or use weighted sums of a few vital signs and their variation from normal values. In both situations, the likelihood of deterioration and poor outcome anticipated with worsening values is increased. In some cases, these models have been developed and validated only for specific and narrow patient groups [11-13], whereas other general models are applicable to wider adult or pediatric ward patients [3-5,10,14].

Objectives

The aim of this review is to identify studies conducted within a general hospital setting that have attempted to develop prediction algorithms for detecting a deteriorating ward patient in real time, based primarily on routinely collected EMR data. This review includes model statistical validation and, where available, the results of digital hospital implementation of new or existing prediction models or rule-based systems for predicting patient deterioration. A secondary aim is to review those successful examples for common data elements, approaches, and statistical or machine learning techniques that were associated with successful clinical use.

Methods

This review followed the recommendations of the Center for Reviews and Dissemination [15] and PRISMA (Preferred Reporting in Systematic Reviews and Meta-Analyses) [16].

Data Sources and Search Strategy

A systematic search of PubMed, Scopus, Web of Science, IEEE Xplore, and ACM Digital Library using a combination of controlled vocabulary (eg, MeSH [Medical Subject Headings] terms) and free text keywords was conducted. The search strategy was first developed for PubMed, guided by the recommendations of Hausner et al [17] and Fatehi et al [18], and transposed to other databases. Across databases, free text keywords remained the same, but controlled vocabulary was mapped where possible (eg, from MeSH to Index Terms). The search was limited by language (English), date of publication (January 1, 2008, to June 30, 2020) and type of publication (original papers).

Screening and Study Selection

Reports that have developed prediction algorithms for detection of clinical patient deterioration in real time, primarily based on routinely collected patient data, were identified. The included studies had to (1) use any kind of (electronic) patient record, (2) use an early warning tool for patient deterioration, (3) use of a system that was dynamic, or observations of a patient over time, and (4) document the model statistical accuracy and performance. The studies were peer reviewed and published in journals or conference proceedings. The focus of this review was on tools in a general hospital ward and in a real time setting. Excluded studies were those conducted solely in a critical care or an emergency department setting, those limited to a single diagnosis or organ dysfunction patient cohort, those that used a static time point or an observation to assess risk of deterioration, and those that were qualitative with no quantitative assessments.

The results of electronic database searches were exported into an EndNote [19] library, and duplicates were removed. Title and abstract screening was coordinated on Rayyan web application (Qatar Computing Research Institute) [20]. Two independent reviewers screened titles and abstracts for relevance, and unresolved differences were adjudicated by a third reviewer. Full texts of remaining papers were assessed by 2 independent reviewers against the inclusion and exclusion criteria. Unresolved differences were adjudicated by a third reviewer. Where more than one publication was found for the same project, the papers were grouped, and the publication with the most comprehensive findings was included.

Data Extraction and Synthesis

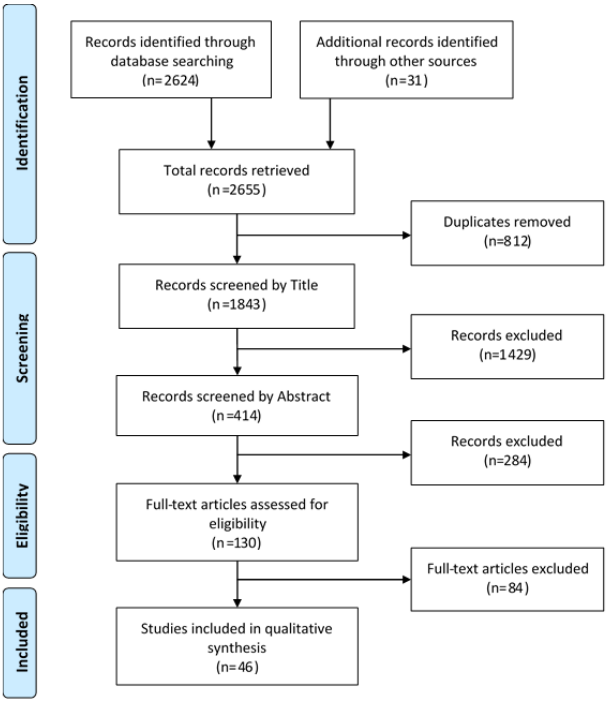
An electronic web form was developed according to the aims of this review and was used for data extraction from full-text papers. Data extraction was conducted by one reviewer and checked by a second reviewer. Owing to the heterogeneity of the included studies in terms of study designs, aims, and outcome measures, it was not possible to conduct a meta-analysis. Thus, a narrative approach was adopted for data synthesis. An assessment of study quality, including risks of bias, was conducted using PROBAST (Prediction Model Risk of Bias Assessment Tool; UMC Utrecht) [21] and ROBINS-I (Risk of Bias in Non-Randomised Intervention Studies; Cochrane) tool [22].

Results

Overview

The initial electronic search of five databases yielded 2624 records, and an additional 31 records were identified from other sources (Figure 1). After removing duplicates, 1843 unique records remained for eligibility assessment. Following the screening of titles and abstracts, 130 papers were deemed relevant, and their full texts were obtained. After a full-text inspection, 46 papers met the eligibility criteria and were included in this review.

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



Characteristics of the Studies

Of the 46 papers in this review, the majority (37/46, 80%) were retrospective studies that attempted to develop and/or evaluate the performance of a prediction model or compare the performance of a number of predictive models using historical patient data [23-59]. One study reported both the retrospective development of a prediction model and a prospective observational study of the developed model [60]. The remaining studies included one randomized controlled trial [61], 4 before-after (implementation) studies [62-65], and 3 prospective observational studies [66-68].

Overview of Studies

Studies from the same setting and institution were grouped and ordered according to the chronological model development, evaluation, and implementation, if available.

Of the identified studies, 6 described a body of research undertaken by researchers from Kaiser Permanente in California, as well as the University of Chicago [34,35,40,44,62,63] (Table 1). Kaiser Permanente is a large integrated managed care organization in the United States and began the deployment of EMRs in its hospitals in 2006. Escobar et al [35] reported the development of the early detection of impending physiological deterioration algorithm, which was further developed and subsequently named the Advanced Alert Monitor [44] and the early detection of impending physiological deterioration version 2 algorithm [40]. These models were developed to predict the

risk of unplanned ICU transfer or mortality, using historical data of hundreds of thousands of patients retrieved from the Epic system, have also been retrospectively tested in a simulation study for feasibility [40], piloted in 2 hospitals [62], and more recently implemented as the Advanced Alert Monitor program and evaluated across the remaining Kaiser Permanente Hospitals [63,69,70].

Table 1. Studies undertaken on Kaiser Permanente Hospitals.

Author and year	Study design	Settings	Study aim	Model type or used	Outcome measure	Key findings
Escobar et al, 2012 [35]	RS[a]: tool development	102,422 hospitalizations; 14 KP[b] hospitals; 2006-2009	To develop a model for the prediction of unplanned ICU[c] transfer using EMR[d] data	Pooled logistic regression models	Transfer to ICU or death	EMR-based detection of impending deterioration outside the ICU is feasible. The overall performance of the model incorporating physiology, diagnosis, and longitudinal data was superior to MEWS[e]. Model *c* statistic 0.85, validation model 0.78
Escobar et al, 2013 [34]	RS: tool development	391,584 hospitalizations; 248,383 patients; 21 KP hospitals; 2008-2011	To develop a risk adjustment methodology applicable to all hospitalized patients	Pooled logistic regression models	Death	Risk adjustment of hospital mortality using EMR is feasible. Incorporation of physiological data increased model discrimination and explanatory power. Model *c* statistic 0.80, validation model 0.88
Kipnis et al, 2016 [44]	RS: tool development and evaluation	649,418 hospitalizations; 374,838 patients; 21 KP hospitals; 2010-2013	To describe the development and performance of an automated EWS[f] based on EMR data: The AAM[g]	Discrete time logistic regression	Transfer to ICU or death	The AAM had better performance compared with NEWS[h] and eCART[i] in all metrics and prediction intervals (AUC[j] 0.82). Around half the alerts triggered occurred within 12 hours of the event and almost two-thirds within 24 hours.
Escobar et al, 2016 [62]	Prospective study: tool implementation	2 KP hospitals	To detail technical and operational challenges of deploying an EWS	AAM	Pre- and postimplementation metrics	The pilot implementation in 2 hospitals was successful and further deployment to other hospitals will go ahead.
Hu et al, 2018 [40]	RS: tool simulation and feasibility	174,632 hospitalizations; 21 KP hospitals	To evaluate the impact of proactive transfer to ICU based on EDIP2[k] score, on mortality rate and LOS[l]	EDIP2	Transfer to ICU or death	Proactively transferring of the most severe patients could reduce mortality rates without sacrificing other patient outcomes.
Escobar et al, 2020 [63]	RS: tool implementation and evaluation	548,838 hospitalizations; 326,816 patients; 21 KP hospitals	To evaluate a staggered deployment of an automated predictive model; identifying patients at high risk for clinical deterioration.	AAM	Pre- and postimplementation outcomes: transfer to ICU, 30-day mortality, LOS, favorable status	30-day mortality after an alert was lower in the intervention compared with control (three deaths avoided per 1000 eligible patients). The intervention was also associated with lower incidence of ICU admission, higher percentage of patients with favorable status 30 days after alert, shorter LOS, and longer survival.

[a]RS: retrospective study.

[b]KP: Kaiser Permanente.

[c]ICU: intensive care unit.

[d]EMR: electronic medical record.

[e]MEWS: Modified Early Warning Score.

[f]EWS: early warning score.

[g]AAM: Advanced Alert Monitor.

[h]NEWS: National Early Warning Score.

[i]eCART: electronic Cardiac Arrest Risk Triage.

[j]AUC: area under the receiver operating characteristic curve.

[k]EDIP2: early detection of impending deterioration version 2.

[l]LOS: length of stay.

Another group of papers from the University of Chicago program was based on experiments using the electronic patient records from a number of hospitals in the Chicago area [29-33,37,66] (Table 2). Churpek et al [29] in several papers reported the development and validation of the electronic Cardiac Arrest Risk Triage score and other tools for predicting cardiac arrest and ICU transfer in Chicago (Table 2). They also showed that mortality and cardiac arrest were easier to predict than ICU transfer [29], and machine learning methods were more accurate than logistic regression for predicting patient deterioration [32].

Table 2. Studies conducted in Chicago hospitals.

Author and year	Study design	Setting	Study aim	Model type or used	Prediction event	Key findings
Churpek et al, 2012 [31]	RS[a]: tool development and evaluation	47,427 patients, 1 hospital, 2008-2011	To develop a CART[b] score and compare with the MEWS[c]	Logistic regression	CA[d] or transfer to ICU[e]	The CART score more accurately predicted cardiac arrest than the MEWS. Model AUC[f] 0.84
Churpek et al, 2013 [29]	RS: tool development	59,643 patients, 1 hospital, 2008-2011	To assess the impact of outcome selection on the performance of prediction algorithms	Logistic regression (4 models)	CA, transfer to ICU, death, all combined	Mortality is the easiest outcome to predict (AUC range 0.73-0.82), and ICU transfer is the most difficult.
Churpek et al, 2014 [30]	RS: tool development	59,301 patients, 1 hospital, 2008-2011	To derive and validate a prediction model for CA	Logistic regression	CA and transfer to ICU	The model can simultaneously predict the risk of CA and ICU transfer and was more accurate than ViEWS[g]. Model AUC 0.88 for CA, 0.77 for ICU
Churpek et al, 2014 [33]	RS: tool development	269,999 admissions, 5 hospitals, 2008-2011	To develop and validate eCART[h] score using commonly collected EMR[i] data	Survival analysis	CA, transfer to ICU, or death	eCART score was more accurate than MEWS for detecting CA, ICU transfer, or death. Model AUC 0.83 for CA, 0.75 for ICU transfer, 0.93 for death and 0.77 all combined
Somanchi et al, 2015 [56]	RS: tool development	133,000 patients, 4 hospitals, 2006-2011	To develop a prediction model for Code Blue, using EMR data, and compare with MEWS	SVM[j] and logistic regression	Code blue event in the next x hours	The model was able to predict Code Blue with ~80% recall and 20% false positive rate 4 hours ahead of the event. It out-performed MEWS.
Churpek et al, 2016 [32]	RS: tool development and evaluation	269,999 patients, 5 hospitals, 2008-2013	To compare the accuracy of different techniques for detecting clinical deterioration on the wards	Logistic, decision trees, SVM, K-NN[k], neural net, MEWS	CA, transfer to ICU, or death	This multicenter study showed that several machine learning methods can more accurately predict clinical deterioration than logistic regression.
Kang et al, 2016 [66]	Prospective study: feasibility study	3889 admissions, 3 wards, 2013-2014	To assess the feasibility of a real-time risk stratification tool	eCART	CA, transfer to ICU, or death	eCART score identified more CA and ICU transfers, many hours in advance, compared with standard RRT[l] activation.
Green et al, 2018 [37]	RS: tool evaluation	107,868 admissions, 5 hospitals, 2008-2013	To compare the BTF[m] calling criteria to MEWS, NEWS[n] and eCART score	BTF, NEWS, MEWS, eCART	CA, transfer to ICU, or death (24 hours)	eCART was more accurate than BTF, MEWS, NEWS for predicting the composite outcome of CA, ICU transfer and death. eCART AUC 0.80 (0.79-0.80)
Bartkowiak et al, 2019 [25]	RS: tool evaluation	32,537 admissions, 1 hospital, 2008-2016	To assess the accuracy of three EWS[o] postoperatively	NEWS, MEWS, eCART	CA, ICU transfer or ward, or death	The eCART score was the most accurate followed by MEWS. Maximum respiratory rate was the most predictive vital sign.
Mayampurath et al, 2019 [48]	RS: tool development	115,825 admissions, 1 hospital, 2008-2016	To develop a model from visual timelines to predict mortality	Convolutional neural network	Death	The model was more accurate than MEWS and SOFA[p], validation model AUC 0.91, and visual timelines enabled interpretation of a deep neural network.

[a]RS: retrospective study.

[b]CART: Cardiac Arrest Risk Triage.

[c]MEWS: Modified Early Warning Score.

[d]CA: cardiac arrest.

[e]ICU: intensive care unit.

[f]AUC: area under the receiver operating characteristic curve.

[g]ViEWS: VitalPAC Early Warning Score.

[h]eCART: electronic Cardiac Arrest Risk Triage.

[i]EMR: electronic medical record.

[j]SVM: support vector machine.

[k]K-NN: K-nearest neighbors.

[l]RRT: rapid response team.

[m]BTF: Between the Flags.

[n]NEWS: National Early Warning Score.

[o]EWS: early warning score.

[p]SOFA: Sequential Organ Failure Assessment.

Several other studies have evaluated screening tools such as the National Early Warning Score, Modified Early Warning Score, Rothman Index, and Sequential Organ Failure Assessment, for the early warning and detection of patient deterioration in hospital settings across different countries [26,39,41,50,52,54,57,59,64,65,67] (Table 3). The evaluation included applying tools retrospectively on historical clinical data to assess the feasibility of future use as well as assessing tools prospectively alongside the standard clinical systems. The studies of screening tools reported differing results of both good and poor predictive accuracy and usefulness for escalation of care through medical emergency team (MET) or rapid response team (Table 3).

Of the studies detailing the implementation of scoring tools across institutions [50-52,61,62,64,65], high-risk patients were appropriately identified as aiding in clinical response (Table 4). However, when comparing intervention and control patient cohorts, differing results were seen with either significant reductions or no impact on the assessed deterioration events.

Within the reviewed papers, deterioration prediction methodologies included single and multiparameter scoring tools, such as the National Early Warning Score and Modified Early Warning Score, as well as statistical and machine learning methods. Single and multiparameter scores were derived from a set of vital sign threshold derangements (Tables 3 and 4). The statistical and machine learning methods included logistic regression, survival models, Cox regression, Gaussian process regression, Markov models, decision trees (random forest and gradient boosted trees), K-nearest neighbor, support vector machine, and neural networks (Tables 1 and 2, and Multimedia Appendix 1 [23,24,27,28,36,38,42,43,45-47,49,51,53,55,58,60,68]). Most of these models attempted to account for changes in physiologic measures over time using novel model frameworks, for example, taking a sliding time window looking forward or backward in time to predict outcomes.

Across all the studies, models of greater complexity and statistical and machine learning methods were shown to have superior performance in predicting deterioration than scoring tools (Multimedia Appendix 1). For example, there was timelier detection and earlier warning of high deterioration risk in more complex models. Furthermore, the discrimination of patient deterioration using statistical and machine learning methods (as assessed with the area under the receiver operating characteristic curve or *c* statistic) outperformed conventional tools.

All patient deterioration prediction tools used vital sign measures, most commonly blood pressure, heart rate, respiratory rate, temperature, oxygen saturation, and a level of consciousness measure (Multimedia Appendix 2, Tables S2 and S3 [23-39,41-45,47-61,63-65,67,68]). In addition, most of the models included basic patient demographic data, such as age and gender, as well as administrative measures, such as admission status, time since admission, length of stay, and patient location. Many of the models attempted to incorporate various pathology or laboratory test results where available, noting that they would experience some level of time delay. Composite indices or scores for severity of illness, longitudinal comorbidities, and combined laboratory results were often included in models where available. The more complex models considered higher-level features of their data such as the frequency, change, minimum, maximum, moving average, and patterns of physiological measures over time [23,24,38,42,45,56,58].

Of the studies that reported missing data, the majority were filled by propagating a previous value carried forward if a current value within a set time window was not measured. If no prior value was available, values were imputed with a representative value such as a population-based estimate or normal value. In models that consider dynamic irregularly sampled physiologic time series data, Gaussian process models were used to deal with sparsity in the data [23,24,55].

Overall, the quality of the studies included in this review was high, with low and unclear risks of bias and concern for the applicability of prediction models in addressing our review questions (Multimedia Appendix 2, Table S2). The majority of prediction model studies appropriately selected model data for inclusion, assessed model predictors and outcomes, approached model development, and adequately tested models. Studies assessed as unclear were because of ambiguous details in reporting the number of participants or samples per derivation and validation data sets, handling of missing data, and lack of detail in model validation. The study quality of the implementation studies we identified was also high, with low to moderate risk of bias identified (Multimedia Appendix 2, Table S4). Moderate concerns were because of the lack of adjustment for potential confounding, participant selection, and lack of detail for handling of missing data.

Table 3. Retrospective studies evaluating scoring tools.

Author, year, and country	Settings	Study aim	Scoring tools	Prediction event	Key findings
Lighthall et al, 2009 [67], United States	1089 patients, 1 hospital, 2006	To evaluate vital signs and association with critical events	MET[a] call criteria	CA[b], ICU[c] transfer or death	Even a single recording of an abnormal vital sign increases the risk of critical events in hospitalized patients.
Huh et al 2014, [41], Korea	3030 events, 1 hospital, 2008-2010	To evaluate the efficacy of screening triggered alerts for MET management	Medical alert system criteria	MET activation	The automatic alert system triggers, along with a skilled intervention team, were successful in managing the MET
Romeo-Brufau et al, 2014 [54], United States	34,898 patients, 2 hospitals, 2011	Comparative analysis of the performance of common EWS[d] methods and how they would function if automated	MEWS[e], SEWS[f], GMEWS[g], Worthing, ViEWS[h], NEWS[i]	Resuscitation call, RRS[j] activation or ICU transfer	The evaluated scores did not offer good predictive capabilities for an automated alarm system. Positive predictive values ranged from <0.01-0.21, and sensitivity ranged from 0.07-0.75.
Yu et al, 2014 [59], United States	328 cases, 328 controls, 1 hospital, 2009-2010	To compare the ability of 9 risk prediction scores in detecting clinical deterioration among non-ICU ward patients	SOFA[k], PIRO[l], ViEWS, SCS[m], MEDS[n], MEWS, SAPS II[o], REMS[p], APACHE II[q]	Critical care consult, ICU transfer or death	Prediction scores can be used to estimate a ward patient's risk of clinical deterioration, with good discriminatory ability comparable with that of existing track-and-trigger systems. 0-12 hours before clinical deterioration, 7 of 9 scores performed with acceptable discrimination (AUC[r]>0.70).
Wengerter et al, 2018 [57], United States	217 cases, 868 controls, 1 hospital, 2013-2015	To evaluate whether Rothman Index variability can predict RRT[s] activation in surgical patients	Rothman Index	RRT activation, mortality	Rothman Index variability predicted likelihood of RRT activation.
Bedoya et al, 2019 [26], United States	85,322 patients, 2 hospitals, 2014-2016	To determine the effectiveness of NEWS implementation on predicting and preventing patient deterioration	NEWS	ICU transfer or death	No change after implementing NEWS. At both academic and community hospitals, NEWS had poor performance characteristics and was generally ignored by nursing staff.
Heller et al [39], 2020, Germany	3827 patients, 2 wards, 2016-2017	To develop a prediction model for Code Blue, using EMR[t] data, and compare with MEWS	MEWS with paging functionality	CA or ICU transfer	The rate of CA and ICU transfers significantly decreased after implementing MEWS with paging functionality.

[a]MET: medical emergency team.

[b]CA: cardiac arrest.

[c]ICU: intensive care unit.

[d]EWS: early warning score.

[e]MEWS: Modified Early Warning Score.

[f]SEWS: Standardized Early Warning Score.

[g]GMEWS: Global Modified Early Warning Score.

[h]ViEWS: VitalPAC Early Warning Score.

[i]NEWS: National Early Warning Score.

[j]RRS: rapid response system.

[k]SOFA: Sequential Organ Failure Assessment.

[l]PIRO: Predisposition, Infection, Response, Organ, Dysfunction Score.

[m]SCS: simple clinical score.

[n]MEDS: Mortality in Emergency Department Sepsis.

[o]SAPS II: Simple Acute Physiology Score II.

[p]REMS: Rapid Emergency Medicine Score.

[q]APACHE II: Acute Physiology and Chronic Health Evaluation Score II.

[r]AUC: area under the receiver operating characteristic curve.

[s]RRT: rapid response team.

[t]EMR: electronic medical record.

Table 4. Studies of scoring tool implementation.

Author, year, and country	Study design	Setting	Study aim	Scoring tool	Intervention assessment	Implementation outcome	Key findings or conclusions
Bailey et al, 2013 [61], United States	PS[a]: RCT[b]	19,116 patients, 1 hospital, 2007-2011	To validate the EWS[c] in general medical wards and trial real time alerting.	Two-tiered EWS [38]	Comparison of alerts between intervention and control patients.	Among patients identified by EWS, there were no differences in proportion transferred to ICU[d] or died in the intervention group compared with control.	Alerts generated for patients meeting the threshold were highly specific for ICU transfer and death. However, sending real time alerts to the nurse manager did not improve event outcomes.
Evans et al, 2015 [64], United States	PS: observational study	6289 patients, 1 hospital, 2012-2013	To develop and evaluate a detection and alert system for monitoring patients every 5 min	MEWS[e] with pager alerts	Comparison of events from patients in 2 wards pre and post intervention	Ward A patients had more ICU transfers, MET[f] calls and greater LOS[g] but fewer deaths during intervention compared with preintervention. No differences were seen in ward B.	Implementation of the predictive model increased appropriate MET calls. Mortality decreased in the ward with older patients and multiple comorbidities, but not in the other ward.
Subbe et al, 2017, [65], United Kingdom	PS: before and after study	4402 patients, 1 hospital, 2014-2015	To assess the effect of a vital sign monitoring and alert system on outcomes	Vital sign monitoring system	Comparison of serious events between control and intervention periods	Deaths, CAs[h] and, for patients transferred to ICU, severity of illness scores were lower in the intervention compared with control.	Deploying an EWS based on vital signs increased RRT[i] calls and decreased mortality and CAs.
Oh et al, 2018 [52], Korea	RS[j]: before and after study	207,054 surgeries, 1 hospital, 2008-2016	To evaluate whether a RRS[k] reduces incidence of postoperative CA	RRS with thresholds and calling criteria	Change in cardiopulmonary arrest rate in patients before and after intervention	Cardiopulmonary arrest relative risk (pre vs post intervention) was 0.56 during RRS operational hours but was unchanged during nonoperational hours. These associations remained after comorbidity adjustment.	Implementation of the RRS reduced postoperative cardiopulmonary arrest incidence but only during RRS operational hours.
Morgan et al, 2020 [50], United States	PS: quality improvement study	30,292 patients, 1 hospital, 2017-2018	To evaluate the implementation of a continuous cloud-based EWS to activate an RRT.	Cloud-based modified NEWS[l] with RRT call threshold	Comparison of intervention with control patients for time to first lactate order, ICU[d] transfer and mortality.	The intervention group had improved the time to the first lactate order within 24 hours of modified NEWS ≥7. There was no significant improvement in time to ICU transfer, ICU length of stay, or hospital mortality.	The study provides preliminary evidence for a pragmatic integration of cloud-based, automated monitoring with standardized and timely RRT intervention.

[a]PS: prospective study.

[b]RCT: randomized controlled trial.

[c]EWS: early warning score.

[d]ICU: intensive care unit.

[e]MEWS: Modified Early Warning Score.

[f]MET: medical emergency team.

[g]LOS: length of stay.

[h]CA: cardiac arrest.

[i]RRT: rapid response team.

[j]RS: retrospective study.

[k]RRS, rapid response system.

[l]NEWS: National Early Warning Score.

Discussion

Principal Findings

This review examines the use of data sets collected in EMRs to develop and implement decision support for clinicians to predict and prevent inpatient deterioration. The current literature confirms that it is possible for routinely collected EMR data to be used to anticipate patient deterioration. However, there are few reports on the performance and efficacy of these systems when used in clinical settings. Despite the wide and increasing adoption of EMRs, the successful implementation of EMR-based early warning systems or their impact on patient-centered outcomes is not commonly reported.

The studies that met the eligibility criteria and the variability in methods created a narrative review rather than a quantitative review. There was considerable variation in the models developed, methodological approach, and data collected. The common study designs (retrospective, small cohorts, and before and after studies) pose a risk of bias. The institutions where models were developed and implemented predominately from the United States, which may pose additional risks of bias because of the nature of the health care system and environment. The heterogeneity in the methodology of the developed models makes comparison of their statistical accuracy and performance difficult. The area under the receiver operating characteristic curve statistic has been used as a comparison metric, but it is important to note the differences in model specification, complexity, and outcome when interpreting model output (eg, scores) and their usefulness for implementation in a real-time clinical setting. In addition, the reporting of superior performance in statistics from models of greater complexity may be the result of model overfitting if this has not been appropriately addressed in model development. Consideration should be made such that data used to develop early warning tools reflect previous cohorts of patients and may not be relevant for future cohorts. It is reasonable that retrospective patterns of deterioration will repeat themselves in the future; however, with advances in health care technology and the potential for new and emerging health disease trends, adapting or updating tools will be required to maintain relevance. We conclude that the effectiveness of current EMR-based digital early warning tools remains promising but has not been reproducibly demonstrated.

We propose that there are multiple factors that make a patient-centered outcome evaluation of EMR-based deterioration prediction difficult. Demonstration of improved patient outcomes following early warning tool implementation relies on the successful performance of all four components of implementation, recognition, escalation, and response. Where there has been successful statistical validation for the EWSs, validation of the escalation and response (ie, implementation) has been elusive. Several factors may contribute, which are not unique to digital warning implementation. The current metrics for outcome measures are problematic. Death, an outcome that is easy to count, does not always account for treatment limitations. Cardiac arrest events are rare; therefore, they are insensitive measures. ICU transfer can reflect MET resourcing deficiencies, ICU bed availability, and local admission practices or end-of-life planning rather than preventable deterioration. Emergency callout rates may reflect individual clinician sentiment rather than true risk or deterioration, or alternatively, not be called when they should be. There is inconsistency in efferent limb performance (assessment, intervention, transfer, monitoring, and follow-up). The time required for implementation and translation into improved outcomes will frequently be confounded by other system improvements, changes within an individual health service, or the maturity of the rapid response system itself. Therefore, the relationship between implementation and patient outcomes remains unstudied, rather than unproven by the current studies.

These issues will continue to challenge the empirical validation of a complete rapid response system. The lifecycle of these technologies has yet to mature to the point where evaluation and assessment of a possible effective intervention can be performed. One example of practical limitations is to recognize which data are contemporaneously available. Data can only inform a ward-based, real-time prediction model once the information is available in the EMR for analysis. This is relevant to model development and evaluation. For a pathology result, for example, there are times of sample collection, arrival at the laboratory or analysis, and the time the result is available for integration into the clinical picture. The results of the test are only available for integration once they are available in the clinical space. However, the metadata confirming that the sample was actually ordered or collected, where it was done, and the number of samples, is immediately available. The result of a positive blood culture can take hours to days, and a definitive negative result can only be finalized 2-5 days after collection. Providing actionable decision support with such dynamic data sets is a challenge, and it could be argued that the performance of the model on an experimental data set is irrelevant unless, once deployed to the EMR, the relevant decision support is actionable by clinicians in time to avoid an adverse outcome.

As maturity grows in the development cycle within EMRs, there will be potential opportunities to improve EMR-based deterioration tools and assess their impact on patient care. The development of algorithms can help monitor in real time rich clinical data from our patients, including vital signs, investigation results, drug prescription, and provide useful decision support to improve care trajectories. These clinical data can then be used at the patient level to provide visibility of deteriorating or at-risk patients to individual clinicians, wards, clinical teams, and the MET responders, and place decision support back into the EMR to support clinicians in accurately predicting and preventing inpatient deterioration. At the system level, the data could provide feedback to guide the optimization of the digital or clinician interaction and maximize response efficiency. Successful implementation should improve

patient-centered outcomes, reduce suffering, incapacity, and mortality, as well as reduce length of stay, increase hospital capacity, improve efficiency, and increase care delivery to meet ever-increasing demands.

One final, but highly important, factor for consideration in the future implementation of EMR-based deterioration tools is the aspect of medical device regulation. For example, with the recent introduction of regulations for software based medical devices in Australia by the Therapeutic Goods Administration, clinical decision support software such as algorithms to predict patient deterioration that meet the definition of a medical device would be subject to regulation and possibly inclusion in the Australian Register of Therapeutic Goods. The legislation also allows for clinical decision support software to be considered as an exempt medical device subject to how it is intended to be used. Either way, it is expected that regulation will lead to increased clinical acceptance and uptake of such algorithms.

Conclusions

The development and accuracy of digital EWSs is increasing, facilitated by the growing availability of digital data sets. However, despite the relative performance of algorithms that can predict patient deterioration, the current literature shows that limited deployment of such algorithms into clinical practice is associated with improvement in patient outcomes. There is a paucity of quality studies in this area, and further work is needed to explore potential clinical benefits, including optimizing of the digital or clinician interaction, consideration of limitations in implementation, such as the requirement for real-time data availability, and use of standardized measures.

Acknowledgments

The work presented in this manuscript was supported by Clinical Excellence Queensland, Queensland Health, and by the National Health and Medical Research Council Centre for Research Excellence in Digital Health (1134919). However, these sponsors had no influence on the study design; on the collection, analysis, and interpretation of data, on the writing of the manuscript; or decision to submit the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Studies of more complex patient deterioration models.
[DOCX File , 22 KB - jmir_v23i9e28209_app1.docx]

Multimedia Appendix 2
Full details of the papers included in this review, including information on prediction model input variables, performance metrics, and quality assessment.
[DOCX File , 49 KB - jmir_v23i9e28209_app2.docx]

References

1. Gao H, McDonnell A, Harrison DA, Moore T, Adam S, Daly K, et al. Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. Intensive Care Med 2007 Apr;33(4):667-679. [doi: 10.1007/s00134-007-0532-3] [Medline: 17318499]
2. McGaughey J, Alderdice F, Fowler R, Kapila A, Mayhew A, Moutray M. Outreach and Early Warning Systems (EWS) for the prevention of intensive care admission and death of critically ill adult patients on general hospital wards. Cochrane Database Syst Rev 2007 Jul 18(3):CD005529. [doi: 10.1002/14651858.CD005529.pub2] [Medline: 17636805]
3. Smith MB, Chiovaro JC, O'Neil M, Kansagara D, Quiñones AR, Freeman M, et al. Early warning system scores for clinical deterioration in hospitalized patients: a systematic review. Ann Am Thorac Soc 2014 Nov;11(9):1454-1465. [doi: 10.1513/AnnalsATS.201403-102OC] [Medline: 25296111]
4. Chapman SM, Wray J, Oulton K, Peters MJ. Systematic review of paediatric track and trigger systems for hospitalised children. Resuscitation 2016 Dec;109:87-109. [doi: 10.1016/j.resuscitation.2016.07.230] [Medline: 27496259]
5. Jacob N, Moriarty Y, Lloyd A, Mann M, Tume LN, Sefton G, et al. Optimising paediatric afferent component early warning systems: a hermeneutic systematic literature review and model development. BMJ Open 2019 Nov 14;9(11):e028796 [FREE Full text] [doi: 10.1136/bmjopen-2018-028796] [Medline: 31727645]
6. O'Connell A, Flabouris A, Thompson CH. Optimising the response to acute clinical deterioration: the role of observation and response charts. Intern Med J 2020 Jul;50(7):790-797. [doi: 10.1111/imj.14444] [Medline: 31389119]
7. Varghese J, Kleine M, Gessner SI, Sandmann S, Dugas M. Effects of computerized decision support system implementations on patient outcomes in inpatient care: a systematic review. J Am Med Inform Assoc 2018 May 01;25(5):593-602 [FREE Full text] [doi: 10.1093/jamia/ocx100] [Medline: 29036406]

8. Alam N, Hobbelink EL, van Tienhoven AJ, van de Ven PM, Jansma EP, Nanayakkara PW. The impact of the use of the Early Warning Score (EWS) on patient outcomes: a systematic review. Resuscitation 2014 May;85(5):587-594. [doi: 10.1016/j.resuscitation.2014.01.013] [Medline: 24467882]
9. Fang AH, Lim WT, Balakrishnan T. Early warning score validation methodologies and performance metrics: a systematic review. BMC Med Inform Decis Mak 2020 Jun 18;20(1):111 [FREE Full text] [doi: 10.1186/s12911-020-01144-8] [Medline: 32552702]
10. Gerry S, Bonnici T, Birks J, Kirtley S, Virdee PS, Watkinson PJ, et al. Early warning scores for detecting deterioration in adult hospital patients: systematic review and critical appraisal of methodology. Br Med J 2020 May 20;369:m1501 [FREE Full text] [doi: 10.1136/bmj.m1501] [Medline: 32434791]
11. Amland RC, Sutariya BB. Quick Sequential [Sepsis-Related] Organ Failure Assessment (qSOFA) and St. John sepsis surveillance agent to detect patients at risk of sepsis: an observational cohort study. Am J Med Qual 2018;33(1):50-57 [FREE Full text] [doi: 10.1177/1062860617692034] [Medline: 28693336]
12. Li L, Walter S, Rathnayake K, Westbrook J. Evaluation and Optimisation of Risk Identification Tools for the Early Detection of Sepsis in Adult Inpatients. Sydney: Australian Institute of Health Innovation, Macquarie University; 2018.
13. Delahanty RJ, Kaufman D, Jones SS. Development and evaluation of an automated machine learning algorithm for in-hospital mortality risk adjustment among critical care patients. Crit Care Med 2018 Jun;46(6):e481-e488. [doi: 10.1097/CCM.0000000000003011] [Medline: 29419557]
14. Malycha J, Bonnici T, Clifton DA, Ludbrook G, Young JD, Watkinson PJ. Patient centred variables with univariate associations with unplanned ICU admission: a systematic review. BMC Med Inform Decis Mak 2019 May 15;19(1):98 [FREE Full text] [doi: 10.1186/s12911-019-0820-1] [Medline: 31092256]
15. Centre for Reviews and Dissemination. Systematic Reviews: CRD's Guidance for Undertaking Reviews in Healthcare. York, England: University of York NHS Centre for Reviews & Dissemination; Dec 01, 2008.
16. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-269 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
17. Hausner E, Waffenschmidt S, Kaiser T, Simon M. Routine development of objectively derived search strategies. Syst Rev 2012 Feb 29;1:19 [FREE Full text] [doi: 10.1186/2046-4053-1-19] [Medline: 22587829]
18. Fatehi F, Gray LC, Wootton R. How to improve your PubMed/MEDLINE searches: 3. advanced searching, MeSH and My NCBI. J Telemed Telecare 2014 Mar;20(2):102-112. [doi: 10.1177/1357633X13519036] [Medline: 24614997]
19. EndNote: EndNote X9.: Clarivate URL: https://clarivate.libguides.com/endnote_training/users/enx9 [accessed 2021-08-10]
20. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
21. Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, PROBAST Group†. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019 Jan 01;170(1):51-58 [FREE Full text] [doi: 10.7326/M18-1376] [Medline: 30596875]
22. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. Br Med J 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
23. Alaa A, Van Der Schaar M. A hidden absorbing semi-Markov model for informatively censored temporal data: learning and inference. J Mach Learn Res 2018 Jan 1;19(1):108-169. [doi: 10.5555/3291125.3291129]
24. Alaa AM, Yoon J, Hu S, van der Schaar M. Personalized risk scoring for critical care prognosis using mixtures of gaussian processes. IEEE Trans Biomed Eng 2018 Jan;65(1):207-218. [doi: 10.1109/TBME.2017.2698602] [Medline: 28463183]
25. Bartkowiak B, Snyder AM, Benjamin A, Schneider A, Twu NM, Churpek MM, et al. Validating the Electronic Cardiac Arrest Risk Triage (eCART) score for risk stratification of surgical inpatients in the postoperative setting: retrospective cohort study. Ann Surg 2019 Jun;269(6):1059-1063 [FREE Full text] [doi: 10.1097/SLA.0000000000002665] [Medline: 31082902]
26. Bedoya AD, Clement ME, Phelan M, Steorts RC, O'Brien C, Goldstein BA. Minimal impact of implemented early warning score and best practice alert for patient deterioration. Crit Care Med 2019 Jan;47(1):49-55 [FREE Full text] [doi: 10.1097/CCM.0000000000003439] [Medline: 30247239]
27. Capan M, Ivy JS, Rohleder T, Hickman J, Huddleston JM. Individualizing and optimizing the use of early warning scores in acute medical care for deteriorating hospitalized patients. Resuscitation 2015 Aug;93:107-112 [FREE Full text] [doi: 10.1016/j.resuscitation.2014.12.032] [Medline: 25597507]
28. Cho KJ, Kwon O, Kwon JM, Lee Y, Park H, Jeon KH, et al. Detecting patient deterioration using artificial intelligence in a rapid response system. Crit Care Med 2020 Apr;48(4):285-289. [doi: 10.1097/CCM.0000000000004236] [Medline: 32205618]
29. Churpek MM, Yuen TC, Edelson DP. Predicting clinical deterioration in the hospital: the impact of outcome selection. Resuscitation 2013 May;84(5):564-568 [FREE Full text] [doi: 10.1016/j.resuscitation.2012.09.024] [Medline: 23022075]

30. Churpek MM, Yuen TC, Park SY, Gibbons R, Edelson DP. Using electronic health record data to develop and validate a prediction model for adverse outcomes in the wards*. Crit Care Med 2014 Apr;42(4):841-848 [FREE Full text] [doi: 10.1097/CCM.0000000000000038] [Medline: 24247472]
31. Churpek MM, Yuen TC, Park SY, Meltzer DO, Hall JB, Edelson DP. Derivation of a cardiac arrest prediction model using ward vital signs*. Crit Care Med 2012 Jul;40(7):2102-2108 [FREE Full text] [doi: 10.1097/CCM.0b013e318250aa5a] [Medline: 22584764]
32. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. Crit Care Med 2016 Feb;44(2):368-374 [FREE Full text] [doi: 10.1097/CCM.0000000000001571] [Medline: 26771782]
33. Churpek MM, Yuen TC, Winslow C, Robicsek AA, Meltzer DO, Gibbons RD, et al. Multicenter development and validation of a risk stratification tool for ward patients. Am J Respir Crit Care Med 2014 Sep 15;190(6):649-655 [FREE Full text] [doi: 10.1164/rccm.201406-1022OC] [Medline: 25089847]
34. Escobar GJ, Gardner MN, Greene JD, Draper D, Kipnis P. Risk-adjusting hospital mortality using a comprehensive electronic record in an integrated health care delivery system. Med Care 2013 May;51(5):446-453. [doi: 10.1097/MLR.0b013e3182881c8e] [Medline: 23579354]
35. Escobar GJ, LaGuardia JC, Turk BJ, Ragins A, Kipnis P, Draper D. Early detection of impending physiologic deterioration among patients who are not in intensive care: development of predictive models using data from an automated electronic medical record. J Hosp Med 2012;7(5):388-395. [doi: 10.1002/jhm.1929] [Medline: 22447632]
36. Fejza A, Geneves P, Layaida N, Bosson J. Scalable and interpretable predictive models for electronic health records. In: Proceedings of the IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA). 2018 Presented at: 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA); Oct 1-3, 2018; Turin, Italy URL: https://hal.inria.fr/hal-03124966 [doi: 10.1109/DSAA.2018.00045]
37. Green M, Lander H, Snyder A, Hudson P, Churpek M, Edelson D. Comparison of the between the flags calling criteria to the MEWS, NEWS and the electronic Cardiac Arrest Risk Triage (eCART) score for the identification of deteriorating ward patients. Resuscitation 2018 Feb;123:86-91 [FREE Full text] [doi: 10.1016/j.resuscitation.2017.10.028] [Medline: 29169912]
38. Hackmann G, Chen M, Chipara O, Lu C, Chen Y, Kollef M, et al. Toward a two-tier clinical warning system for hospitalized patients. AMIA Annu Symp Proc 2011;2011:511-519 [FREE Full text] [Medline: 22195105]
39. Heller AR, Mees ST, Lauterwald B, Reeps C, Koch T, Weitz J. Detection of deteriorating patients on surgical wards outside the ICU by an automated MEWS-based early warning system with paging functionality. Ann Surg 2020 Jan;271(1):100-105. [doi: 10.1097/SLA.0000000000002830] [Medline: 29771722]
40. Hu W, Chan CW, Zubizarreta JR, Escobar GJ. An examination of early transfers to the ICU based on a physiologic risk score. Manuf Serv Oper Manag 2018 Jul;20(3):531-549. [doi: 10.1287/msom.2017.0658]
41. Huh JW, Lim C, Koh Y, Lee J, Jung Y, Seo H, et al. Activation of a medical emergency team using an electronic medical recording-based screening system*. Crit Care Med 2014 Apr;42(4):801-808. [doi: 10.1097/CCM.0000000000000031] [Medline: 24335439]
42. Keim-Malpass J, Clark MT, Lake DE, Moorman JR. Towards development of alert thresholds for clinical deterioration using continuous predictive analytics monitoring. J Clin Monit Comput 2020 Aug;34(4):797-804. [doi: 10.1007/s10877-019-00361-5] [Medline: 31327101]
43. Kia A, Timsina P, Joshi HN, Klang E, Gupta RR, Freeman RM, et al. MEWS++: enhancing the prediction of clinical deterioration in admitted patients through a machine learning model. J Clin Med 2020 Jan 27;9(2):343 [FREE Full text] [doi: 10.3390/jcm9020343] [Medline: 32012659]
44. Kipnis P, Turk BJ, Wulf DA, LaGuardia JC, Liu V, Churpek MM, et al. Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the ICU. J Biomed Inform 2016 Dec;64:10-19 [FREE Full text] [doi: 10.1016/j.jbi.2016.09.013] [Medline: 27658885]
45. Kirkland LL, Malinchoc M, O'Byrne M, Benson JT, Kashiwagi DT, Burton MC, et al. A clinical deterioration prediction tool for internal medicine patients. Am J Med Qual 2013;28(2):135-142. [doi: 10.1177/1062860612450459] [Medline: 22822159]
46. Kwon J, Lee Y, Lee Y, Lee S, Park J. An algorithm based on deep learning for predicting in-hospital cardiac arrest. J Am Heart Assoc 2018 Jun 26;7(13):e008678 [FREE Full text] [doi: 10.1161/JAHA.118.008678] [Medline: 29945914]
47. Loekito E, Bailey J, Bellomo R, Hart GK, Hegarty C, Davey P, et al. Common laboratory tests predict imminent death in ward patients. Resuscitation 2013 Mar;84(3):280-285. [doi: 10.1016/j.resuscitation.2012.07.025] [Medline: 22863543]
48. Mayampurath A, Sanchez-Pinto LN, Carey KA, Venable L, Churpek M. Combining patient visual timelines with deep learning to predict mortality. PLoS One 2019 Jul 31;14(7):e0220640 [FREE Full text] [doi: 10.1371/journal.pone.0220640] [Medline: 31365580]
49. Mohamadlou H, Panchavati S, Calvert J, Lynn-Palevsky A, Le S, Allen A, et al. Multicenter validation of a machine-learning algorithm for 48-h all-cause mortality prediction. Health Informatics J 2020 Sep;26(3):1912-1925 [FREE Full text] [doi: 10.1177/1460458219894494] [Medline: 31884847]

50. Morgan CK, Amspoker AB, Howard C, Razjouyan J, Siddique M, D'Avignon S, et al. Continuous cloud-based early warning score surveillance to improve the safety of acutely ill hospitalized patients. J Healthc Qual 2021;43(1):59-66. [doi: 10.1097/JHQ.0000000000000272] [Medline: 32604130]
51. O'Brien C, Goldstein BA, Shen Y, Phelan M, Lambert C, Bedoya AD, et al. Development, implementation, and evaluation of an in-hospital optimized early warning score for patient deterioration. MDM Policy Pract 2020 Jan 10;5(1):2381468319899663 [FREE Full text] [doi: 10.1177/2381468319899663] [Medline: 31976373]
52. Oh TK, Kim S, Lee DS, Min H, Choi YY, Lee EY, et al. A rapid response system reduces the incidence of in-hospital postoperative cardiopulmonary arrest: a retrospective study. Can J Anaesth 2018 Dec;65(12):1303-1313. [doi: 10.1007/s12630-018-1200-5] [Medline: 30076577]
53. Redfern OC, Pimentel MA, Prytherch D, Meredith P, Clifton DA, Tarassenko L, et al. Predicting in-hospital mortality and unanticipated admissions to the intensive care unit using routinely collected blood tests and vital signs: development and validation of a multivariable model. Resuscitation 2018 Dec;133:75-81 [FREE Full text] [doi: 10.1016/j.resuscitation.2018.09.021] [Medline: 30253229]
54. Romero-Brufau S, Huddleston JM, Naessens JM, Johnson MG, Hickman J, Morlan BW, et al. Widely used track and trigger scores: are they ready for automation in practice? Resuscitation 2014 Apr;85(4):549-552 [FREE Full text] [doi: 10.1016/j.resuscitation.2013.12.017] [Medline: 24412159]
55. Shamout FE, Zhu T, Sharma P, Watkinson PJ, Clifton DA. Deep interpretable early warning system for the detection of clinical deterioration. IEEE J Biomed Health Inform 2020 Feb;24(2):437-446. [doi: 10.1109/JBHI.2019.2937803] [Medline: 31545746]
56. Somanchi S, Adhikari S, Lin A, Eneva E, Ghani R. Early prediction of cardiac arrest (code blue) using electronic medical records. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2015 Aug 10 Presented at: KDD '15: 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 10 - 13, 2015; Sydney NSW Australia p. 2119-2126. [doi: 10.1145/2783258.2788588]
57. Wengerter BC, Pei KY, Asuzu D, Davis KA. Rothman Index variability predicts clinical deterioration and rapid response activation. Am J Surg 2018 Jan;215(1):37-41. [doi: 10.1016/j.amjsurg.2017.07.031] [Medline: 28818297]
58. Wong J, Taljaard M, Forster AJ, Escobar GJ, van Walraven C. Addition of time-dependent covariates to a survival model significantly improved predictions for daily risk of hospital death. J Eval Clin Pract 2013 Apr;19(2):351-357. [doi: 10.1111/j.1365-2753.2012.01832.x] [Medline: 22409151]
59. Yu S, Leung S, Heo M, Soto GJ, Shah RT, Gunda S, et al. Comparison of risk prediction scoring systems for ward patients: a retrospective nested case-control study. Crit Care 2014 Jun 26;18(3):R132 [FREE Full text] [doi: 10.1186/cc13947] [Medline: 24970344]
60. Ye C, Wang O, Liu M, Zheng L, Xia M, Hao S, et al. A real-time early warning system for monitoring inpatient mortality risk: prospective study using electronic medical record data. J Med Internet Res 2019 Jul 05;21(7):e13719 [FREE Full text] [doi: 10.2196/13719] [Medline: 31278734]
61. Bailey TC, Chen Y, Mao Y, Lu C, Hackmann G, Micek ST, et al. A trial of a real-time alert for clinical deterioration in patients hospitalized on general medical wards. J Hosp Med 2013 May;8(5):236-242. [doi: 10.1002/jhm.2009] [Medline: 23440923]
62. Escobar GJ, Turk BJ, Ragins A, Ha J, Hoberman B, LeVine SM, et al. Piloting electronic medical record-based early detection of inpatient deterioration in community hospitals. J Hosp Med 2016 Nov;11 Suppl 1(Suppl 1):18-24 [FREE Full text] [doi: 10.1002/jhm.2652] [Medline: 27805795]
63. Escobar GJ, Liu VX, Schuler A, Lawson B, Greene JD, Kipnis P. Automated identification of adults at risk for in-hospital clinical deterioration. N Engl J Med 2020 Nov 12;383:1951-1960. [doi: 10.1056/nejmsa2001090]
64. Evans RS, Kuttler KG, Simpson KJ, Howe S, Crossno PF, Johnson KV, et al. Automated detection of physiologic deterioration in hospitalized patients. J Am Med Inform Assoc 2015 Mar;22(2):350-360 [FREE Full text] [doi: 10.1136/amiajnl-2014-002816] [Medline: 25164256]
65. Subbe CP, Duller B, Bellomo R. Effect of an automated notification system for deteriorating ward patients on clinical outcomes. Crit Care 2017 Mar 14;21(1):52 [FREE Full text] [doi: 10.1186/s13054-017-1635-z] [Medline: 28288655]
66. Kang MA, Churpek MM, Zadravecz FJ, Adhikari R, Twu NM, Edelson DP. Real-time risk prediction on the wards: a feasibility study. Crit Care Med 2016 Aug;44(8):1468-1473 [FREE Full text] [doi: 10.1097/CCM.0000000000001716] [Medline: 27075140]
67. Lighthall GK, Markar S, Hsiung R. Abnormal vital signs are associated with an increased risk for critical events in US veteran inpatients. Resuscitation 2009 Nov;80(11):1264-1269. [doi: 10.1016/j.resuscitation.2009.08.012] [Medline: 19744762]
68. Arnold J, Davis A, Fischhoff B, Yecies E, Grace J, Klobuka A, et al. Comparing the predictive ability of a commercial artificial intelligence early warning system with physician judgement for clinical deterioration in hospitalised general internal medicine patients: a prospective observational study. BMJ Open 2019 Oct 10;9(10):e032187 [FREE Full text] [doi: 10.1136/bmjopen-2019-032187] [Medline: 31601602]

69. Dummett BA, Adams C, Scruth E, Liu V, Guo M, Escobar GJ. Incorporating an early detection system into routine clinical practice in two community hospitals. J Hosp Med 2016 Nov;11 Suppl 1(Suppl 1):25-31 [FREE Full text] [doi: 10.1002/jhm.2661] [Medline: 27805798]
70. Paulson SS, Dummett BA, Green J, Scruth E, Reyes V, Escobar GJ. What do we do after the pilot is done? Implementation of a hospital early warning system at scale. Jt Comm J Qual Patient Saf 2020 Apr;46(4):207-216 [FREE Full text] [doi: 10.1016/j.jcjq.2020.01.003] [Medline: 32085952]

Abbreviations

EMR: electronic medical record
EWS: early warning score
ICU: intensive care unit
MeSH: Medical Subject Headings
MET: medical emergency team
PRISMA: Preferred Reporting in Systematic Reviews and Meta-Analyses
PROBAST: Prediction Model Risk of Bias Assessment Tool
ROBINS-I: Risk of Bias in Non-Randomised Intervention Studies

Edited by R Kukafka; submitted 24.02.21; peer-reviewed by V Liu, M Dugas, E Lee; comments to author 05.04.21; revised version received 14.07.21; accepted 27.07.21; published 30.09.21.

Please cite as:
Mann KD, Good NM, Fatehi F, Khanna S, Campbell V, Conway R, Sullivan C, Staib A, Joyce C, Cook D
Predicting Patient Deterioration: A Review of Tools in the Digital Hospital Setting
J Med Internet Res 2021;23(9):e28209
URL: https://www.jmir.org/2021/9/e28209
doi:10.2196/28209
PMID:34591017

Original Paper

Patients' Perspectives on the Usability of a Blended Approach to an Integrated Intervention for Patients With Medically Unexplained Physical Symptoms: Mixed Methods Study

Suze Adriana Johanna Toonders[1,2,3], MSc; Paula Elisabeth van Westrienen[1,2,3], PhD; Sophie Konings[1,4], MSc; Marianne E Nieboer[1], MSc; Cindy Veenhof[3,4,5], PhD; Martijn F Pisters[1,2,3,4], PhD

[1]Department of Health Innovation and Technology, Fontys University of Applied Sciences, Eindhoven, Netherlands

[2]Center for Physical Therapy Research and Innovation in Primary Care, Leidsche Rijn Julius Health Care Centers, Utrecht, Netherlands

[3]Physical Therapy Research Group, Department of Rehabilitation, Physical Therapy Science and Sport, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands

[4]Physical Therapy Sciences Program in Clinical Health Sciences, University Medical Center Utrecht, Utrecht, Netherlands

[5]Research Group Innovation of Human Movement Care, University of Applied Sciences Utrecht, Utrecht, Netherlands

Corresponding Author:
Suze Adriana Johanna Toonders, MSc
Department of Health Innovation and Technology
Fontys University of Applied Sciences
Dominee Theodor Fliednerstraat 2
Eindhoven, 5600 AH
Netherlands
Phone: 31 641621939
Email: s.a.j.toonders@umcutrecht.nl

Abstract

Background: Medically unexplained physical symptoms are physical symptoms, such as pain, fatigue, and dizziness, that persist for more than a few weeks and cannot be explained after adequate medical examination. Treatment for preventing the chronicity of symptoms is recommended. A promising approach is identifying patients who are at risk and subsequently offering a blended care intervention that focuses on promoting self-management while using eHealth as a supportive tool. When these interventions match with a patient's expectations, their effectiveness grows.

Objective: This study aimed to obtain more insights into usability from the patient perspective to improve future interventions.

Methods: A mixed methods design (ie, the use of qualitative and quantitative data) was used. Through semistructured interviews, in-depth insights were gained into patients' perspectives on usability. The analysis process was continuous and iterative. Data were synthesized and categorized into different themes. The System Usability Scale, which measures the usability of a system, was used to compare participants that found usability to be low, medium, or high. This study was approved by the Medical Ethical Committee Utrecht (approval number: 17-391/C).

Results: Saturation was reached after interviewing 13 participants. The following four themes emerged from the interviews: motivations and expectations prior to participating in the program, the applicability of e-coaching, the role of health care professionals, and the integrated design of the blended approach.

Conclusions: The successful implementation of integrated blended care interventions based on patients' perspectives requires matching treatments to patients' individual situations and motivations. Furthermore, personalizing the relative frequency of face-to-face appointments and e-coaching can improve usability.

(J Med Internet Res 2021;23(9):e19794) doi:10.2196/19794

KEYWORDS

usability; medically unexplained physical symptoms; blended care

Introduction

Medically unexplained physical symptoms (MUPS) are physical symptoms that persist for more than a few weeks and cannot be explained after adequate medical examination [1]. MUPS are a serious concern, since approximately 25% to 50% of symptoms remain unexplained in primary care [2,3]. Patients with MUPS experience symptoms such as pain, fatigue, and dizziness [4]. These symptoms often have a major impact on daily life and result in a high burden for patients with MUPS [5]. MUPS can be divided into the following three stages: mild, moderate, and chronic [6]. These stages are based on the frequency of consulting a general practitioner, the duration of symptoms, and the physical and psychological dysfunctions experienced [6]. Existing research on treatment for the chronic stages of MUPS has provided valuable insights, and recommended interventions have included cognitive behavioral therapy, exercise therapy, and neuroscience education [7]. Treatment for preventing the chronicity of symptoms has been recommended in order to reduce the severity of symptoms and the direct and indirect costs of care [8,9]. This is in line with the general trend in health care policy; policies nationwide aim to strengthen health programs to prevent diseases and address risk factors [10]. Health care is thereby changing its focus from cures and care to behavior and health [11].

In order for programs to succeed in shifting their focus to behavior and health, these programs must include proactive and indicated prevention [12]. A first step is identifying patients who are at risk for developing chronicity [13,14]. Moreover, literature has shown that programs and interventions should focus on promoting patients' self-management [15,16]. eHealth can serve as a supportive tool for both personalization and the promotion of self-management [17,18]. eHealth is not only supportive of usual therapeutic guidance but is also a substantial element of interventions as a whole [19]. This is referred to as *blended care*—the combination of face-to-face contact with integrated web-based applications [20]—or as *e-coaching*, which is defined as "the use of technology during coaching to motivate and stimulate (groups of) people to change attitudes, behaviors, and rituals" [21,22].

When these interventions match patients' expectations, sustainable changes in patients are achieved more effectively [23]. More insights into usability from the patient perspective can further improve these interventions [24,25]. For example, from the patient perspective, interventions should be easy to use and acceptable. This usability, which is defined as "the quality of a system with respect to ease of learning, ease of use and user satisfaction" [26], can be measured.

The objective of this study was to gain more understanding into patients' perspectives on the usability of integrated blended care interventions. In order to do so, this study analyzed a recent proactive, multidisciplinary, and integrated blended care intervention that was developed to prevent chronicity in patients with MUPS in primary care [27,28]. At-risk patients were identified by using electronic medical records [29]. e-Coaching was used to integrate technology into the intervention. The main goals were to (1) promote self-management among patients and (2) provide patients with insights into dealing with their complaints.

Methods

Study Design and Setting

A mixed methods design (ie, the use of qualitative and quantitative data) was used. Through semistructured interviews, qualitative data were gathered in order to gain an in-depth understanding of usability from patients' perspectives. System Usability Scale (SUS) scores (low, medium, and high) were compared to responses in the interviews, which allowed us to gain better insight into the relationship between identified themes from interviews and experienced usability. This study was approved by the Medical Ethical Committee of University Medical Center Utrecht (approval number: 17-391/C).

Participants

Patients who participated in the PARASOL intervention were eligible for inclusion. To be included in the intervention, all patients (aged ≥18 years) must have had ≥5 consultations with their general practitioner in the past 12 months. Of these consultations, ≥3 had to be classified as "suggestive of MUPS" based on 1 of the 104 International Classification of Primary Care codes. Patients with medical and psychiatric diagnoses were excluded [28]. Only participants in the PARASOL intervention who provided informed consent for this follow-up study were invited. In order to obtain rich data, stratified purposeful sampling was conducted based on the outcomes of the SUS. Patients with validated SUS scores of <70, between 70 and 80, and >80 were included; these represent low, medium, and high scores for usability, respectively [30].

Measurements

Qualitative data were collected in one-to-one semistructured interviews, which were conducted at an agreed-upon location. A second researcher was available to play the role of observer. The topic list for the interviews was based on the theoretical construct of De Bleser et al [26] and supplemented with the determinants of health care innovation that were selected and developed by the Netherlands Organization for Applied Scientific Research [31] (Textbox 1). The quantitative data consisted of the outcomes of the SUS. The SUS has high reliability [30] and contains 10 questions on the usability of a system [32]. Questions were answered on a numeric rating scale with scores that range from 1 to 5 ("strongly agree" to "strongly disagree"). The SUS was administered at the end of the intervention. The demographic data consisted of age, gender, and educational level (basic, intermediate, and high). Educational levels were derived from the Standard Classification of Education used by Statistics Netherlands [33].

Textbox 1. Outline of the interview guide [26]. The key areas are shown.

Performance

- Impact of use environment
- Impact of user characteristics
- Ease of the manipulation of the device

Satisfaction

- Physical dimension
- Privacy dimension
- Human interaction
- Self-concept
- Routine
- Sustainability

Acceptability

- Acceptance for daily life use
- Willingness to pay for device

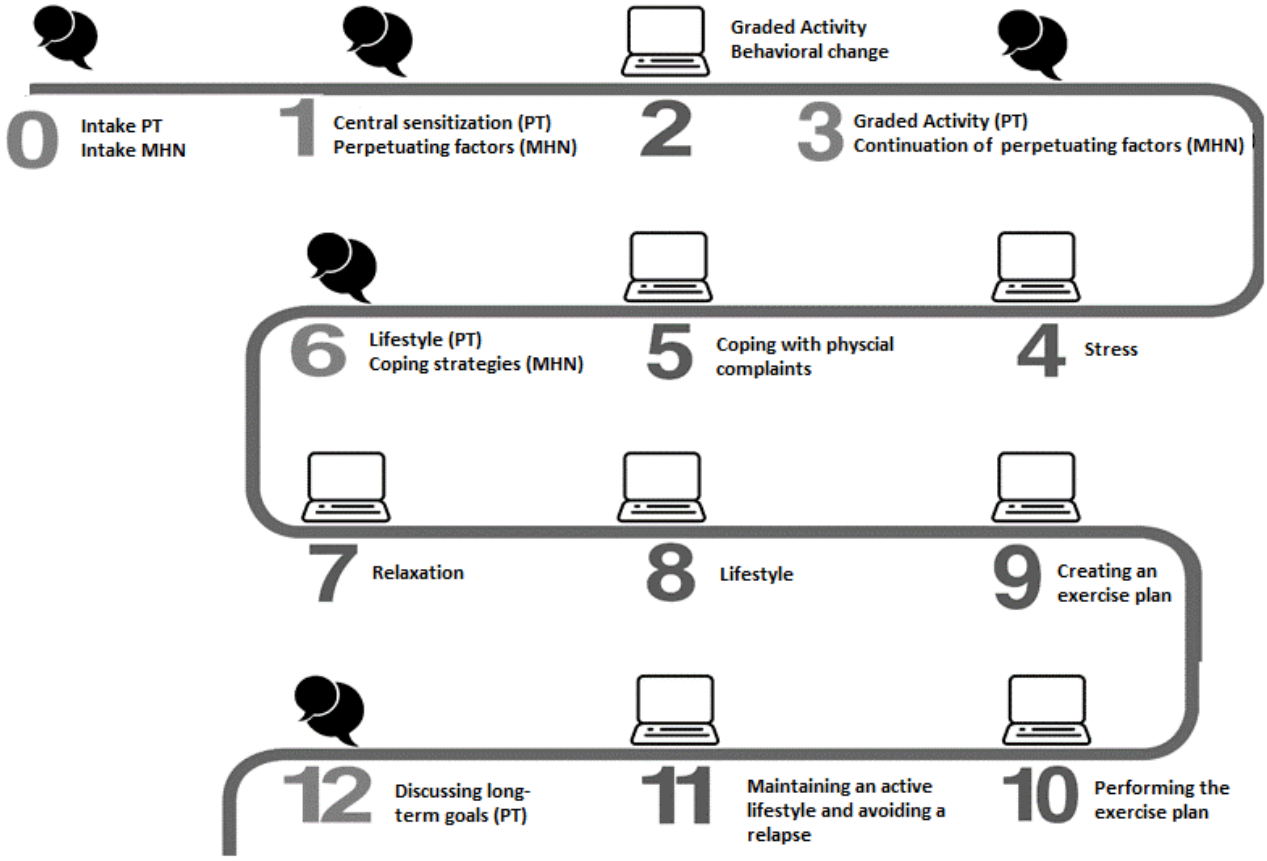
Procedure

Qualitative data were collected from semistructured interviews within 4 weeks after participants completed the PARASOL intervention to avoid recall bias. Interviews took place in patients' homes or in one of the participating health care centers, depending on the preferences of the patients. Before the interview started, procedures regarding sound recording and the coding of data were explained, after which permission was requested from the participants. Quantitative data were collected for the randomized controlled trial PARASOL (Evaluation of a Proactive Preventive Program in Patients With MUPS; trial registration number: NL57931.041.16) [28]. Demographic data were retrieved from baseline measurements. After 3 months, upon the completion of the PARASOL intervention, SUS scores were gathered.

PARASOL Intervention

The PARASOL intervention was a 12-week integrated blended care intervention that consisted of 4 face-to-face consultations with a mental health nurse and 5 physical therapy sessions and was supplemented with e-coaching (Figure 1). e-Coaching consisted of information modules and videos on self-management and educative themes, videos and instructions on prescribed home exercises, and assignments for gradually increasing physical activity. The intervention aimed to improve patients' perceptions of symptoms and identify modifiable risk factors of chronicity by providing therapeutic neuroscience education and promoting self-management. The intervention also aimed to promote an active lifestyle by using a cognitive behavioral approach and graded activities. Health care professionals were instructed on how to treat patients with moderate MUPS during a 2-day training session. Beyond the program itself, instructions included presentations on the study population, central sensitization, therapeutic neuroscience education, graded activities, and perpetuating factors. Furthermore, health care professionals were instructed on how to integrate e-coaching during the intervention. They were, for instance, guided on how to personalize general themes and instructed to ask patients about whether they understood information that is given on web-based platforms. All health care professionals received a guideline after finishing the training.

Figure 1. Overview of the PARASOL intervention. The text cloud images indicate face-to-face contact with a PT and MHN. The computer images indicate modules of e-coaching. MHN: mental health nurse; PT: physical therapist.



Data Analysis

Interviews were recorded and transcribed verbatim, and transcriptions were checked by 2 researchers. Within 1 week after completing the interviews, a summary was sent to all participants. This member check verified whether interpretations were correct. After the initial interviews were conducted, the interviewer added other questions based on the themes that emerged from these interviews. Both researchers encoded meaningful text fragments independently, and a set of preliminary concepts and codes was generated. The analysis process was continuous and iterative. Data were synthesized and categorized into 4 different themes. In the last stage of the analysis, for each theme, interview responses were compared on the basis of participants' SUS ratings. This allowed us to gain better insight into the relationship between identified themes from interviews and experienced usability.

Results

Interview Results

Saturation was reached after 13 interviews. Interviews lasted for approximately 20 to 50 minutes and had a mean duration of 33 minutes. Participants' mean age was 42 years. A majority of participants were female (10/13, 77%). Further, 5 participants had an SUS score of <70, 5 participants had a score of between 70 and 80, and 3 participants had a score of >80. The demographic characteristics of the study population can be found in Table 1.

Table 1. Demographic characteristics.

Participant number	Age (years)	Sex	Educational level	Previous experience in blended care	Interest in technology in the field of health care	System Usability Scale score
1	35	Female	Intermediate	No	Yes	67.5
2	48	Female	Intermediate	No	No	60.0
3	38	Female	Intermediate	No	Neutral	77.5
4	23	Female	Intermediate	No	Neutral	57.5
5	42	Female	Basic	No	Neutral	55.0
6	42	Male	Intermediate	No	Yes	50.0
7	48	Female	High	Yes	Yes	77.5
8	43	Female	Intermediate	No	Yes	85.0
9	47	Female	High	Yes	Yes	80.0
10	38	Male	High	No	No	72.5
11	31	Female	High	No	Yes	72.5
12	52	Male	High	No	Yes	87.5
13	57	Female	High	No	No	95.0

The 13 interviewees formed a subset of participants from the PARASOL intervention arm (n=80; age: mean 47 years; female: 57/80, 71%). The overall averages of the PARASOL intervention participants were hence comparable to those who were selected for interviews on the basis of purposeful sampling. The overall mean SUS score in the PARASOL intervention arm (n=55) was 74.6. A total of 19 participants had an SUS score of <70, 15 participants had an SUS score of between 70 and 80, and 21 participants had an SUS score of >80. Further, 20 participants in the PARASOL intervention did not complete the intervention, and 5 questionnaires were not submitted.

As the use of e-coaching integrated in treatment is relatively new, participants were asked about their general experience with and interest in technology in health care. Every participant had used some form of technology (in the broadest sense of the word). The use of a PC, smartphone, and tablet were mentioned. The integration of technology in health care was only previously experienced by 2 of the participants. When asked about technology in health care, participants mentioned the use of pedometers, health apps, and websites. Participants' interest in technology differed, as can be seen in Table 1.

A total of 4 themes emerged from the interviews. These themes provided insight into the usability of a blended approach to an integrated intervention from patients' perspectives.

Theme 1: Motivations and Expectations Prior to Participation in the Intervention

There was no consensus on participants' expectations prior to the intervention. Some participants stated that they had no expectations or that they had no expectations that their complaints would disappear by participating in the intervention. Others expected fewer complaints and more physical activity, and some expected that their pain would go away. A recurring statement reflected the hope that someone would seriously consider their complaints:

> *That someone finally thinks about the fact that these complaints are really there, and that a program is being made.* [Participant #3]

In terms of motivation, some participants participated mainly for personal interest. Other participants were just curious and saw no disadvantages, and some started the intervention because of a referral from their general practitioner. Experiencing intense pain was a motivation for participating in the intervention, and some participants mentioned that there were no other options for treatment with regard to their complaints. One participant stated:

> *I take this, because elsewhere a program is never really offered.* [Participant #8]

When the results were analyzed based on SUS score groups, they showed that higher overall SUS scores were related to quotes regarding autonomy and intrinsic motivation (Textbox 2). In terms of expectations related to the intervention program, there was no difference among SUS groups (Textbox 3).

Textbox 2. Quotes related to motivation. The quotes are stratified by System Usability Scale (SUS) score groups.

SUS score group: <70

- "I participate to stay active"
- "Advice from GP"

SUS score group: 70-80

- "Interesting to see whether the mental and physical aspects come together"
- "I don't understand my complaints and want to know what they are, and how I can deal with them"

SUS score group: >80

- "I have to make use of this opportunity, as I have been looking for ways to deal with my complaints for two years"
- "I had no way to resolve my complaints, and perhaps this will help me"

Textbox 3. Quotes related to expectations. The quotes are stratified by System Usability Scale (SUS) score groups.

SUS score group: <70

- "I don't know if it will work"
- "I have no idea what to expect"

SUS score group: 70-80

- "I'm curious, rather than have any expectations"
- "I thought, this must really work"

SUS score group: >80

- "I was open to something new"

Theme 2: Applicability of e-Coaching

References were made to e-coaching during interviews twice. The first reference concerned the look and feel of the application, and the second concerned the application's acceptability. Some participants mentioned that they spent a long time searching within the application and found the web-based portion to be confusing. For example:

> *I had to watch instruction videos but I could not find them.* [Participant #3]

Other participants however found the site to be well structured. There was no consensus on the ease with which documents or instruction videos could be found. Many participants had problems with logging in. In addition, the application often had bugs. This did not promote the use of e-coaching. One participant said:

> *I did my exercises every day but the program did not work so I just did not fill it in.* [Participant #13]

Another participant missed an evaluation that would have given insight into their progress. The ability to ask questions on web-based platforms and the fact that people can use the intervention anywhere were mentioned as facilitators. Participants stated that the planning assignments and exercises were clear every week. One participant said:

> *What I found very clear was that you could just click and do your exercises and activities on a weekly and daily basis.* [Participant #8]

Participants appreciated the ability to tick off the followed modules so that it was immediately clear which modules had been completed and which were still open. There was no consensus on whether obtaining information through text or film was preferred. Participants gave the following tips for the use of e-coaching:

> *Add forms on the site to leave notes on progress, e.g. how many minutes one walked.* [Participant #9 and Participant #11]
>
> *Make assignments more accessible by using visual support (colors, shapes).* [Participant #7]

The higher the satisfaction (as measured by the SUS), the more participants understood and used the web-based environment (Textbox 4).

Textbox 4. Quotes related to the applicability of e-coaching. The quotes are stratified by System Usability Scale (SUS) score groups.

SUS score group: <70

- "It is unclear for me how to use the website"
- "I can't enter the system, I never accessed the online part"
- "I often did not fill out the online sections, I prefer face-to-face treatments"

SUS score group: 70-80

- "I could not find the video, so I used text"
- "Clear and easy to use"

SUS score group: >80

- "The videos are clear and easy to use in daily life"
- "The site was clear"
- "It was easy to get the hang of the application"

Theme 3: The Role of Health Care Professionals

An often-mentioned facilitator of the treatment was agreement among health care professionals. Participants felt that they were receiving the same information from different angles. In face-to-face treatments, which participants felt to be useful, health care professionals provided psychoeducation, in which reminders and repetition were introduced to patients. A participant stated:

> *Because both the mental health nurse and the physical therapist spoke about interpreting pain, for example, and the physical therapist explains it more anatomically.* [Participant #11]

The important roles of health care professionals were found to be discussing exercises, providing information, setting goals, and helping patients reach these goals. Participants also appreciated the fact that health care professionals supported reflections on behaviors and thoughts via confrontation, convincement, and motivation. For example:

> *Holding up a mirror to me, that there was a confrontation, it was very helpful that the physical therapist was confrontational.* [Participant #12]

Another facilitator was the approachability of the mental health nurse. Participants recommended increasing the involvement of the general practitioner to increase the amount of feedback and encouragement that they receive. One participant stated:

> *I can imagine that people with these complaints do not always immediately think the mental health nurse and the physical therapist are going to solve the problem, so I think that the GP is still important for encouragement.* [Participant #3]

Participants also did not expect physical therapists to engage in conversations as much as they did:

> *I think physical therapy is important only when giving exercises and not for conversations.* [Participant #10]

The higher the SUS score, the more patients understood that health care professionals acted as coaches rather than as therapists (Textbox 5). There was no difference among subgroups with regard to interprofessional collaboration (Textbox 6).

Textbox 5. Quotes related to the role of professionals. The quotes are stratified by System Usability Scale (SUS) score groups.

SUS score group: <70

- "I feel the need to have my own say more"
- "Sometimes I feel I have the same conversation twice, the physical therapist and I were a better match and we could converse more easily"

SUS score group: 70-80

- "The physical therapist remember me and my story, and that made me feel good"
- "I expected more from the physical therapist, just conversing and no exercises"

SUS score group: >80

- "The professionals were very involved"
- "It's good that the professionals held up a mirror to me"

Textbox 6. Quotes related to interprofessional collaboration. The quotes are stratified by System Usability Scale (SUS) score groups.

SUS score group: <70

- "Good cooperation, same advice"
- "The same advices, did not notice cooperation, I did know they coordinated amongst the two of them"

SUS score group: 70-80

- "The combination of the mental health nurse and the physical therapist was good"
- "There was an overlap, but that did not bother me, it was complementary"

SUS score group: >80

- "I know they coordinated, they did not enter each other's domains"
- "One was more physical, the other was more psychological"

Theme 4: Integrated Design of the Blended Approach

Given that only 2 participants had previous experience with blended care, interview questions about this new method of delivering health care were asked. Some participants were satisfied with the higher frequency of face-to-face appointments at the start of the intervention, while others were not. The time between appointments increases the chance of forgetting parts of the treatments. The face-to-face sessions served as a reminder:

> *Because I forget a lot, so it's nice that I can have feedback reminder.* [Participant #5]

Participants suggested making the number of face-to-face sessions dependent on individual preferences. One participant said:

> *I think you should personally consult with each individual on the number of appointments.* [Participant #13]

Others indicated that the number of face-to-face appointments should be made dependent on one's experience with web-based applications. For example:

> *I think for me personally I could have done with fewer appointments, as I am used to work online.* [Participant #8]

Participants also mentioned that it was important for face-to face sessions and e-coaching to be coordinated. One participant stated:

> *You are encouraged to do the online program and then you come to practice and can get the information again, it connects.* [Participant #7]

Another stated that face-to-face sessions filled the gap that was left on web-based platforms:

> *In fact, I first had to read the explanation on the website and then my questions were discussed.* [Participant #10]

The possibility to schedule therapy based on personal preferences however was seen as an advantage. For example:

> *I liked the times. It was possible for me to make an appointment at the end of the day.* [Participant #7]

The advantage of e-coaching was that participants could prepare specific questions that could be asked during the face-to-face sessions (eg, "I could ask specific questions I prepared myself" [Participant #10]). Further, participants generally perceived blended care as positive (eg, "But that you can check it yourself at home. I think this is very good" [Participant #7]).

Participants appreciated the integrated design of the intervention across all of the different SUS score groups (Textbox 7).

Textbox 7. Quotes related to the integrated design of the blended approach. The quotes are stratified by System Usability Scale (SUS) score groups.

SUS score group: < 70

- "Because feedback is more specific for my own situation"
- "Face-to-face was a reminder...I find personal contact to be very important"

SUS score group: 70-80

- "The proportion [face-to-face and online] and frequency was good"
- "Face-to-face and online matched"
- "Repetition made it easier to remember"

SUS score group: >80

- "I find it easy to combine with other activities, I could do with less appointments"
- "The number of appointments should be based on personal preferences"

Overall, the results of this study show that participants experienced the intervention positively. This integrated blended care intervention aimed to promote self-management among patients and provide patients with insights into dealing with their complaints. Participants stated that they learned about self-management:

> *Now, I can estimate what I can do and cannot do.* [Participant #9]
>
> *I can actually do it all by myself.* [Participant #8]

Participants also gained more insights into dealing with their complaints:

> *Knowing nothing is broken, that idea has reassured me.* [Participant #4]
>
> *Because of graded activity, pain turns into pride; I am happier, undertake more, sing more; I'm enjoying more.* [Participant #11]

Textbox 8 includes all of the core themes that emerged from the semistructured interviews and hence summarizes usability from patients' perspectives. It shows the factors that were appreciated and lessons learned for improving usability.

Textbox 8. Summary of findings.

Factors that patients appreciated

- Information being recognizable
- The intervention as an incentive
- The personal approach
- The holistic approach
- Interprofessional collaboration

Lessons learned for improving usability

- Connect the intervention to the individual's situation and motivation
- Improve the accessibility of and technology support in e-coaching
- Introduce the possibility of asking questions on web-based platforms
- Personalize the intervention with respect to the amount of personal guidance alongside e-coaching

Discussion

Principal Findings

In this study, we evaluated patients' perspectives on the usability of an integrated blended care intervention. All included patients participated in a 12-week proactive blended care intervention in primary care with the aim of preventing the chronicity of MUPS. Participants were all generally positive about the received care. Various aspects of usability were highlighted, and responses were categorized into 4 themes.

The first theme that arose from interviews was the motivations and expectations of patients prior to the intervention. Existing literature shows that interventions that match patients' expectations are more effective in achieving sustainable changes in patients [23]. This especially holds true for intrinsic motivation rather than extrinsic motivation, which increases one's willingness to spend more time on assignments [34] and results in better health care outcomes [35]. Motivation also seems to be a factor of patients' adherence to eHealth [36]. In this study, we found differences in motivation related to satisfaction. When the overall results of the interviews were compared based on SUS scores, intrinsic motivation seemed to be an important factor related to experienced usability. Another factor that may influence a patient's motivation is patient selection. In this study, an electronic screening method involving the use of data from the electronic medical records of general practitioners' patients was used [29]. All eligible patients who were at risk for the chronicity of complaints were proactively approached by their general practitioners via an invitation letter. The selection of patients via this approach also has implications for patients' motivation, as the chance of approaching patients who may be less motivated may increase. To achieve adherence in patients, one should therefore take motivation into account in future interventions.

Many participants were not satisfied with the technical support provided in e-coaching, as technical functions did not work and logging in was a problem. The degree of satisfaction, which was measured with the SUS, increases when the web-based environment is understood and can be used. When patients were uncertain about the usefulness of e-coaching, the e-coaching modules were not used. This phenomenon has also been found in literature. Adapting eHealth to users' understanding and capabilities leads to a more usable and useful system [23]. When comparing the ages and educational levels of the participants in the low and high SUS score groups, a finding that stood out was that those with lower satisfaction were substantially younger and had lower educational levels. Existing literature shows that individuals with less education have worse actual and self-rated skills for evaluating the quality of web-based health information and lower trust in web-based health information compared to those with more education [37]. Studies however have found no consensus regarding the relationship between satisfaction and age [37].

Irrespective of the differences in satisfaction with e-coaching, participants were satisfied with the interprofessional collaboration. The holistic approach, through which physical therapists and mental health nurses provided information from different angles, was positively received by the participants. The expectations of participants regarding the role of health care professionals however differed among the SUS score groups. The higher the SUS score, the more patients understood that health care professionals acted as coaches rather than as therapists. Participants in the lower SUS score group, for instance, felt that they had to explain their complaints twice and expected that the roles of physical therapists would include more than just engaging in conversations and providing exercises. As the organization of health care has changed (ie, focusing more on prevention) [38], the role of health care professionals will also change; health care professionals will shift their focus from being a therapist to being more of a coach [39]. It seems important to explain this new role at the start of integrated blended care interventions in order to better shape the expectations of patients. Aside from interprofessional collaboration, attention should also be given to the collaboration between professionals and patients. Shared decision-making can support this process [40].

Participants appreciated the integrated design of the intervention across all of the different SUS score groups. They positively evaluated the possibility of saving texts and videos for future reference and the repetition of information in e-coaching combined with face-to-face sessions. The ability to personalize face-to-face sessions by allowing patients to prepare specific questions after studying the general information in the e-coaching modules was appreciated. Earlier studies have underlined the importance of face-to-face treatment combined with web-based care, as this has been found to improve and preserve outcomes [35,36,41]. The extent to which the intervention was tailored to participants made interventions and information recognizable. Participants also mentioned that an important yet missing part of the intervention was a diary or a free space for taking notes on exercises. The option to tick off exercises and modules and the explanation of exercises were considered to be helpful. These findings are supported by literature stating that the key components of the positive effect that eHealth has on health outcomes are personalization, stimulation, goal setting, and the integration of e-coaching [21]. All of these elements were available in the integrated blended care intervention.

The results of this study demonstrate the usability of an integrated blended care program for patients with MUPS. More research is needed to investigate whether these results are patient specific or whether the results of this patient population are unique. What remains important is ensuring that the use of technology in treatment fits the participants [42]. A checklist can help health care professionals, together with patients, to decide whether a patient is eligible for this program and whether the program matches a patient's characteristics (eg, abilities, needs, and preferences) and prior experiences with blended care [41].

Strengths and Limitations

A limitation of this qualitative study is that all information is based on a specific integrated blended care intervention—the PARASOL intervention. Therefore, some items of the core themes are directly linked to this specific intervention. However, recommendations are insightful in general when starting an integrated intervention with a blended approach.

The theoretical construct of Bleser et al [26] was chosen. This construct contains the performance, satisfaction, and acceptability features. Other theoretical constructs for gathering insights into usability also exist, such as the Unified Theory of Acceptance and Use of Technology and the Technology Acceptance Model. These other constructs however largely overlap [43,44]. The Unified Theory of Acceptance and Use of Technology focuses more on social influences related to behavioral intention, whereas the Technology Acceptance Model focuses on perceived usefulness and ease of use. Given the findings of this study, including other measuring instruments, such as the Intrinsic Motivation Inventory and the Rotter locus of control scale, could be an interesting addition in future research. These could shed more light on patients' motivations at the start of the program. The strengths of this study are the use of an iterative process during the analysis of the results and the use of triangulation methods during the whole research process. Furthermore, patient involvement was sought in all research phases.

Conclusions

The successful implementation of integrated blended care interventions based on patients' perspectives requires matching treatments to patients' individual situations and motivations. In addition, personalizing the relative frequency of face-to-face appointments and e-coaching is of importance.

Acknowledgments

This work was supported by SIA-RAAK-public (grant 2015-02-24P).

Authors' Contributions

SAJT, PEVW, MFP, and CV initiated this study and contributed to the concept and design of this study. SAJT organized the interviews and participated as an interviewer, and PEVW and SK participated as observers. SAJT and SK analyzed and interpreted the data. MEN was the qualitative interview expert. All authors revised the manuscript and approved the final version for submission.

Conflicts of Interest

None declared.

References

1. Hartman TCO, Blankenstein AH, Molenaar AO, van den Berg DB, van der Horst HE, Arnold IA, et al. NHG Guideline on medically unexplained symptoms (MUS). Huisarts Wet 2013;56(5):222-230.
2. Kroenke K, Jackson JL. Outcome in general medical patients presenting with common symptoms: a prospective study with a 2-week and a 3-month follow-up. Fam Pract 1998 Oct;15(5):398-403. [doi: 10.1093/fampra/15.5.398] [Medline: 9848423]
3. Jackson JL, Passamonti M. The outcomes among patients presenting in primary care with a physical symptom at 5 years. J Gen Intern Med 2005 Nov;20(11):1032-1037 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.0241.x] [Medline: 16307629]
4. Verhaak PFM, Meijer SA, Visser AP, Wolters G. Persistent presentation of medically unexplained symptoms in general practice. Fam Pract 2006 Aug;23(4):414-420. [doi: 10.1093/fampra/cml016] [Medline: 16632487]
5. Zonneveld LNL, Sprangers MAG, Kooiman CG, van 't Spijker A, Busschbach JJV. Patients with unexplained physical symptoms have poorer quality of life and higher costs than other patient groups: a cross-sectional study on burden. BMC Health Serv Res 2013 Dec 17;13(1):520 [FREE Full text] [doi: 10.1186/1472-6963-13-520] [Medline: 24344899]
6. Hartman TCO, Borghuis MS, Lucassen PLBJ, van de Laar FA, Speckens AE, van Weel C. Medically unexplained symptoms, somatisation disorder and hypochondriasis: course and prognosis. A systematic review. J Psychosom Res 2009 May;66(5):363-377. [doi: 10.1016/j.jpsychores.2008.09.018] [Medline: 19379952]
7. van Dessel N, den Boeft M, van der Wouden JC, Kleinstäuber M, Leone SS, Terluin B, et al. Non-pharmacological interventions for somatoform disorders and medically unexplained physical symptoms (MUPS) in adults. Cochrane Database Syst Rev 2014 Nov 01(11):CD011142. [doi: 10.1002/14651858.CD011142.pub2] [Medline: 25362239]
8. Konnopka A, Schaefert R, Heinrich S, Kaufmann C, Luppa M, Herzog W, et al. Economics of medically unexplained symptoms: a systematic review of the literature. Psychother Psychosom 2012;81(5):265-275. [doi: 10.1159/000337349] [Medline: 22832397]
9. Khan AA, Khan A, Harezlak J, Tu W, Kroenke K. Somatic symptoms in primary care: etiology and outcome. Psychosomatics 2003;44(6):471-478. [doi: 10.1176/appi.psy.44.6.471] [Medline: 14597681]
10. Time to deliver in Europe. Meeting noncommunicable disease targets to achieve the Sustainable Development Goals. World Health Organization Europe. 2019 Apr. URL: https://www.euro.who.int/__data/assets/pdf_file/0006/413259/WHO-TKM-Outcome-Report-WEB.pdf [accessed 2021-09-07]
11. van der Kamp J, Plochg T. The health system quartet: Four basic systems—cure, care, heal and deal—to foster the co-production of sustained health. In: Embracing Complexity in Health. Cham, Switzerland: Springer; 2019:113-123.
12. Musich S, Wang S, Hawkins K, Klemes A. The impact of personalized preventive care on health care quality, utilization, and expenditures. Popul Health Manag 2016 Dec;19(6):389-397 [FREE Full text] [doi: 10.1089/pop.2015.0171] [Medline: 26871762]
13. Neuwirth EEB, Schmittdiel JA, Tallman K, Bellows J. Understanding panel management: a comparative study of an emerging approach to population care. Perm J 2007;11(3):12-20 [FREE Full text] [doi: 10.7812/tpp/07-040] [Medline: 21461107]
14. Boersma K, Linton SJ. Screening to identify patients at risk: profiles of psychological risk factors for early intervention. Clin J Pain 2005;21(1):38-43. [doi: 10.1097/00002508-200501000-00005] [Medline: 15599130]
15. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0]
16. Lorig KR, Sobel DS, Ritter PL, Laurent D, Hobbs M. Effect of a self-management program on patients with chronic disease. Eff Clin Pract 2001;4(6):256-262. [Medline: 11769298]
17. Alpay L, van der Boog P, Dumaij A. An empowerment-based approach to developing innovative e-health tools for self-management. Health Informatics J 2011 Dec;17(4):247-255 [FREE Full text] [doi: 10.1177/1460458211420089] [Medline: 22193825]
18. van der Vaart R, Witting M, Riper H, Kooistra L, Bohlmeijer ET, van Gemert-Pijnen LJEWC. Blending online therapy into regular face-to-face therapy for depression: content, ratio and preconditions according to patients and therapists using a Delphi study. BMC Psychiatry 2014 Dec 14;14(1):355 [FREE Full text] [doi: 10.1186/s12888-014-0355-z] [Medline: 25496393]
19. Kloek C, Bossen D, de Bakker DH, Veenhof C, Dekker J. Blended interventions to change behavior in patients with chronic somatic disorders: Systematic review. J Med Internet Res 2017 Dec 21;19(12):e418 [FREE Full text] [doi: 10.2196/jmir.8108] [Medline: 29269338]
20. Erbe D, Eichert HC, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: Systematic review. J Med Internet Res 2017 Sep 15;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]

21. Lentferink AJ, Oldenhuis HK, de Groot M, Polstra L, Velthuijsen H, van Gemert-Pijnen JE. Key components in eHealth interventions combining self-tracking and persuasive eCoaching to promote a healthier lifestyle: A scoping review. J Med Internet Res 2017 Aug 01;19(8):e277 [FREE Full text] [doi: 10.2196/jmir.7288] [Medline: 28765103]
22. van Gemert-Pijnen JEWC, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
23. Kayser L, Kushniruk A, Osborne RH, Norgaard O, Turner P. Enhancing the effectiveness of consumer-focused health information technology systems through eHealth literacy: A framework for understanding users' needs. JMIR Hum Factors 2015 May 20;2(1):e9 [FREE Full text] [doi: 10.2196/humanfactors.3696] [Medline: 27025228]
24. Rosson MB, Carroll JM. Usability engineering. Scenario-based development of human-computer interaction. In: The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications. Blacksburg, Virginia: Lawrence Erlbaum Associates; 2002:1032-1050.
25. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
26. de Bleser L, de Geest S, Vincke B, Ruppar T, Vanhaecke J, Dobbels F. How to test electronic adherence monitoring devices for use in daily life: a conceptual framework. Comput Inform Nurs 2011 Sep;29(9):489-495. [doi: 10.1097/NCN.0b013e31821a1555] [Medline: 21532468]
27. van Westrienen PE, Pisters MF, Gerrits M, Veenhof C, de Wit NJ. Identifying treatment modalities for a multidisciplinary and blended care intervention for patients with moderate medically unexplained physical symptoms: Qualitative study among professionals. JMIR Ment Health 2019 Apr 12;6(4):e12203 [FREE Full text] [doi: 10.2196/12203] [Medline: 30977737]
28. van Westrienen PE, Pisters MF, Toonders SA, Gerrits M, Veenhof C, de Wit NJ. Effectiveness of a blended multidisciplinary intervention for patients with moderate medically unexplained physical symptoms (PARASOL): Protocol for a cluster randomized clinical trial. JMIR Res Protoc 2018 May 08;7(5):e120 [FREE Full text] [doi: 10.2196/resprot.9404] [Medline: 29739735]
29. van Westrienen PE, Pisters MF, Veenhof C, de Wit NJ. Identification of patients with moderate medically unexplained physical symptoms in primary care with a five years follow-up. BMC Fam Pract 2019 May 21;20(1):66 [FREE Full text] [doi: 10.1186/s12875-019-0950-7] [Medline: 31113381]
30. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
31. Fleuren MAH, Paulussen TGWM, Van Dommelen P, Van Buuren S. Towards a measurement instrument for determinants of innovations. Int J Qual Health Care 2014 Oct;26(5):501-510 [FREE Full text] [doi: 10.1093/intqhc/mzu060] [Medline: 24951511]
32. Brooke J. SUS: A 'quick and dirty' usability scale. In: Usability Evaluation In Industry. Boca Raton, Florida, Verenigde Staten: CRC Press; 1996.
33. Standaard onderwijsindeling Editie 2016/ 17. Centraal Bureau voor de Statistiek. URL: https://docplayer.nl/47234471-Standaard-onderwijsindeling-editie-2016-17.html [accessed 2021-09-09]
34. Vallerand RJ. Deci and Ryan's Self-Determination Theory: A view from the Hierarchical Model of Intrinsic and Extrinsic Motivation. Psychol Inq 2000;11(4):312-318.
35. Michalak J, Klappheck MA, Kosfelder J. Personal goals of psychotherapy patients: The intensity and the "why" of goal-motivated behavior and their implications for the therapeutic process. Psychother Res 2004;14(2):193-209. [doi: 10.1093/ptr/kph017]
36. Talboom-Kamp EPWA, Verdijk NA, Kasteleyn MJ, Numans ME, Chavannes NH. From chronic disease management to person-centered eHealth; a review on the necessity for blended care. Clinical eHealth 2018 Mar;1(1):3-7 [FREE Full text] [doi: 10.1016/j.ceh.2018.01.001]
37. Diviani N, van den Putte B, Giani S, van Weert JC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015 May 07;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
38. Hibbard JH. Moving toward a more patient-centered health care delivery system. Health Aff (Millwood) 2004;Suppl Variation:VAR133-VAR135. [doi: 10.1377/hlthaff.var.133] [Medline: 15471780]
39. Wouters M, Swinkels I, van Lettow B, de Jong J, Sinnige J, Brabers A, et al. E-health in verschillende snelheden: eHealth-monitor 2018. Nivel. 2018. URL: https://www.nivel.nl/nl/publicatie/e-health-verschillende-snelheden-ehealth-monitor-2018 [accessed 2021-09-09]
40. Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: a model for clinical practice. J Gen Intern Med 2012 Oct;27(10):1361-1367 [FREE Full text] [doi: 10.1007/s11606-012-2077-6] [Medline: 22618581]

41. Wentzel J, van der Vaart R, Bohlmeijer ET, van Gemert-Pijnen JEWC. Mixing online and face-to-face therapy: How to benefit from blended care in mental health care. JMIR Ment Health 2016 Feb 09;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]
42. Kloek CJJ, Janssen J, Veenhof C. Development of a checklist to assist physiotherapists in determination of patients' suitability for a blended treatment. Telemed J E Health 2020 Aug;26(8):1051-1065. [doi: 10.1089/tmj.2019.0143] [Medline: 31804904]
43. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Q 2012;36(1):157-178. [doi: 10.2307/41410412]
44. Venkatesh V, Davis FD. A theoretical extension of the Technology Acceptance Model: Four longitudinal field studies. Manage Sci 2000 Feb 01;46(2):169-332. [doi: 10.1287/mnsc.46.2.186.11926]

Abbreviations

MUPS: medically unexplained physical symptoms
SUS: System Usability Scale

Edited by G Eysenbach; submitted 01.05.20; peer-reviewed by O Blanson Henkemans, J Brooke; comments to author 29.06.20; revised version received 17.09.20; accepted 28.10.20; published 28.09.21.

Please cite as:
Toonders SAJ, van Westrienen PE, Konings S, Nieboer ME, Veenhof C, Pisters MF
Patients' Perspectives on the Usability of a Blended Approach to an Integrated Intervention for Patients With Medically Unexplained Physical Symptoms: Mixed Methods Study
J Med Internet Res 2021;23(9):e19794
URL: https://www.jmir.org/2021/9/e19794
doi:10.2196/19794
PMID:34581674

Original Paper

TikTok as a Health Information Source: Assessment of the Quality of Information in Diabetes-Related Videos

Wenwen Kong[1*], MD; Shijie Song[2*], MSc; Yuxiang Chris Zhao[3], PhD; Qinghua Zhu[2], PhD; Ling Sha[4], MD

[1]Nanjing Hospital of Chinese Medicine Affiliated to Nanjing University of Chinese Medicine, Nanjing, China

[2]School of Information Management, Nanjing University, Nanjing, China

[3]School of Economics and Management, Nanjing University of Science and Technology, Nanjing, China

[4]Nanjing Jiangning Maternal and Child Health and Family Planning Service Center, Nanjing, China

[*]these authors contributed equally

Corresponding Author:
Shijie Song, MSc
School of Information Management
Nanjing University
Xianlin Ave 163
Nanjing, 210023
China
Phone: 1 15951973800
Email: ssong@smail.nju.edu.cn

Abstract

Background: Diabetes has become one of the most prevalent chronic diseases, and many people living with diabetes use social media to seek health information. Recently, an emerging social media app, TikTok, has received much interest owing to its popularity among general health consumers. We notice that there are many videos about diabetes on TikTok. However, it remains unclear whether the information in these videos is of satisfactory quality.

Objective: This study aimed to assess the quality of the information in diabetes-related videos on TikTok.

Methods: We collected a sample of 199 diabetes-related videos in Chinese. The basic information presented in the videos was coded and analyzed. First, we identified the source of each video. Next, 2 independent raters assessed each video in terms of the completeness of six types of content (the definition of the disease, symptoms, risk factors, evaluation, management, and outcomes). Then, the 2 raters independently assessed the quality of information in the videos, using the DISCERN instrument.

Results: In regard to the sources of the videos, we found 6 distinct types of uploaders; these included 3 kinds of individual users (ie, health professionals, general users, and science communicators) and 3 types of organizational users (ie, news agencies, nonprofit organizations, and for-profit organizations). Regarding content, our results show that the videos were primarily about diabetes management and contained limited information on the definition of the disease, symptoms, risk factors, evaluation, and outcomes. The overall quality of the videos was acceptable, on average, although the quality of the information varied, depending on the sources. The videos created by nonprofit organizations had the highest information quality, while the videos contributed by for-profit organizations had the lowest information quality.

Conclusions: Although the overall quality of the information in the diabetes videos on TikTok is acceptable, TikTok might not fully meet the health information needs of patients with diabetes, and they should exercise caution when using TikTok as a source of diabetes-related information.

(J Med Internet Res 2021;23(9):e30409) doi:10.2196/30409

KEYWORDS

diabetes; information quality; infodemiology; social media; short video apps; TikTok

Introduction

Diabetes has become one of the most prevalent chronic diseases throughout the world. According to a recent report by the International Diabetes Federation, the estimated global prevalence of diabetes in people aged 20-79 years reached 493 million in 2019, accounting for 9.3% of the total world population [1]. Crude estimates of diabetes prevalence are 13.0%

for US adults [2] and 12.8% for Chinese adults [3]. This high prevalence of diabetes results in huge financial burdens and losses for societies. In 2019, diabetes and related complications led to approximately 4.2 million deaths globally and resulted in US $760 billion of health expenditures [1]. Therefore, there is a pressing need to take action in managing diabetes.

Individuals living with diabetes can actively manage their chronic condition. Early studies suggest that intensive blood glucose control can greatly reduce the risk of complications from the disease [4-6]. However, effective blood glucose control is not an easy task among people living with diabetes. Adequate glycemic controls require a constellation of actions, such as a customized diet, exercise plans, regular self-assessments of blood glucose levels, and optimized medication [7]. According to one study, in 2013, only 25.8% of patients with diabetes had received treatment in China, and only 39.7% of those treated had adequate blood glucose control [8]; such low treatment and adherence rates may be associated with people's limited knowledge of the disease [9]. People living with diabetes often have diverse needs for information regarding their chronic condition, such as basic information on diabetes and the effectiveness of treatment options, on the sequelae of diabetes, blood glucose control, etc [10]. Nevertheless, they usually encounter many difficulties finding relevant and easy-to-understand information on their conditions [11].

Emerging internet technologies provide opportunities for better health communication and patient education. The internet has shifted the role of patients from passive information recipients to active information seekers [12]. General health information consumers use social media platforms (eg, discussion forums, microblogs, and group chatting) to seek both instrumental advice and emotional support [13-15]. Patients with diabetes who actively use social media for information, according to recent evidence, are associated with having lower glycated hemoglobin values [16]. A possible explanation for this is that social media provides patients with many opportunities to gain health knowledge, thereby increasing patient activation (ie, the ability and willingness that equip patients to take active action in managing their health care) [17]. Therefore, it is essential to utilize social media for better health communication for managing diabetes conditions.

Despite the considerable benefits of social media, its use for health communication has some limitations. In the literature, the quality of the information is the most extensively mentioned concern [18]. The possibility of encountering faulty health information on social media increases risks for patients, who may make health decisions on the basis of inaccurate information [19]. The quality of unmoderated information poses challenges for both patients and health care providers. On one hand, patients need to be able to distinguish high-quality information sources from low-quality ones [20,21]; one the other, health professionals and institutions are expected to respond to and combat health misinformation to protect the public [22]. Therefore, it is important to examine the quality of health information on social media.

Recently, an emerging short video app, TikTok, has attracted the interest of health care researchers [23]. During the COVID-19 pandemic, health-related videos on TikTok were widely viewed and shared. For example, COVID-19–related videos on the app have been watched approximately 93.1 billion times [24]. TikTok contains many videos about diabetes; however, their quality remains unstudied. To address this gap, this study aimed to systematically assess the quality of the information in diabetes-related videos on TikTok.

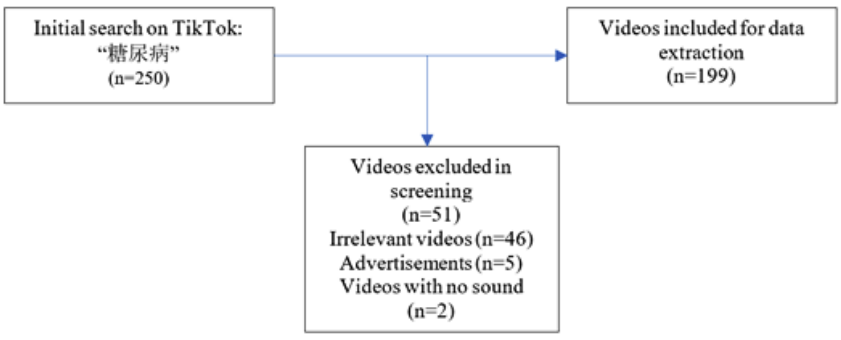
Methods

Search Strategy and Data Extraction

Using the keyword "糖尿病" ("diabetes" in Chinese), we searched TikTok during the period from January 20-25, 2021, and we retrieved the first 250 videos delivered by TikTok's recommended sorting process. We included the videos directly related to diabetes and excluded videos on other topics, commercial advertisements, and videos with no sound. After the screening, we obtained 199 videos for further data extraction and analysis (Figure 1).

We retrieved and extracted the basic information for each video, including the URL, publication date, name of the uploader, type of uploader (individual vs organization), uploaders' verification status, length of the video, number of times it was shared, and number of "likes" and comments it received. The extracted data were recorded in Excel (Microsoft Inc).

Figure 1. Video screening procedure.



Measures

We measured 2 aspects of diabetes-related videos on TikTok: their content and the quality of their information. First, we adopted the coding schema proposed by Goobie et al [25] to rate the quality of six types of content: the definition of the disease, signs/symptoms, risk factors, evaluation, management, and outcomes. Two raters assessed each video independently

and scored how sufficiently the video addressed each of the content types on a 3-item scale: 0 points (no content), 1 point (some content), 2 points (extensive content).

To rate the quality of the information, we adopted the DISCERN instrument. According to a systematic review [26], since its publication in 1998, DISCERN has been one of the most widely adopted instruments for assessing the quality of health information. It is a brief questionnaire that enables its users to assess the quality of health information concerning treatment choices [27]. The instrument consists of 16 questions, with response choices based on a 5-point scale, ranging from 1=poor to 5=good. These 16 questions are divided into 3 sections. The first 8 concern the reliability of the publication, such as whether its aims are clear and whether it is relevant, balanced, and unbiased. The scores for this section indicate whether the publication can be trusted as a source of information for choosing a treatment for a particular disease. The 7 questions in the second section focus on the details of treatment choices, such as whether the publication describes how each treatment works and explains its risks and benefits. The scores for this section reflect the quality of the publication's information about treatment choices (including self-care). The third section consists of 1 final question, based on all the previous ones; it asks users to rate the overall quality of the publication as a source of information about treatment choices. Of note is the fact that, although the original DISCERN instrument was designed for evaluating written publications, it has been widely used for assessing health-related videos. For example, researchers have used it to evaluate YouTube videos informing patients about treatments for cancer [28,29] and diabetes [30].

Rating process

Two authors (WK and LS) worked on the rating tasks; both are certified physicians who work at endocrinology departments at 2 local hospitals. The raters independently scored each video for its coverage of the 6 types of content and applied the 16 questions of the DISCERN instrument. Interrater reliability was assessed with SPSS (version 22, IBM Corp). The interrater reliability for each of the 6 items relating to video content ranged from 0.813 to 0.981, and all of the reliability coefficients are highly significant at an error margin of 0.1%. The interrater reliability for each of the 16 items of the DISCERN instrument ranges from 0.898 to 0.991, and all of the reliability coefficients are highly significant at an error margin of 0.1%. These results indicate satisfactory interrater reliability.

Results

Video Sources

We identified 2 primary sources of the videos: individual and organizational users. We further identified 3 types of video creators among individual users: health professionals, science communicators, and general users. Among organizational users, we identified three types of sources: news agencies, nonprofit organizations, and for-profit organizations.

The results suggest that individual users published most of the videos (n=156, 78.4%). Among individual users, health professionals contributed the most videos (n=138, 69.3%), followed by general users (n=12, 6.0%), and science communicators (n=6, 3.0%). We noted that only 43 videos were uploaded by organizational users, and these accounted for 21.6% of the videos in our sample. Among organizational users, news agencies contributed the most videos (n=31, 15.6%), followed by nonprofit organizations (n=7, 3.5%) and for-profit organizations (n=5, 2.5%) (Table 1).

In our sample, the shortest video lasts only 13 seconds, while the longest lasts 407 seconds. On average, the videos are approximately 1 minute long. All videos were published after 2019. The earliest video had been on TikTok for 589 days, while the latest one was published 3 days prior to the day we collected the data. The videos in the sample received 2.75 million "likes" and 157,700 comments and were shared 305,200 times. Table 2 shows the characteristics of the videos, described by the median numbers across different sources.

Table 1. Descriptions of video sources.

Source type	Source description	Videos, n (%)
Individual users (n=156)		
Health professionals	Individuals who identify themselves as health professionals (eg, doctors and nurses)	138 (69.3)
General users	General users (eg, general health consumers)	12 (6.0)
Science communicators	Individuals who are engaged in scientific communication (eg, popular science writers)	6 (3.0)
Organizational users (n=43)		
News agencies	News agencies and the press	31 (15.6)
Nonprofit organizations	Organizations operated for collective, public, or social benefit and public hospitals	7 (3.5)
For-profit organizations	Organizations that pursue commercial interests	5 (2.5)

Table 2. Characteristics of the videos across sources (median numbers).

Source type	Length of video (seconds), median	Days on TikTok, median	"Likes," median	Comments, median	Times shared, median
Health professionals (n=138)	50.5	131	4191	164	802.5
General users (n=12)	135.5	161	3089.5	78.5	891
Science communicators (n=6)	43	135	24,000	593	4966
News agencies (n=31)	43	90	9917	167	3505
Nonprofit organizations (n=7)	52	200	23,000	194	3023
For-profit organizations (n=5)	63	317	33,000	405	6759

Video Content

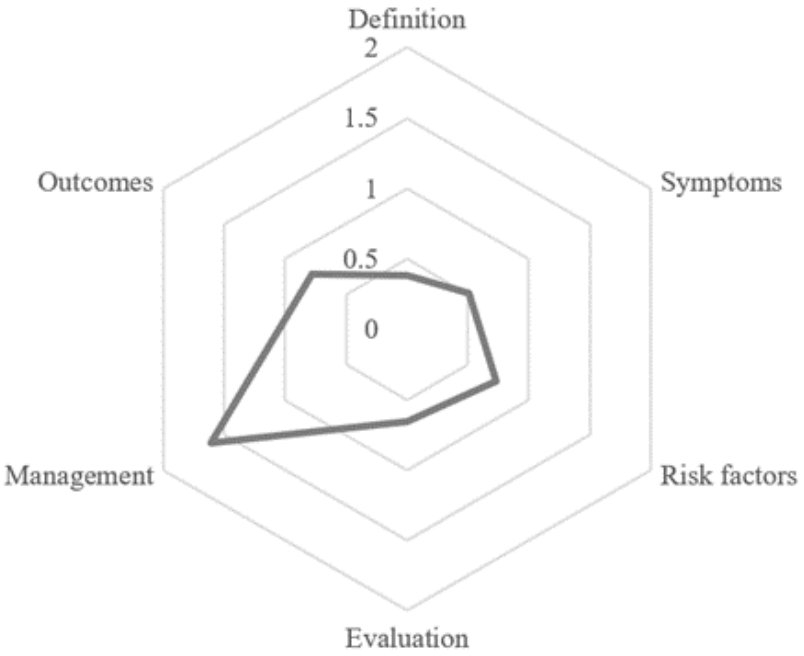
It was not our intention to exclude any type of diabetes during searching and screening. However, we found that most of the videos were about type 2 diabetes mellitus (n=193, 97%). We identified only 4 (2%) videos about gestational diabetes mellitus and 2 (1%) videos about type 1 diabetes mellitus.

Moreover, we averaged the scores of the 2 raters for each aspect of the video content and obtained scores that ranged over the full 5-point scale, from "no content" to "extensive content." The results show that more than half of the videos contain little or no content on the definition of the disease, symptoms, or evaluations of diabetes. Overall, 46.2% of the videos contain little or no content on diabetes-related risk, and 66.8% have some or more information on outcomes. Management of diabetes was the most frequent topic in the sample. Overall, 67.8% of the videos sufficiently introduced diabetes management (Table 3). The overall scores for all the videos are given in Figure 2.

Table 3. Completeness of video content.

Content	Definition, n (%)	Symptoms, n (%)	Risk factors, n (%)	Evaluation, n (%)	Management, n (%)	Outcomes, n (%)
No content (0 points)	113 (56.8)	115 (57.8)	79 (39.7)	100 (50.3)	14 (7.0)	54 (27.1)
Little content (0.5 points)	27 (13.6)	9 (4.5)	13 (6.5)	8 (4.0)	3 (1.5)	12 (6.0)
Some content (1 point)	52 (26.1)	47 (23.6)	66 (33.2)	54 (27.1)	44 (22.1)	115 (57.8)
Most content (1.5 points)	7 (3.5)	14 (7.0)	17 (8.5)	4 (2.0)	3 (1.5)	5 (2.5)
Extensive content (2 points)	0 (0)	14 (7.0)	24 (12.1)	33 (16.6)	135 (67.8)	13 (6.5)

Figure 2. Completeness of video content.



Information Quality

Our results suggest that the general quality of the diabetes information videos on TikTok is acceptable. Overall, the videos published by the nonprofit organizations had the highest DISCERN scores, followed by those published by the health professionals and news agencies. The videos published by the for-profit organizations had the lowest total DISCERN scores, followed by those of the science communicators and general TikTok users. The mean numbers for the whole instrument indicate significant differences across the video sources, at α=.01 (Table 4).

Regarding reliability, videos published by the nonprofit organizations had the highest scores, while those from the for-profit organizations had the lowest scores. Our results suggest that nonprofit organizations, news agencies, and individual health professionals also contributed videos with relatively high reliability. The differences in reliability across the different video sources are significant, at α=.01.

In regard to treatment choices, diabetes-related videos on TikTok were of medium to low quality. Nonprofit organizations and health professionals contributed higher-quality videos on treatment choices than other sources; however, the differences are not significant.

Table 4. DISCERN scores of diabetes-related TikTok videos by source.

Video source	Reliability of the videos (items 1-8)[a], mean (SD)	Quality of treatment choices (items 9-15)[b], mean (SD)	Overall information quality (item 16)[c], mean (SD)	Total DISCERN scores[d], mean (SD)
Health professionals (n=138)	28.10 (3.59)	16.37 (5.00)	3.26 (0.67)	47.74 (7.71)
General users (n=12)	25.38 (3.19)	13.42 (3.63)	2.58 (0.51)	41.38 (6.14)
Science communicators (n=6)	25.00 (3.21)	15.25 (5.29)	3.08 (0.20)	43.33 (8.23)
News agencies (n=31)	28.48 (3.76)	16.02 (5.14)	3.23 (0.71)	47.73 (7.39)
Nonprofit organizations (n=7)	29.14 (2.25)	18.00 (3.40)	3.50 (0.50)	50.64 (4.61)
For-profit organizations (n=5)	24.20 (30.50)	13.20 (5.03)	2.60 (0.55)	40.00 (7.11)

[a]*P*=.005 (1-way analysis of variance).

[b]*P*=.23 (1-way analysis of variance).

[c]*P*=.004 (1-way analysis of variance).

[d]*P*=.009 (1-way analysis of variance).

Discussion

Principal Findings

This study systematically evaluated the information quality of diabetes-related videos on TikTok. According to a recent systematic review [13], the use of social media as a source of information is gaining in popularity among patients with diabetes. The various social media channels provide patients with a convenient means to seek medical knowledge and get social support [31]. While traditional social platforms (eg, Facebook, Twitter, and Instagram) have been widely investigated as channels of diabetes-related health communication [32], the role of emerging, mobile-based apps in disseminating diabetes knowledge is not yet fully understood. Our results reveal that TikTok is a powerful platform for disseminating diabetes-related information. The 199 diabetes videos examined in our study received 2.75 million “likes” and were commented on and shared thousands of times, which indicates that TikTok is a promising channel for health communication.

We identified 2 main categories of video uploaders (ie, individual and organizational users), each containing several more specific types of users. Individual users included health professionals, science communicators, general TikTok users; organizational users comprised news agencies, nonprofit organizations, and for-profit organizations. Health professionals contributed the most videos, while the for-profit organizations contributed the least. Many prior studies have suggested that health professionals and organizations can utilize social media for effective health communication and public health promotion [33,34]. Our study revealed that health professionals in China have been actively engaged in promoting diabetes knowledge via TikTok; however, nonprofit health organizations use this emerging video-based channel less frequently.

In terms of video content, the study found 3 types of imbalances. First, most of the videos were about type 2 diabetes mellitus, while very few videos discussed gestational diabetes mellitus and type 1 diabetes mellitus. Second, most of the videos were about disease management, but few fully addressed other aspects of content, such as the definition and symptoms of the disease, risk factors, evaluation, and outcomes. Third, when many videos are generally reliable, these videos were of average to fair quality concerning treatment choices. Prior studies suggest that patients with diabetes have various health information needs, including a need for information about treatment, course of the disease, abnormalities in glucose metabolism, progression of diabetes through their life cycle, coping techniques, and prevention [35,36]. Moreover, these information needs vary, depending on the type of diabetes mellitus. For example, young people with type 1 diabetes mellitus may be particularly interested in “diabetes through the life circle” [35]. Given the observed imbalances in video content, we suspect that current diabetes-related videos on TikTok cannot fully meet patients’ information needs. Therefore, we call for more pertinent videos to address patients’ comprehensive information needs.

Our study found that the quality of information in the videos differed with the type of source. Videos published by nonprofit organizations had the highest quality, while those from the for-profit organizations had the lowest quality. This finding is consistent with those of prior studies, which suggest that government-sponsored platforms are more likely to publish high-quality information than for-profit organizations [25]. Unfortunately, the videos contributed by the nonprofit organizations account for a mere portion of the total corpus of diabetes-related videos on TikTok. We suggest that government departments and public hospitals contribute more high-quality material and leverage the power of this social media channel to promote public health. Given the large variations in information quality from the different sources, we also suggest that patients exercise caution when using TikTok to obtain diabetes-related information.

Limitations and Future Directions

The findings of this study should be viewed in light of several limitations. First, the study looked only at the quality of diabetes information, not the quality of communication. For example, we observed that the communication modalities varied largely in the TikTok videos. Some videos used rich materials (eg, illustrative images or cases and visual data) to communicate the information, while some were based on plain narratives. Unfortunately, the instruments employed in this paper targeted information quality and overlooked the quality of communication. We call for more empirical studies in the future to investigate the communication quality of diabetes-related videos on TikTok. Second, the videos included in our study were in Chinese; therefore, the findings cannot be applied to diabetes-related videos on TikTok in other languages (eg, English). We encourage future researchers to assess the information quality of diabetes-related videos in other languages to obtain deeper insight into quality issues with diabetes-related videos on TikTok. Third, there are many instruments for assessing the quality of health-related information, such as DISCERN, JAMA benchmarks, and the HONcode principles. This study employed the DISCERN instrument because it has proved effective for assessing the quality of videos on other platforms and apps (eg, YouTube). However, we encourage more studies using a variety of instruments to triangulate the validity of these findings in the future.

Conclusions

This study assessed the information quality of 199 diabetes-related videos on the short video mobile app TikTok. The results show that the videos primarily addressed diabetes management but contained limited information on other types of content, such as the definition and symptoms of the disease, risk factors, evaluation, and outcomes. The overall quality of the diabetes videos was found to be acceptable on average, although it varied significantly, depending on the type of source. We conclude that the health information needs of patients with diabetes might not be fully met by watching TikTok videos, and patients should exercise caution when using TikTok for diabetes-related information.

Acknowledgments

We thank the anonymous reviewers and editors for their helpful comments on the manuscript. The study was supported by the National Science Foundation of China (72074112, 72174083) and the Key Projects of Philosophy and Social Sciences Research of Chinese Ministry of Education (19JZD021).

Conflicts of Interest

None declared.

References

1. IDF Diabetes Atlas 2019. International Diabetes Federation. URL: https://www.diabetesatlas.org/en/resources/ [accessed 2021-08-16]
2. National Diabetes Statistics Report, 2020. Atlanta, GA. Centers for Disease Control and Prevention, U.S. Dept of Health and Human Services. 2020. URL: https://www.cdc.gov/diabetes/data/statistics-report/index.html [accessed 2021-08-16]
3. Li Y, Teng D, Shi X, Qin G, Qin Y, Quan H, et al. Prevalence of diabetes recorded in mainland China using 2018 diagnostic criteria from the American Diabetes Association: national cross sectional study. BMJ 2020 Apr 28;369:m997 [FREE Full text] [doi: 10.1136/bmj.m997] [Medline: 32345662]
4. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). The Lancet 1998 Sep 12;352(9131):837-853. [doi: 10.1016/S0140-6736(98)07019-6] [Medline: 9742976]
5. Dluhy RG, McMahon GT. Intensive glycemic control in the ACCORD and ADVANCE trials. N Engl J Med 2008 Jun 12;358(24):2630-2633. [doi: 10.1056/NEJMe0804182] [Medline: 18539918]
6. ADVANCE Collaborative Group. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. N Engl J Med 2008 Jun 12;358(24):2560-2572. [doi: 10.1056/NEJMoa0802987] [Medline: 18539916]
7. Greenwood DA, Gee PM, Fatkin KJ, Peeples M. A Systematic Review of Reviews Evaluating Technology-Enabled Diabetes Self-Management Education and Support. J Diabetes Sci Technol 2017 Sep;11(5):1015-1027 [FREE Full text] [doi: 10.1177/1932296817713506] [Medline: 28560898]

8. Xu Y, Wang L, He J, Bi Y, Li M, Wang T, 2010 China Noncommunicable Disease Surveillance Group. Prevalence and control of diabetes in Chinese adults. JAMA 2013 Sep 04;310(9):948-959. [doi: 10.1001/jama.2013.168118] [Medline: 24002281]
9. Bains SS, Egede LE. Associations between health literacy, diabetes knowledge, self-care behaviors, and glycemic control in a low income population with type 2 diabetes. Diabetes Technol Ther 2011 Mar;13(3):335-341 [FREE Full text] [doi: 10.1089/dia.2010.0160] [Medline: 21299402]
10. Weymann N, Härter M, Dirmaier J. Information and decision support needs in patients with type 2 diabetes. Health Informatics J 2016 Mar;22(1):46-59 [FREE Full text] [doi: 10.1177/1460458214534090] [Medline: 24916569]
11. Lee DYL, Armour C, Krass I. The development and evaluation of written medicines information for type 2 diabetes. Health Educ Res 2007 Dec;22(6):918-930. [doi: 10.1093/her/cym048] [Medline: 17947249]
12. Nordfeldt S, Ängarne-Lindberg T, Berterö C. To use or not to use--practitioners' perceptions of an open web portal for young patients with diabetes. J Med Internet Res 2012 Nov 09;14(6):e154 [FREE Full text] [doi: 10.2196/jmir.1987] [Medline: 23137767]
13. Elnaggar A, Ta Park V, Lee SJ, Bender M, Siegmund LA, Park LG. Patients' Use of Social Media for Diabetes Self-Care: Systematic Review. J Med Internet Res 2020 Apr 24;22(4):e14209 [FREE Full text] [doi: 10.2196/14209] [Medline: 32329745]
14. Magnezi R, Bergman YS, Grosberg D. Online activity and participation in treatment affects the perceived efficacy of social health networks among patients with chronic illness. J Med Internet Res 2014 Jan 10;16(1):e12 [FREE Full text] [doi: 10.2196/jmir.2630] [Medline: 24413148]
15. Allen C, Vassilev I, Kennedy A, Rogers A. Long-Term Condition Self-Management Support in Online Communities: A Meta-Synthesis of Qualitative Papers. J Med Internet Res 2016 Mar 10;18(3):e61 [FREE Full text] [doi: 10.2196/jmir.5260] [Medline: 26965990]
16. Oser SM, Stuckey HL, Parascando JA, McGinley EL, Berg A, Oser TK. Glycated Hemoglobin Differences Among Blog-Reading Adults With Type 1 Diabetes Compared With Those Who Do Not Read Blogs: Cross-Sectional Study. JMIR Diabetes 2019 Apr 02;4(2):e13634 [FREE Full text] [doi: 10.2196/13634] [Medline: 30938693]
17. Grosberg D, Grinvald H, Reuveni H, Magnezi R. Frequent Surfing on Social Health Networks is Associated With Increased Knowledge and Patient Health Activation. J Med Internet Res 2016 Aug 10;18(8):e212 [FREE Full text] [doi: 10.2196/jmir.5832] [Medline: 27511272]
18. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
19. Song S, Zhang Y, Yu B. Interventions to support consumer evaluation of online health information credibility: A scoping review. Int J Med Inform 2021 Jan;145:104321. [doi: 10.1016/j.ijmedinf.2020.104321] [Medline: 33202372]
20. Song S, Zhao Y, Song X, Zhu Q. The Role of Health Literacy on Credibility Judgment of Online Health Misinformation. 2019 Presented at: 2019 IEEE International Conference on Healthcare Informatics (ICHI); June 10-13, 2019; Xi'an. [doi: 10.1109/ichi.2019.8904844]
21. Zhang Y, Song S. Older Adults' Evaluation of the Credibility of Online Health Information. 2020 Presented at: CHIIR '20: Conference on Human Information Interaction and Retrieval; March 14-18, 2020; Vancouver, BC. [doi: 10.1145/3343413.3377997]
22. Bautista JR, Zhang Y, Gwizdka J. Healthcare professionals' acts of correcting health misinformation on social media. Int J Med Inform 2021 Apr;148:104375. [doi: 10.1016/j.ijmedinf.2021.104375] [Medline: 33461008]
23. Song S, Zhao YC, Yao X, Ba Z, Zhu Q. Short video apps as a health information source: an investigation of affordances, user experience and users' intention to continue the use of TikTok. INTR 2021 Jun 24;ahead-of-print(ahead-of-print):1-23. [doi: 10.1108/intr-10-2020-0593]
24. Ostrovsky AM, Chen JR. TikTok and Its Role in COVID-19 Information Propagation. J Adolesc Health 2020 Nov;67(5):730 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.07.039] [Medline: 32873499]
25. Goobie GC, Guler SA, Johannson KA, Fisher JH, Ryerson CJ. YouTube Videos as a Source of Misinformation on Idiopathic Pulmonary Fibrosis. Ann Am Thorac Soc 2019 May;16(5):572-579. [doi: 10.1513/AnnalsATS.201809-644OC] [Medline: 30608877]
26. Zhang Y, Sun Y, Xie B. Quality of health information for consumers on the web: A systematic review of indicators, criteria, tools, and evaluation results. J Assn Inf Sci Tec 2015 Apr 29;66(10):2071-2084. [doi: 10.1002/asi.23311]
27. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
28. ReFaey K, Tripathi S, Yoon JW, Justice J, Kerezoudis P, Parney IF, et al. The reliability of YouTube videos in patients education for Glioblastoma Treatment. J Clin Neurosci 2018 Sep;55:1-4. [doi: 10.1016/j.jocn.2018.07.001] [Medline: 30075980]
29. Radonjic A, Fat Hing NN, Harlock J, Naji F. YouTube as a source of patient information for abdominal aortic aneurysms. J Vasc Surg 2020 Feb;71(2):637-644. [doi: 10.1016/j.jvs.2019.08.230] [Medline: 31611104]

30. Leong AY, Sanghera R, Jhajj J, Desai N, Jammu BS, Makowsky MJ. Is YouTube Useful as a Source of Health Information for Adults With Type 2 Diabetes? A South Asian Perspective. Can J Diabetes 2018 Aug;42(4):395-403.e4. [doi: 10.1016/j.jcjd.2017.10.056] [Medline: 29282200]
31. Zhang Y, He D, Sang Y. Facebook as a platform for health information and communication: a case study of a diabetes group. J Med Syst 2013 Jun;37(3):9942. [doi: 10.1007/s10916-013-9942-7] [Medline: 23588823]
32. Gabarron E, Larbi D, Dorronzoro E, Hasvold PE, Wynn R, Årsand E. Factors Engaging Users of Diabetes Social Media Channels on Facebook, Twitter, and Instagram: Observational Study. J Med Internet Res 2020 Sep 29;22(9):e21204 [FREE Full text] [doi: 10.2196/21204] [Medline: 32990632]
33. Klassen KM, Borleis ES, Brennan L, Reid M, McCaffrey TA, Lim MS. What People "Like": Analysis of Social Media Strategies Used by Food Industry Brands, Lifestyle Brands, and Health Promotion Organizations on Facebook and Instagram. J Med Internet Res 2018 Jun 14;20(6):e10227 [FREE Full text] [doi: 10.2196/10227] [Medline: 29903694]
34. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2013 Jan;14(1):15-23. [doi: 10.1177/1524839911405850] [Medline: 21558472]
35. Biernatzki L, Kuske S, Genz J, Ritschel M, Stephan A, Bächle C, et al. Information needs in people with diabetes mellitus: a systematic review. Syst Rev 2018 Feb 14;7(1):27 [FREE Full text] [doi: 10.1186/s13643-018-0690-0] [Medline: 29444711]
36. Kuske S, Schiereck T, Grobosch S, Paduch A, Droste S, Halbach S, et al. Diabetes-related information-seeking behaviour: a systematic review. Syst Rev 2017 Oct 24;6(1):212 [FREE Full text] [doi: 10.1186/s13643-017-0602-8] [Medline: 29065919]

Edited by G Eysenbach; submitted 13.05.21; peer-reviewed by A Alasmari, M Lim, R Fang; comments to author 04.06.21; revised version received 21.07.21; accepted 07.08.21; published 01.09.21.

Please cite as:
Kong W, Song S, Zhao YC, Zhu Q, Sha L
TikTok as a Health Information Source: Assessment of the Quality of Information in Diabetes-Related Videos
J Med Internet Res 2021;23(9):e30409
URL: https://www.jmir.org/2021/9/e30409
doi:10.2196/30409
PMID:34468327

Original Paper

Privacy Practices of Health Information Technologies: Privacy Policy Risk Assessment Study and Proposed Guidelines

Haley M LaMonica[1], BSc, MA, PhD, ABPP-CN; Anna E Roberts[1], BAppSc, MExSC, MBMSc; Grace Yeeun Lee[1], MPH, MHM; Tracey A Davenport[1], BA(Hons), eMBA; Ian B Hickie[1], AM, MD, FRANZCP, FASSA

Brain and Mind Centre, The University of Sydney, Camperdown, Australia

Corresponding Author:
Haley M LaMonica, BSc, MA, PhD, ABPP-CN
Brain and Mind Centre
The University of Sydney
88 Mallett Street
Camperdown, 2050
Australia
Phone: 61 0426955658
Email: haley.lamonica@sydney.edu.au

Abstract

Background: Along with the proliferation of health information technologies (HITs), there is a growing need to understand the potential privacy risks associated with using such tools. Although privacy policies are designed to inform consumers, such policies have consistently been found to be confusing and lack transparency.

Objective: This study aims to present consumer preferences for accessing privacy information; develop and apply a privacy policy risk assessment tool to assess whether existing HITs meet the recommended privacy policy standards; and propose guidelines to assist health professionals and service providers with understanding the privacy risks associated with HITs, so that they can confidently promote their safe use as a part of care.

Methods: In phase 1, participatory design workshops were conducted with young people who were attending a participating *headspace* center, their supportive others, and health professionals and service providers from the centers. The findings were knowledge translated to determine participant preferences for the presentation and availability of privacy information and the functionality required to support its delivery. Phase 2 included the development of the 23-item privacy policy risk assessment tool, which incorporated material from international privacy literature and standards. This tool was then used to assess the privacy policies of 34 apps and e-tools. In phase 3, privacy guidelines, which were derived from learnings from a collaborative consultation process with key stakeholders, were developed to assist health professionals and service providers with understanding the privacy risks associated with incorporating HITs as a part of clinical care.

Results: When considering the use of HITs, the participatory design workshop participants indicated that they wanted privacy information to be easily accessible, transparent, and user-friendly to enable them to clearly understand what personal and health information will be collected and how these data will be shared and stored. The privacy policy review revealed consistently poor readability and transparency, which limited the utility of these documents as a source of information. Therefore, to enable informed consent, the privacy guidelines provided ensure that health professionals and consumers are fully aware of the potential for privacy risks in using HITs to support health and well-being.

Conclusions: A lack of transparency in privacy policies has the potential to undermine consumers' ability to trust that the necessary measures are in place to secure and protect the privacy of their personal and health information, thus precluding their willingness to engage with HITs. The application of the privacy guidelines will improve the confidence of health professionals and service providers in the privacy of consumer data, thus enabling them to recommend HITs to provide or support care.

(J Med Internet Res 2021;23(9):e26317) doi:10.2196/26317

KEYWORDS

privacy; mental health; technology; digital tools; smartphone; apps

Introduction

Health Information Technologies

Digital health has quickly become an integral component of best practice health care, transforming the way care is delivered. By capitalizing on digital infrastructure, it is widely recognized that digital health solutions improve access to care, particularly for individuals with mobility or transport restrictions, or for those who live remotely where health care resources may be limited [1,2]. The availability of health information technologies (HITs) is proving invaluable during the COVID-19 pandemic, where face-to-face mental health care is often delivered digitally (eg, videoconferencing) [3]. Beyond access issues, however, digital health also has the potential to optimize or eliminate waitlists and facilitate routine outcome monitoring to strengthen and maintain patient-health professional relationships [4], allowing for shared, data-driven decision-making on appropriate treatment plans [5]. With a greater need for and reliance on digital health solutions for screening, treatment, and ongoing maintenance of health, there is now an increased focus on the privacy and security of personal and health information collected via HITs, such as health-related apps and e-tools (eg, websites and web-based courses).

Legal and Ethics Rights of Individuals

It is crucial to consider the legal and ethical rights of individuals who choose to both explicitly and passively share their web-based health information. This is essential, particularly in the area of mental health care, where data often contain highly personal information that could cause significant harm and distress if not handled appropriately. There is increasing documentation and guidance in this area, such as the recent release of the National Safety and Quality Digital Mental Health Standards (consultative draft) by the Australian Commission on Safety and Quality in Health Care [6], which includes educational brochures that provide tips to consumers, caregivers [7], and clinicians [8] on choosing a digital mental health service. Specific to privacy, the Australian Privacy Principles require that all organizations have a clearly expressed and current privacy policy detailing how personal information is managed [9]. As personal and health information is deemed particularly sensitive, extra protection concerning its handling was established under the Privacy Act 1988 [10]. The World Economic Forum also has highlighted *trust* as one of the primary issues that needs to be addressed on a global scale to ensure consumers' and health professionals' trust in the privacy and security standards of new digital tools and technology-based therapies [11].

Privacy Practices of HITs

The use of health-related apps has rapidly increased in recent years, with 47% of Australian consumers using apps in 2018 [12]. Health professionals are also increasingly recommending HITs as part of clinical practice. For example, approximately half of Australian general practitioners responding to an annual technology survey indicated that they recommend HITs for at least monthly use by patients, with mindfulness and mental health apps recommended most often [13]. Although HITs, including apps and e-tools, have gained considerable favor with consumers and health professionals for promoting self-management of health and well-being, the privacy of personal and health information remains a notable area of concern. A systematic review of 79 health and wellness apps certified as clinically safe and trustworthy by the United Kingdom National Health Service Health Apps Library found that 66% of apps transmitting personal information on the internet did so without encryption, and 20% did not have any form of privacy policy [14]. Furthermore, although no app collected or shared information in a manner that was not explicitly stated in the privacy policy, the nature of the personal information included in such transmission was not described in 78% of policies [14]. A recent cross-sectional assessment of 36 top-ranked apps found that 92% of them transmitted data to a third party; however, only 64% of privacy policies made this explicit. In addition, only 43% and 50% of privacy policies disclosed that apps were transmitting data to Google and Facebook, respectively [15]. This begs the question, can users trust that personal and health information collected via HITs will be kept private and secure?

Objectives

This study aims to use co-design methodologies to better understand young people's preferences for learning about how their personal and health information will be handled by HITs and create prototypes for the InnoWell Platform. The InnoWell Platform was developed by InnoWell, a joint venture between the University of Sydney and PricewaterhouseCoopers (PwC; Australia) through Project Synergy, an Aus $30 million (US $22.1 million) Australian government–funded initiative [16]. As described previously [5,17], the InnoWell Platform is a co-designed digital tool embedded within traditional in-clinic and web-based mental health services to support person-centered, measurement-based care. This study also seeks to develop and apply a privacy policy risk assessment tool to assess whether existing HITs meet the recommended privacy policy standards and present guidelines to assist health professionals and service providers to ask the appropriate questions for themselves and HIT manufacturers to ensure that they can confidently promote the safe use of HITs as part of care.

Methods

Phase 1

Participatory Design Workshops

Participatory design (ie, co-design) methodologies are routinely used to ensure that digital tools are designed to meet the needs of the intended user base, thus increasing uptake and engagement [18,19]. Our research team has extensive experience in the use of participatory design, including workshop design and facilitation as well as knowledge translation [19-23].

Our research team conducted a series of 10 participatory design workshops from July to September 2018 in 9 urban and rural *headspace* centers across Australia (Ashfield, Bathurst, Broken Hill, Dubbo, Orange, Wagga Wagga, and Wollongong, New South Wales [NSW]; Townsville, Queensland; Edinburgh North, South Australia). *headspace* centers are primary mental health

services providing support to young Australians and their families to promote mental health and engagement with the community. The methods and results of these workshops were previously reported in detail by Cheng et al [24]. In summary, the workshops brought together key stakeholders from the participating *headspace* communities, including help-seeking young people, supportive others, health professionals, and service providers, to collaboratively discuss technology designs, ideas, and principles to support mental health and well-being. In these workshops, technology designs, ideas, and principles, including the concepts of data privacy and security, were evaluated by the participants. Due to varying numbers of participants and other contextual factors, workshops ranged from 2.5 to 4 hours in duration and consisted of the following stages: discovery, evaluation, and prototype. The discovery stage focused on current ways in which technology is used by participants, including for the purposes of supporting mental health, internet access issues in regional communities, and concerns about sharing personal and health information via web-based programs or tools. During the evaluation stage, participants were presented with images of components of the InnoWell Platform and asked to document their feedback. Finally, in the prototype stage, participants were given the opportunity to brainstorm new items, functionalities, and wireframes using sketchbooks for the components or functionalities of the InnoWell Platform. In this paper, we present findings related to the development and inclusion of privacy information in HITs as co-designed by the participants. Importantly, these results were not included in the original publication [24].

Participants

Participants included individuals from the participating *headspace* communities, including young people attending a participating *headspace* center, a supportive other of a young person attending a participating *headspace* center (eg, family member, caregiver, or friend), or a health professional or service provider working at a participating *headspace* center. The inclusion criteria for participation in the study required participants to be aged ≥12 years, proficient in reading and speaking English, and having completed the participant consent process. Details of recruitment, screening, and informed consent processes have been previously documented by Cheng et al [24].

Knowledge Translation

The InnoWell Platform consists of a multidimensional assessment evaluating a range of biopsychosocial domains (eg, psychological distress, sleep, alcohol use, and physical health) to provide a holistic view of the consumer. The assessment results are available in real time and designed to be reviewed collaboratively by the consumer and their health professional to promote shared decision-making about care options, accounting for consumer preferences. A consumer's progress can then be routinely tracked and monitored over time using assessment tools to inform treatment planning, clinical review, and coordinated care within and between services.

As previously described [24], workshop notes and descriptive artifacts were reviewed by an independent knowledge translation team with 2 young researchers without previous knowledge of the InnoWell Platform or the fundamental principles underpinning its design. Each team member, taking note of their general observations, reviewed all data independently and subsequently identified the key concepts noted by workshop participants to then produce prototype designs of the components of the InnoWell Platform.

Ethics

This study was approved by the University of Sydney's Human Research Ethics Committee (protocol number: 2018/130).

Phase 2

Development of the Privacy Policy Risk Assessment Tool

Drawing from digital health privacy and security criteria published by existing research and professional associations, our research team developed a privacy policy risk assessment tool (Multimedia Appendix 1) to evaluate privacy policies for HITs, including apps and e-tools. Developed in 2018 by a multidisciplinary expert panel of health professionals, informaticists, medical students, and consumers with lived experience of mental illness, the American Psychiatric Association app evaluation model includes risk, privacy and safety questions as part of their simple four-stage hierarchal app review process [25]. Although the American Psychiatric Association model served as the primary reference, to ensure an all-encompassing assessment tool, we also incorporated details from the National and Safety Quality Digital Mental Health Standards developed by the Australian Commission on Safety and Quality in Health Care to improve the quality of digital mental health care and protect the privacy of service users [6]. The final privacy policy risk assessment tool consists of 23 items covering (1) privacy policy (6 items), (2) personal health information (PHI; 4 items), (3) data security and storage (9 items), and (4) other aspects of privacy (4 items). As the questions and concepts included in the privacy policy risk assessment tool were drawn from previously published privacy and security standards, we were confident that the measure has face and construct validity; however, a specific validity analysis was not conducted.

Review of Digital Health Tool Privacy Policies

Importantly, the privacy policy risk assessment tool is broadly applicable to HITs. To demonstrate its utility, in this study, we evaluated the privacy policies of the apps and e-tools in the *youth* configuration of the InnoWell Platform [16]. Within the InnoWell Platform, there are two types of care options: clinical and nonclinical. Clinical care options require health professionals' involvement, such as individual therapy and group therapy, whereas a consumer can immediately access and begin using nonclinical care options, such as apps and e-tools, without the support of a health professional [26]. During the co-design process, care options are tailored to the consumer population; in this case, young people receiving care through primary youth mental health services (eg, *headspace* centers).

Aligned with established evaluation processes [15], privacy policies and any related material that may contain privacy-related content, such as terms and conditions, were

sourced from associated websites and app store links. A nonrestrictive process was used, allowing all hyperlinks from the app store or within the privacy policy to be considered. All available information was collated and reviewed in accordance with the privacy policy risk assessment tool developed primarily by a member of the research team (AER), with support from a research intern (Toby Wong).

Assessing Readability

The readability levels of each privacy policy were assessed as part of the evaluation process. There are multiple readability formulas available; however, for the purpose of this study, we used the Flesh-Kincaid readability tests because of their acceptance in the health care literature [27]. The readability tests were designed to indicate how difficult an English passage of writing is to understand using a formula calculated from the average number of syllables per word and the average number of words per sentence [28]. Both the Flesch reading ease and the Flesch-Kincaid reading grade level were calculated using the web-based readability tools [29]. The scores indicate the readability of a passage of text on a scale from “very easy to read" to "very difficult to read” and “fifth grade reading level” to “professional reading level,” respectively. Scores were calculated using the first two paragraphs of the privacy policies. The paragraphs (at least 100 words) were copied into a readability formula calculator, and a score was given.

Data Analysis

Descriptive statistics were used to analyze all aspects of the assessment data. SPSS version 25 (IBM Corp) was used for all analyses.

Phase 3

Development of Privacy Guidelines: Consultations

As a result of Project Synergy and the development of the InnoWell Platform, a set of core principles and privacy guidelines were used as the starting point to formalize a more encompassing set of privacy guidelines. A series of consultations were held between 2014 and 2016 (phase 1) and then again between 2017 and 2018 (phase 2) to develop a set of privacy guidelines for Project Synergy. Initial consultations were conducted by Orygen (the National Centre of Excellence in Youth Mental Health) [30]. The subsequent consultations were conducted by the Project Synergy research and development team (led by the authors TAD and IBH). These consultations were held at the University of Sydney’s Brain and Mind Centre, either in-person or via teleconference, and brought together key stakeholders across relevant organizations.

Participants

Participants included key stakeholder groups, including Orygen (the National Centre of Excellence in Youth Mental Health), the Young and Well Cooperative Research Centre, Mental Health Commission of NSW (Pacific Privacy Consulting Pty Ltd), the Project Synergy research and development team (the University of Sydney’s Brain and Mind Centre), InnoWell, and PwC (Australia). Select individuals were nominated by each organization, with participants contributing diverse expertise and experience, such as the involvement of 2 ex-serving privacy commissioners for NSW and Victoria (Australia).

Preliminary Development of Privacy Guidelines for Phase 1 Project Synergy (2014-2016)

From the outset of the development cycle of the prototype and as part of phase 1 of Project Synergy (2014-2016) [16], this ongoing series of consultations were envisaged to inform the development of the privacy guidelines. Therefore, an initial set of guidelines was developed; these guidelines were produced by Orygen and supported by the Young and Well Cooperative Research Centre and the Mental Health Commission of NSW [30]. They were also reviewed by the Project Synergy research and development team at the University of Sydney’s Brain and Mind Centre (the authors) and Pacific Privacy Consulting.

Ongoing Development of Privacy Guidelines for Phase 2 of Project Synergy (2017-2020)

As a result of the review of the initial guidelines developed in phase 1 of Project Synergy, a narrower focus was decided upon and used as the starting point for the development of more encompassing privacy guidelines for phase 2 of Project Synergy (2017-20). Specifically, upon review of the initial guidelines whereby the University of Sydney and Pacific Privacy Consulting determined that privacy concerns were the most important priority, a narrower focus was given to 8 core foundation principles to be followed by organizations using the prototype in phase 1 of Project Synergy (2014-2016) [16]. The core foundation principles included responsibility for legal compliance, anonymous or pseudonymous services wherever practicable, individual control, transparency, interaction with individuals, encryption, deidentification, and cross-border processing risk. Development of the broad structure and content of the guidelines was guided by the Project Synergy research and development team, with review and input by Pacific Privacy Consulting and InnoWell.

This paper presents privacy guidelines to assist health service providers in considering the privacy of their consumers when using HITs as part of care. The guidelines were first drafted by the Project Synergy research and development team based on the information gathered through the initial collaborative consultation process. The checklist was then reviewed, discussed, and evaluated by the research team, ultimately resulting in agreement by consensus.

Results

Knowledge Translation Findings

The results of the knowledge translation process highlighted that participants wanted privacy information to be presented before being required to create an account. Specifically, they emphasized the need for privacy information to be readily available, allowing a user to be completely comfortable when entering more sensitive information into a HIT, such as the InnoWell Platform (eg, “Always ask could this site be more secure with my information” [Wollongong workshop]). This included the ability to change permissions concerning data sharing at their discretion (eg, “[I] would want privacy settings

in place so that not everyone that shares the system can see" [Broken Hill workshop]). Participants noted that privacy information is frequently confusing and difficult to understand, leaving them unsure whether they should trust the HIT to protect their personal and health information. Thus, multiple participants suggested a pin code or password (eg, "Consider password security like in bank apps." [Townsville workshop]) to access certain data so the consumer controls who has access to their information in the InnoWell Platform. Importantly, the idea of consumer control extended beyond HIT manufacturers such as InnoWell and included health professionals and supportive others (ie, family members and carers) accessing personal and health information (eg, "Need privacy setting like Facebook...can filter who can see the information" [Bathurst workshop]).

App and e-Tool Privacy Policy Assessment

We evaluated 34 privacy policies using the privacy policy risk assessment tool. Most of these apps and e-tools were designed for both youth and adult users (28/34, 82%), whereas the remaining 18% (6/34) were specifically designed for youth audiences (aged ≤25 years). Most apps and e-tools (20/34, 59%) were self-help or self-management tools supporting mental health and well-being, including three specifically using cognitive behavioral therapy techniques. There were also 12% (4/34) symptom trackers, 6% (2/34) web-based counseling services, 6% (2/34) planning and time management tools, 6% (2/34) psychoeducational websites, and 2% (1/34) mindfulness and meditation app. The remaining apps and e-tools supported fitness (2/34, 6%) and relationships (1/34, 2%).

Summary of Privacy Policy Information

Overview

The summary results from the review of privacy policies are presented in Table 1. All apps and e-tools had privacy policies. Some of the policies were not readily accessible directly from the app or e-tool but rather were hosted on an external website that the app or e-tool privacy policy fell under (ie, a privacy policy for a hospital or government department). Importantly, most of the policies (26/34, 76%) explicitly stated that they met the standards of the Privacy Act 1988 (Australia) or international equivalent (ie, Health Insurance Portability and Accountability Act). In addition, most manufacturers (31/34, 91%) introduced the purpose of the privacy policy, stating that the policy explained the manufacturers' approach to privacy, protection, and management of personal information. Similarly, 97% (33/34) of privacy policies were noted to provide adequate information for all potential users. In contrast, manufacturers (26/34, 76%) frequently did not provide adequate information about their organization and how or why they operated.

Table 1. Summary results of the privacy policy assessment results (N=34).

Privacy policy questions and responses	Value, n (%)
Is there a privacy policy?	
No	0 (0)
Yes	34 (100)
Does the app or e-tool claim to meet the standards of the Privacy Act 1988 (Australia), HIPAA[a] (United States), or another international equivalent?	
No	8 (24)
Yes	26 (76)
Did the manufacturer introduce the purpose of the privacy policy?	
No	3 (9)
Yes	31 (91)
Does the privacy policy provide an introduction to the organization, including its vision and purpose?	
No	26 (76)
Some information	3 (9)
Yes	5 (15)
Does the privacy policy provide adequate information (both targeted and general) relevant for all users, including consumers seeking care and health professionals?	
No	0 (0)
Some information	1 (3)
Yes	33 (97)

[a]HIPAA: Health Insurance Portability and Accountability Act.

Readability of Privacy Policies

In relation to the Flesch reading ease, the privacy policies were all found to fall into the top three of the seven available score categories, with 9% (3/34) rated as *fairly difficult to read*, an additional 53% (18/34) rated as *difficult to read*, and the remaining 38% (13/34) rated as *very difficult to read*. The

Flesh-Kincaid grade-level test illustrated similar results with only the top two levels represented out of eight possible levels. All but three policies fell into the *college graduate* grade level (31/34, 91%), with the remaining considered *professional-level* reading difficulty (3/34, 9%).

Collection of PHI

Table 2 presents results from the PHI assessment. Most apps and e-tools (32/34, 94%) collected some form of PHI, ranging from simple demographic information to more sensitive information such as information related to mental health. The remaining 6% (2/34) of apps and e-tools were considered informational tools and collected data such as email addresses. Of the 32 apps and e-tools that collected PHI, most (28/32, 87%) shared this information in some manner (data sharing is reviewed in greater detail below). This use was disclosed in most privacy policies (27/32, 85%), explicitly stating that reasonable steps were taken to ensure the security of the PHI; however, for 6% (2/32) and 9% (3/32) of privacy policies how PHI was shared was *somewhat clear* or *not clear*, respectively. Although more than half of the privacy policies (17/32, 53%) clearly stated how and when PHI was deleted, this information was either *somewhat clear* or *not clear* in the remaining 3% (1/32) or 44% (14/32) privacy policies, respectively.

Table 2. PHI[a] assessment results (N=34).

PHI questions and responses	Value, n (%)
Does the app or e-tool collect PHI (ie, demographic information, medical histories, test and laboratory results, mental health conditions, or insurance information)?	
No	2 (6)
Yes	32 (94)
Is PHI (ie, demographic information, medical histories, test and laboratory results, mental health conditions, or insurance information) shared?	
No	4 (13)
Yes	28 (87)
Is it clear if the organization has taken reasonable steps to ensure the security of PHI?	
Not clear	3 (9)
Somewhat clear	2 (6)
Yes—clear	27 (85)
Is it clear how and when their organization will delete PHI?	
No	14 (44)
Somewhat clear	1 (3)
Yes	17 (53)

[a]PHI: personal health information.

Data Sharing and Use of Information

Results of the assessment questions related to data sharing, data preferences, and data storage are presented in Table 3. Most privacy policies (32/34, 94%) declared how data were used and for what purposes, with approximately two-thirds (22/34, 65%) stating that users were allowed to delete data. Notably, only 3% (1/34) of apps allowed data sharing preferences to be changed, although this required the user to contact the manufacturer via email; 9% (3/34) of apps and e-tools enabled users with some permissions specific to data sharing, such as receiving push notifications or allowing the user to choose whether to share data such as location with the mobile app.

Table 3. Data sharing and data use assessment results (N=34).

Data sharing and data use questions and responses	Value, n (%)
Does the privacy policy declare data use and purpose?	
No	2 (6)
Yes	32 (94)
Is shared data deidentified? (ie, is data anonymous—is personal information masked or severed from the identity of the contributor)	
No	11 (33)
Yes	23 (67)
Can the user change their preferences regarding data sharing (ie, switch it on or off)?	
No	30 (88)
Yes	1 (3)
Some permissions	3 (9)
Can the user delete their data from the app or e-tool?	
No	10 (29)
Yes	22 (65)
Informational webpage only	2 (6)
Is user data stored on the device?	
No	27 (79)
Yes	4 (12)
Unspecified	3 (9)
Is user data stored on a server?	
Yes	32 (94)
Unspecified	2 (6)
For how long is user data stored?	
Unspecified or unclear	12 (35)
Until no longer needed	12 (35)
1 year or less	2 (6)
1-3 years	3 (9)
More than 3 years	3 (9)
At user discretion	2 (6)
What type of server is used to store user data? (eg, Amazon Web Services, within Australian borders)	
Secure Australian server	6 (19)
Secure overseas server	11 (34)
Unspecified server	7 (22)
University server	2 (6)
Hospital or PHN[a]	2 (6)
Unclear	4 (13)
In what country is the server located ?	
Australia	13 (38)
United States	9 (26)
Canada	3 (9)
Europe	1 (3)
Multiple countries	2 (6)
Unclear or unspecified	6 (18)

[a]PHN: Primary Health Network.

Most apps and e-tools (32/34, 94%) stored data on a server, with a small number (4/34, 12%) storing data on both the device (ie, PHI) and a server (ie, email address and website activity). In addition, 6% (2/34) of apps and e-tools did not specify where the data were stored. Data storage duration ranged from up to 1 year (2/34, 6%), 1 to 3 years (3/34, 9%), and more than 3 years (3/34, 9%). Although more than one-third of the apps and e-tools were unclear or did not specify for how long data were stored (12/34, 35%), approximately one-third (12/34, 35%) stored the data until no longer needed by the organization. The remaining 6% (2/34) of privacy policies stated that the data would be deleted at the user's discretion.

Of the 32 apps and e-tools that stored data on a server, 11 (34%) stored data on a secure overseas server, 6 (19%) stored data on a secure Australian server, 7 (22%) stored data on an unnamed or unspecified server. In addition, 6% (2/32) of apps and e-tools stored data on a university server, 6% (2/32) stored data on a hospital or primary health network server system, and 13% (4/32) were unclear on the type of server used. The location of data storage was mixed between Australia (13/34, 38%), the United States (9/34, 26%), Canada (3/34, 9%), Europe (1/34, 3%), and multiple locations (2/34, 6%), with the remainder unclear as to where data were stored (6/34, 18%).

Review of How Data Are Shared

Most apps and e-tools (27/34, 79%) shared data with relevant third parties, including but not limited to partners, suppliers, collaborators, advisers, and business associates. The types of data shared varied from PHI to aggregated user data, such as location. A small number of apps and e-tools (4/34, 12%) shared information with irrelevant third parties, including social media platforms such as Facebook. In addition, 38% (13/34) of apps and e-tools shared data with a research partner or university, 15% (5/34) shared information with government departments, and 38% (13/34) shared data with a health-related group or person (eg, a support person or health professional).

Additional Information

All privacy policies were assessed for their inclusion of various other details, which are summarized in Table 4. More than half of the apps and e-tools (20/34, 59%) used third-party vendors, such as Google Analytics, to evaluate and track consumer use of websites, collect demographic data, and evaluate other information related to the apps or e-tool website and the user's device. Less than one-quarter of privacy policies (8/34, 24%) adequately explained how the manufacturer would respond to a data breach, although most policies (31/34, 91%) provided some details as to how to provide feedback or lodge a complaint either with the manufacturer or through an expert third party (ie, Office of the Australian Information Commissioner). Finally, only 24% (8/34) of privacy policies explicitly warned individuals about privacy risks involved in accessing services that are outside the control of the service provider, such as third-party advertisements, with an additional 9% (3/34) of policies providing some detail in this regard.

Table 4. Overview of additional details provided by privacy policies (N=34).

Additional questions and responses	Value, n (%)
Does the app or e-tool use third-party vendors (eg, Google Analytics)?	
No	11 (32)
Yes	20 (59)
Unspecified	3 (9)
Is the manner in which the organization will respond to a data breach adequately explained?	
No	26 (76)
Yes	8 (24)
Does the privacy policy inform users how they can make inquiries and provide feedback and lodge complaints, including both contact details for the relevant party within the organization and a third-party expert (eg, Office of the Australian Information Commissioner)?	
No—does not provide either	3 (9)
Some—provides information for internal or third party only	16 (47)
Yes—provides both internal and third-party expert	15 (44)
Does the privacy policy explicitly warn users about privacy risks involved in accessing services offered that are outside the control of the organization ?	
No	23 (67)
Somewhat	3 (9)
Yes	8 (24)

Privacy Considerations for Health Services

As generated through the collaborative consultation process described previously, Textbox 1 presents privacy guidelines to assist health professionals and service providers to ask the appropriate questions—of themselves and to HIT manufacturers—before confidently promoting the safe use of HITs as part of mental health care.

Textbox 1. Privacy guidelines—health professional and service provider considerations regarding the use of health information technologies (HITs) for care.

Privacy guidelines

- The HIT manufacturer has clearly introduced the purpose of its privacy policy.
- The privacy policy includes an introduction to the HIT manufacturer and how and why their organization operates.
- The privacy policy provides adequate information and addresses my concerns.
  - If no, I am aware who I need to contact to seek clarification...
- The privacy policy is written clearly.
  - If no, I am aware who I need to contact to seek clarification...
- The privacy policy adequately explains how the HIT manufacturer will collect and use personal data.
- The privacy policy adequately explains how and when the HIT manufacturer will disclose personal data to third parties.
  - If the HIT manufacturer shares data with third parties, I am confident that the third-party partners are reputable and will comply with all legislative requirements when collecting, storing, and sharing data.
  - If no, I am aware who I need to contact to seek clarification...
- I am confident the HIT manufacturer has taken the appropriate steps to protect everyone's data, adopting the strongest security measures.
- I have been made aware of how end users can access and correct their personal information on the HIT.
  - If no, I am aware who I need to contact to seek clarification...
- The privacy policy outlines how and when the HIT manufacturer will delete personal data.
- The privacy policy outlines how the HIT manufacturer will respond to any data breaches.
- The privacy policy includes information on how I can inquire, provide feedback, or make complaints.
- There is the opportunity for me to contact a third-party expert to inquire about the privacy policy (such as the Office of the Australian Information Commissioner).
  - If yes, they are...
- From what I read, I feel comfortable using the proposed HIT as part of my clinical care and/or practice.

Discussion

Listening to Consumers

With their increased experience and exposure, consumers are becoming more sophisticated users of HITs. They can offer valuable insights into how privacy information should be presented to ensure that it is clear, informative, and transparent. Participants of our co-design workshops highlighted the need for all HITs to have a privacy policy that provides relevant data security information before collecting information from an individual, preceding the account creation process. In addition, participants stated that privacy policies should be accessible, transparent, and user-friendly, ensuring that consumers understand what personal and health data will be collected, stored, and shared and, in turn, enabling them to trust the HIT to protect their data. These findings align with those reported by Schueller et al [31], which indicated that 70% of mental health app users rate the inclusion of a privacy policy and data encryption as important. Of note, when a mental health app was deemed to be from a *trusted source* (no shared definition of what constituted a trusted source), users assumed that the app adequately protected their data [31], potentially leaving their data vulnerable to unrecognized data sharing pathways.

Privacy Policy Risk Assessment

As evidenced by our co-design results, consumers are calling for greater clarity and transparency in the privacy policies of HITs so that they can be confident that they understand how their personal and health information may be used. Importantly, all apps and e-tools included in this study had a privacy policy. All but 1 of those policies provided explicit details explaining the manufacturer's approach to privacy and how personal information is managed and protected. However, approximately one-quarter of the privacy policies did not meet the standards of the Australian Privacy Act 1988 or other international equivalents, raising concerns regarding undisclosed data sharing and poorly secured data storage. Even when the use of data adheres to privacy standards, issues of transparency often arise. For example, a recent review of the data sharing practices of 24 health-related apps found that data were shared with 55 unique entities, including app developers, their parent companies, and third parties (ie, service providers). Subsequently, third parties shared user data with an additional 216 integrated fourth parties (eg, Facebook sharing data with data brokers to enable targeted advertising) [32].

In addition to poor overall transparency, our results also confirm that privacy policies, when present, are fairly difficult to read

and require a college or professional reading level, essentially rendering them useless for a large portion of potential users (eg, children and young people or individuals with cognitive impairments or intellectual disabilities). This aligns with previous research by Robillard et al [33], which highlighted that the readability of privacy policies of mental health apps is typically written at too high a level for the general population. Strikingly, a longitudinal review of privacy policies found both a decline in readability and a marked increase in length [34]. To improve the readability of privacy policies, it is suggested that HIT manufacturers either compare different versions of their policy to determine which one has the best readability score or that the policy be rewritten until it meets a predetermined grade level. Researchers have suggested that a grade 9-10 reading level is likely to be appropriate for the general population [35].

Most apps and e-tools included in this study collected some form of PHI (32/34, 94%), including, in some cases, information related to mental health, with 87% (28/34) of those apps and e-tools then sharing these data in some manner. Although data sharing was disclosed in most privacy policies (27/32, 84%), how PHI was shared was not transparent in 15% (5/32) of the policies. Most apps and e-tools shared data with relevant third parties (27/34, 79%); however, 12% (4/34) also shared information with third parties deemed to be irrelevant, such as social media platforms (ie, Facebook). Of note, few apps and e-tools (4/31, 13%) allowed users to update their permissions concerning data sharing.

Although it is unlawful in Australia, for example, to share PHI for purposes other than those stated in a privacy policy, the complexities of the web-based environment frequently preclude a full understanding of how data are shared and for what purpose [36]. Users must recognize that shared data enters into a *supply chain*, being passed from apps and e-tools to parent companies and then on to third parties such as data trackers, aggregators, and brokers [37]. Subsequently, data may be sold to researchers and government agencies for advertising purposes. In addition to driving targeted advertising campaigns, this aggregated data may also be used to influence employment and financial and insurance decisions, with potentially marked consequences at the individual level [37], potentially leading to incarceration or human rights abuses in some countries [11]. Given the risks described above, explicit and transparent documentation of how data are shared is critical to ensure that users are able to provide informed consent. Furthermore, it is critical that governance structures and regulatory standards are established globally to ensure ethical practice in digital mental health care, including the handling of PHI. As advised by the World Economic Forum, regulations should not be designed exclusively by governments but rather in collaboration with consumers with lived experience, mental health professionals, technology manufacturers, and policy makers, with the aim of facilitating efficient access to effective and safe digital tools to address growing mental health needs [11].

Most apps and e-tools in our sample stored data on a server (32/34, 94%), with more than half (21/32, 66%) storing data on an unnamed or unspecified server. Although other server types included university servers (2/32, 6%) and a hospital or primary health network server system (2/32, 6%), the type of server used to store data for an additional 22% (7/32) of apps and e-tools was unclear. Once data are transmitted to a third-party server, it is often difficult to determine the robustness of the privacy and confidentiality standards in place to protect the PHI. For example, Cultura Colectiva, a digital media company with access to user information from Facebook, stored data on a publicly accessible server, resulting in the exposure of 540 million individual records, including user IDs and names [38]. In addition, although it is ideal that all information being transmitted is encrypted, personal and health information may still be visible in server logs with few restrictions in terms of access [39]. Notably, when different apps and e-tools use the same server, it may be possible to link different PHIs together to create digital profiles of users [40] with potentially negative repercussions such as impacts on employment and insurance. Despite the frequency of data breaches [41], 76% (26/34) of the privacy policies reviewed in this study failed to document how the organization would respond to data breaches, leaving the user to wonder what steps might be taken to protect their data from exposure and misuse (ie, identify theft).

Privacy Guidelines for Health Professionals and Service Providers

Given the concerns regarding accessibility, transparency, and readability outlined above, through a consultative process with key stakeholders, our team developed privacy guidelines (Textbox 1) to prompt health professionals and service providers to ask informed questions when reviewing an HIT privacy policy to ensure adequate data privacy and security measures are in place. The guidelines aim to support health professionals and service providers to confidently promote the safe use of HITs as part of care and within the broader service. The privacy guidelines have considered the privacy policy of HIT manufacturers to emphasize the importance of building trust between users and HIT manufacturers through transparency [42]. Thus, it is the responsibility of the HIT manufacturer to be aware of all (both current and emerging) regulatory requirements and best practice principles [42] to ensure that the privacy policy is communicated to all users. This not only minimizes potential harm but also allows users to be well informed and to have more control when consenting to use HITs.

As few consumers will review academic literature before accessing HITs, they are more likely to learn about available apps and e-tools via the internet, app stores, social media, word of mouth, or health professionals. In relation to the latter, it is recommended that health professionals and service providers conduct their own risk assessment before implementing HITs into their service to ensure appropriate risk strategies are in place [3]. Not only should providers have an understanding of privacy risks but it is also important that they work with consumers to ensure they are aware of the potential for privacy breaches to ensure they are providing informed consent before engaging with an HIT [37].

Limitations

This study has some limitations that are important to note. Although we engaged in a thorough collaborative consultation

process to develop the broad structure and content of the privacy guidelines for Project Synergy, the development of the privacy policy risk assessment tool and the guidelines for health professionals and service providers was conducted by the research team, independent of this broader stakeholder group. Therefore, we acknowledge that both the assessment tool and the guidelines may benefit from further input or revision by individuals with expertise in data privacy and security, both from a legal perspective and regarding manufacturers of digital tools. As highlighted by the co-design work presented in phase 1, our group recognizes the importance of including the voice of those with lived experience in our work to reform mental health services and systems of care, including the ethical development and application of digital tools. With that being said, we acknowledge that the privacy policy risk assessment tool and guidelines were developed without contributions from consumers with lived experience of mental ill health.

Conclusions

Given the increasing uptake of HITs, both by individuals for the purposes of self-management and by health professionals as a means to complement clinical services, it is essential that all users have a clear understanding of what personal and health information will be collected, how these data will be shared and stored, and what privacy and security measures are in place to ensure it is protected. Our findings highlight the ubiquitous poor readability and lack of transparency in existing privacy policies, a stark contrast to what consumers emphasized as essential factors in the presentation of privacy information. Although consumers, health professionals, and services are becoming increasingly reliant on HITs to deliver, support, or enhance care, concerns regarding the privacy of health and personal information are likely to undermine user confidence and willingness to engage with HITs. Therefore, we provide suggested guidelines that can be easily adopted by health professionals and service providers when considering the implementation of HITs, including apps and e-tools, into their service. We recommend that these guidelines be adopted to ensure that HITs are used to their full potential to maximize patient health outcomes while minimizing risk and that users are informed of privacy and security considerations to make educated decisions as to whether they would like to share their personal and health information.

Acknowledgments

The authors IBH and TAD were integral in securing funding to support this study. This study was designed by HML, AER, TAD, and GL. All data analyses were conducted by HML and AER. All authors have contributed to and approved the final manuscript. This research was conducted on behalf of the Australian Government Department of Health as part of Project Synergy (2017-20). InnoWell was formed by the University of Sydney and PricewaterhouseCoopers (Australia) to deliver the Aus $30 million (US $22.1 million) Australian government-funded Project Synergy. The authors would also like to acknowledge Future Generation Global for funding the Youth Mental Health & Technology Program, which aims to improve young people's access to quality mental health care, including through the use of co-design methodologies. The authors would like to acknowledge and thank Hannah Yee for her contributions to the knowledge translation and Toby Wong for his assistance with data collection. The authors would also like to thank the participants in the co-design workshops for their insights on how to best present privacy information to users.

Conflicts of Interest

IBH is the codirector of health and policy at the Brain and Mind Centre, University of Sydney. The Brain and Mind Centre operates an early-intervention youth service at Camperdown under contract to headspace. He is the chief scientific advisor to and a 5%-equity shareholder in InnoWell Pty Ltd. InnoWell was formed by the University of Sydney (45% equity) and PricewaterhouseCoopers (Australia; 45% equity) to deliver the Aus $30 million (US $22.1 million) Australian government-funded Project Synergy (2017-2020; a 3-year program for the transformation of mental health services) and to lead the transformation of mental health services internationally through the use of innovative technologies. TAD is now the director of research and evaluation at the Design and Strategy Division of the Australian Digital Health Agency. The funding source does not entail any potential conflicts of interest for the other members of the Project Synergy research and development team.

Multimedia Appendix 1
Privacy risk assessment tool.
[DOCX File , 16 KB - jmir_v23i9e26317_app1.docx]

References

1. Patient-centric technology improves access, efficiency, and quality of care. Wolters Kluwer. 2018 Apr 24. URL: https://www.wolterskluwer.com/en/expert-insights/patient-centric-technology-improves-access-efficiency-and-quality-of-care [accessed 2021-08-06]
2. Health technology in Australia. Australian Government Department of Health. URL: https://www.health.gov.au/health-topics/health-technology [accessed 2021-08-06]

3. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. BMC Public Health 2020 Aug 01;20(1):1193 [FREE Full text] [doi: 10.1186/s12889-020-09301-4] [Medline: 32738884]
4. Davenport TA, Cheng VW, Iorfino F, Hamilton B, Castaldi E, Burton A, et al. Flip the clinic: a digital health approach to youth mental health service delivery during the covid-19 pandemic and beyond. JMIR Ment Health 2020 Dec 15;7(12):e24578 [FREE Full text] [doi: 10.2196/24578] [Medline: 33206051]
5. Iorfino F, Cross SP, Davenport T, Carpenter JS, Scott E, Shiran S, et al. A digital platform designed for youth mental health services to deliver personalized and measurement-based care. Front Psychiatry 2019 Aug 23;10:595 [FREE Full text] [doi: 10.3389/fpsyt.2019.00595] [Medline: 31507465]
6. National Safety and Quality Digital Mental Health (NSQDMH) standards. Australian Commission on Safety and Quality in Health Care. 2020 May 29. URL: https://www.safetyandquality.gov.au/sites/default/files/2020-04/UPDATED%20national_safety_and_quality_digital_mental_health_standards_-_consultation_draft_-_feb_2020_0.pdf [accessed 2020-10-16]
7. Tips for choosing a digital mental health service - information for consumers and carers. Australian Commission on Safety and Quality in Health Care. 2020. URL: https://www.safetyandquality.gov.au/sites/default/files/2020-05/tips_for_choosing_a_digital_mental_health_service_-_consumers_and_carers_fact_sheet_-_may_2020.pdf [accessed 2020-10-16]
8. Tips for choosing a digital mental health service - information for clinicians. Australian Commission on Safety and Quality in Health Care. 2020. URL: https://www.safetyandquality.gov.au/sites/default/files/2020-05/tips_for_choosing_a_digital_mental_health_service_-_clinicians_fact_sheet_-_may_2020.pdf [accessed 2020-10-16]
9. Read the Australian privacy principles. Australian Government Office of the Australian Information Commissioner. 2014. URL: https://www.oaic.gov.au/privacy/australian-privacy-principles/read-the-australian-privacy-principles/ [accessed 2021-08-06]
10. Privacy Act 1988. Australian Government Federal Register of Legislation. 1988. URL: https://www.legislation.gov.au/Details/C2018C00456 [accessed 2020-08-28]
11. Empowering 8 billion minds enabling better mental health for all via the ethical adoption of technologies. World Economic Forum. 2019 Jul. URL: http://www3.weforum.org/docs/WEF_Future%20Council_Mental_Health_and_Tech_Report.pdf [accessed 2021-08-06]
12. Granwal L. Penetration rate of health applications on mobiles and tablets in Australia 2014-2018. Statista. 2020 Apr 09. URL: https://www.statista.com/statistics/985116/use-health-apps-mobile-tablet-australia/#statisticContainer [accessed 2020-10-16]
13. Byambasuren O, Beller E, Glasziou P. Current knowledge and adoption of mobile health apps among australian general practitioners: survey study. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e13199 [FREE Full text] [doi: 10.2196/13199] [Medline: 31199343]
14. Huckvale K, Prieto JT, Tilney M, Benghozi P, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med 2015 Sep 07;13:214 [FREE Full text] [doi: 10.1186/s12916-015-0444-y] [Medline: 26404673]
15. Huckvale K, Torous J, Larsen ME. Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation. JAMA Netw Open 2019 Apr 05;2(4):e192542 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.2542] [Medline: 31002321]
16. Hickie IB, Davenport TA, Burns JM, Milton AC, Ospina-Pinillos L, Whittle L, et al. Project Synergy: co-designing technology-enabled solutions for Australian mental health services reform. Med J Aust 2019 Oct;211 Suppl 7:3-39. [doi: 10.5694/mja2.50349] [Medline: 31587276]
17. LaMonica HM, Davenport TA, Braunstein K, Ottavio A, Piper S, Martin C, et al. Technology-enabled person-centered mental health services reform: strategy for implementation science. JMIR Ment Health 2019 Sep 19;6(9):e14719 [FREE Full text] [doi: 10.2196/14719] [Medline: 31538938]
18. Bjerkan J, Hedlund M, Hellesø R. Patients' contribution to the development of a web-based plan for integrated care - a participatory design study. Inform Health Soc Care 2015 Mar;40(2):167-184. [doi: 10.3109/17538157.2014.907803] [Medline: 24786524]
19. Davenport TA, Milton AC, Ospina-Pinillos L, Whittle L, Ricci CS, Burns JM. Project Synergy research and development cycle: iterative processes of participatory design, user testing, implementation and feasibility testing. Med J Aust 2019 Oct 07;211(7 Suppl):8-11 [FREE Full text]
20. LaMonica HM, Davenport TA, Burns J, Cross S, Hodson S, Veitch J, et al. Technology-enabled mental health service reform for open arms - veterans and families counselling: participatory design study. JMIR Form Res 2019 Sep 19;3(3):e13662 [FREE Full text] [doi: 10.2196/13662] [Medline: 31538937]
21. Davenport T, Ospina-Pinillos L, Ricci C, Milton A, Burns J, Hickie I. Fit Uni Life to thrive: an online health and wellbeing prototype for young people. Med J Aust 2019 Oct 07;211(7 Suppl):12-16 [FREE Full text]

22. Ospina-Pinillos L, Davenport TA, Ricci CS, Milton AC, Scott EM, Hickie IB. Developing a mental health eClinic to improve access to and quality of mental health care for young people: using participatory design as research methodologies. J Med Internet Res 2018 May 28;20(5):e188 [FREE Full text] [doi: 10.2196/jmir.9716] [Medline: 29807878]
23. LaMonica HM, Davenport TA, Roberts AE, Hickie IB. Understanding technology preferences and requirements for health information technologies designed to improve and maintain the mental health and well-being of older adults: participatory design study. JMIR Aging 2021 Jan 06;4(1):e21461 [FREE Full text] [doi: 10.2196/21461] [Medline: 33404509]
24. Cheng VW, Piper SE, Ottavio A, Davenport TA, Hickie IB. Recommendations for designing health information technologies for mental health drawn from self-determination theory and co-design with culturally diverse populations: template analysis. J Med Internet Res 2021 Feb 10;23(2):e23502 [FREE Full text] [doi: 10.2196/23502] [Medline: 33565985]
25. Torous JB, Chan SR, Gipson SY, Kim JW, Nguyen T, Luo J, et al. A hierarchical framework for evaluation and informed decision making regarding smartphone apps for clinical care. Psychiatr Serv 2018 May 01;69(5):498-500. [doi: 10.1176/appi.ps.201700423] [Medline: 29446337]
26. Davenport TA, LaMonica HM, Whittle L, English A, Iorfino F, Cross S, et al. Validation of the Innowell Platform: Protocol for a clinical trial. JMIR Res Protoc 2019 May 31;8(5):e13955 [FREE Full text] [doi: 10.2196/13955] [Medline: 31152524]
27. Wang L, Miller MJ, Schmitt MR, Wen FK. Assessing readability formula differences with written health information materials: application, results, and recommendations. Res Social Adm Pharm 2013;9(5):503-516. [doi: 10.1016/j.sapharm.2012.05.009] [Medline: 22835706]
28. Kincaid J, Fishburne JR, Rogers R, Chissom B. Derivation of New Readability Formulas: (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for navy enlisted personnel. Institute for Simulation and Training, University of Central Florida, Millington. 1975. URL: https://stars.library.ucf.edu/istlibrary/56/?utm_source=stars.library.ucf.edu%2Fistlibrary%2F56&utm_medium=PDF&utm_campaign=PDFCoverPages [accessed 2021-08-24]
29. Automatic readability checker. Readability Formulas. 2021. URL: https://readabilityformulas.com/free-readability-formula-tests.php [accessed 2021-08-12]
30. Robinson J, Donaldson A, Hetrick S, Rice S, Bailey E. Guidelines for Project Synergy: Development of a set of medico-legal ethical guidelines. Young and Well CRC, Melbourne. 2016. URL: https://www.youngandwellcrc.org.au/ [accessed 2021-08-25]
31. Schueller SM, Neary M, O'Loughlin K, Adkins EC. Discovery of and interest in health apps among those with mental health needs: survey and focus group study. J Med Internet Res 2018 Jun 11;20(6):e10141 [FREE Full text] [doi: 10.2196/10141] [Medline: 29891468]
32. Grundy Q, Chiu K, Held F, Continella A, Bero L, Holz R. Data sharing practices of medicines related apps and the mobile ecosystem: traffic, content, and network analysis. Br Med J 2019 Mar 20;364:l920 [FREE Full text] [doi: 10.1136/bmj.l920] [Medline: 30894349]
33. Robillard JM, Feng TL, Sporn AB, Lai J, Lo C, Ta M, et al. Availability, readability, and content of privacy policies and terms of agreements of mental health apps. Internet Interv 2019 Mar 06;17:100243 [FREE Full text] [doi: 10.1016/j.invent.2019.100243] [Medline: 30949436]
34. Milne GR, Culnan MJ, Greene H. A longitudinal assessment of online privacy notice readability. J Public Policy Mark 2006 Sep 1;25(2):238-249. [doi: 10.1509/jppm.25.2.238]
35. Krumay B, Klar J. Readability of privacy policies. In: Singhal A, Vaidya J, editors. Data and Applications Security and Privacy XXXIV. Cham: Springer; 2020:388-399.
36. Mobile privacy: a better practice guide for mobile app developers. Office of the Australian Information Commission. 2014 Sep 05. URL: https://www.oaic.gov.au/agencies-and-organisations/guides/guide-for-mobile-app-developers [accessed 2020-10-22]
37. Martin K. Ethical issues in the big data industry. In: Strategic Information Management. Oxfordshire, England: Routledge; Apr 2020:450-471.
38. Lecher C. Facebook app developers leaked millions of user records on cloud servers, researchers say. The Verge. 2019 Apr 03. URL: https://www.theverge.com/2019/4/3/18293978/facebook-app-developers-leak-user-records-data-cloud-servers [accessed 2020-10-23]
39. He D, Naveed M, Gunter CA, Nahrstedt K. Security concerns in android mHealth apps. AMIA Annu Symp Proc 2014 Nov 14;2014:645-654 [FREE Full text] [Medline: 25954370]
40. Vallina-Rodriguez N, Sundaresan S. 7 in 10 smartphone apps share your data with third-party services. The Conversation. 2017 May 30. URL: https://theconversation.com/7-in-10-smartphone-apps-share-your-data-with-third-party-services-72404 [accessed 2021-08-06]
41. 2019 healthcare data breach report. HIPAA Journal. URL: https://www.hipaajournal.com/2019-healthcare-data-breach-report/ [accessed 2020-10-23]
42. Parker L, Karliychuk T, Gillies D, Mintzes B, Raven M, Grundy Q. A health app developer's guide to law and policy: a multi-sector policy analysis. BMC Med Inform Decis Mak 2017 Oct 02;17(1):141 [FREE Full text] [doi: 10.1186/s12911-017-0535-0] [Medline: 28969704]

Abbreviations

HIT: health information technology
NSW: New South Wales
PHI: personal health information
PwC: PricewaterhouseCoopers

Edited by R Kukafka; submitted 07.12.20; peer-reviewed by N Shen, L Ospina-Pinillos; comments to author 05.02.21; revised version received 15.02.21; accepted 04.05.21; published 16.09.21.

Please cite as:
LaMonica HM, Roberts AE, Lee GY, Davenport TA, Hickie IB
Privacy Practices of Health Information Technologies: Privacy Policy Risk Assessment Study and Proposed Guidelines
J Med Internet Res 2021;23(9):e26317
URL: https://www.jmir.org/2021/9/e26317
doi:10.2196/26317
PMID:34528895

Original Paper

Association Between What People Learned About COVID-19 Using Web Searches and Their Behavior Toward Public Health Guidelines: Empirical Infodemiology Study

Ikpe Justice Akpan[1*], BSc, MSc, PhD; Obianuju Genevieve Aguolu[2*], MD, MPH, PhD; Yawo Mamoua Kobara[3*], BSc, MSc; Rouzbeh Razavi[4*], BSc, MSc, PhD; Asuama A Akpan[5*], BSc, MA; Murali Shanker[4*], BSc, MSc, PhD

[1]Department of Management & Information Systems, Kent State University, New Philadelphia, OH, United States

[2]Infectious Disease Internal Medicine Department, Yale School of Medicine, Yale University, New Haven, CT, United States

[3]Statistical and Actuarial Sciences, Western University, London, ON, Canada

[4]Department of Management & Information Systems, Kent State University, Kent, OH, United States

[5]Research and Development, Ibom International Center for Research and Scholarship, Windsor, ON, Canada

[*]all authors contributed equally

Corresponding Author:
Ikpe Justice Akpan, BSc, MSc, PhD
Department of Management & Information Systems
Kent State University
330 University Drive Northeast
New Philadelphia, OH, 44663
United States
Phone: 1 3303393391 ext 7572
Fax: 1 3303393321
Email: iakpan@kent.edu

Abstract

Background: The use of the internet and web-based platforms to obtain public health information and manage health-related issues has become widespread in this digital age. The practice is so pervasive that the first reaction to obtaining health information is to “Google it.” As SARS-CoV-2 broke out in Wuhan, China, in December 2019 and quickly spread worldwide, people flocked to the internet to learn about the novel coronavirus and the disease, COVID-19. Lagging responses by governments and public health agencies to prioritize the dissemination of information about the coronavirus outbreak through the internet and the World Wide Web and to build trust gave room for others to quickly populate social media, online blogs, news outlets, and websites with misinformation and conspiracy theories about the COVID-19 pandemic, resulting in people’s deviant behaviors toward public health safety measures.

Objective: The goals of this study were to determine what people learned about the COVID-19 pandemic through web searches, examine any association between what people learned about COVID-19 and behavior toward public health guidelines, and analyze the impact of misinformation and conspiracy theories about the COVID-19 pandemic on people’s behavior toward public health measures.

Methods: This infodemiology study used Google Trends’ worldwide search index, covering the first 6 months after the SARS-CoV-2 outbreak (January 1 to June 30, 2020) when the public scrambled for information about the pandemic. Data analysis employed statistical trends, correlation and regression, principal component analysis (PCA), and predictive models.

Results: The PCA identified two latent variables comprising past coronavirus epidemics (pastCoVepidemics: keywords that address previous epidemics) and the ongoing COVID-19 pandemic (presCoVpandemic: keywords that explain the ongoing pandemic). Both principal components were used significantly to learn about SARS-CoV-2 and COVID-19 and explained 88.78% of the variability. Three principal components fuelled misinformation about COVID-19: misinformation (keywords “biological weapon,” “virus hoax,” “common cold,” “COVID-19 hoax,” and “China virus”), conspiracy theory 1 (ConspTheory1; keyword “5G” or “@5G”), and conspiracy theory 2 (ConspTheory2; keyword “ingest bleach”). These principal components explained 84.85% of the variability. The principal components represent two measurements of public health safety guidelines—public health measures 1 (PubHealthMes1; keywords “social distancing,” “wash hands,” “isolation,” and “quarantine”) and public health measures 2 (PubHealthMes2; keyword “wear mask”)—which explained 84.7% of the variability. Based on the PCA results and

the log-linear and predictive models, ConspTheory1 (keyword "@5G") was identified as a predictor of people's behavior toward public health measures (PubHealthMes2). Although correlations of misinformation (keywords "COVID-19," "hoax," "virus hoax," "common cold," and more) and ConspTheory2 (keyword "ingest bleach") with PubHealthMes1 (keywords "social distancing," "hand wash," "isolation," and more) were r=0.83 and r=–0.11, respectively, neither was statistically significant (P=.27 and P=.13, respectively).

Conclusions: Several studies focused on the impacts of social media and related platforms on the spreading of misinformation and conspiracy theories. This study provides the first empirical evidence to the mainly anecdotal discourse on the use of web searches to learn about SARS-CoV-2 and COVID-19.

(J Med Internet Res 2021;23(9):e28975) doi:10.2196/28975

KEYWORDS

internet; novel coronavirus; SARS-CoV-2; COVID-19; infodemiology; misinformation; conspiracy theories; public health

Introduction

Overview

A novel coronavirus initially named 2019-nCoV emerged in Wuhan, China, and was formally reported to the World Health Organization (WHO) on December 31, 2019 [1-3]. Further scientific evidence soon unveiled the semblance of the 2019-nCoV's genome sequence to a previous epidemic, the severe acute respiratory syndrome (SARS), a disease epidemic caused by SARS-CoV, which broke out in Foshan, China, in 2002 [4,5]. Some initial studies also identified similar features that were related to the Middle East respiratory syndrome (MERS) epidemic caused by MERS-CoV as the causative agent [6].

The outbreak was originally named 2019-nCoV on January 13, 2020, the same day that the first imported case outside China occurred in the Philippines and other countries [7]. The spread of 2019-nCoV continued across many countries, causing the WHO to declare the outbreak a pandemic [6]. The 2019-nCoV was later renamed SARS-CoV-2 and identified as the causative agent of COVID-19 in February 2020 [8,9]. The highly contagious COVID-19 spread rapidly globally and caught the world unprepared. With no adequately planned health communication strategies, panic ensued, while confirmed cases of infections and deaths from COVID-19 increased rapidly worldwide [3]. The public rushed to internet platforms to learn about the outbreak through Google searches, online news outlets, and social media platforms [10-15].

In March 2020, the WHO launched a free online introductory training course in different languages, including English, French, Spanish, and Chinese, to make the public aware of the contagious COVID-19 [6]. However, it is unclear how many people knew about or used the free training lessons about COVID-19 that the WHO had made available via its website [6]. Instead, several studies suggest that the public flocked to the internet to learn about SARS-CoV-2 and COVID-19 through web searches, online news outlets, and social media [16,17]. Analyzing how people search and navigate the World Wide Web and other internet platforms for health-related information can provide valuable insights into the health-related behavior of populations [18-20]. The public's preference for online health information closely matches the field of infodemiology, a term that is a portmanteau of information and epidemiology. According to Eysenbach [18], the term is defined as the science of distribution and determinants of information in an electronic medium—specifically the internet—or in a population, with the aim to inform public health and public policy. Considering the global spread of COVID-19, using the internet to learn or gain information about the pandemic in this digital age is not surprising, as internet use has become pervasive worldwide [21,22]. Several studies have examined social media's influence on what people learned and the appropriate behaviors toward misinformation and conspiracy theories [14,15]. Similarly, Sulyok et al [12] examined the impact of web searches on the confirmed cases of COVID-19 in Europe, while Neely et al [23] investigated information-seeking behaviors on social media regarding the pandemic.

Miller [14,15] identified political leaders' failure in sensitizing the public as a motivating factor that pushed people to the internet as an alternative information source to learn about the COVID-19 pandemic. Misinformation had started flooding the web right from the initial stage of the emergence of the novel coronavirus, mainly from user-created content on social media [24]. Thus, as people turned to the web to search for information, there was limited nontechnical information for the nonexpert public about the coronavirus. Rather, people either got exposed to learning incorrect information about SARS-CoV-2 and COVID-19 or embraced fake news, misinformation, and conspiracy theories, with grave consequences [16]. Some of the unfounded misinformation included misconstruing COVID-19 as a "common cold" or as a hoax, which made people have a false sense of immunity, while others ignored any public health safety measures [11]. Similarly, the conspiracy theories propagated online included that COVID-19 was a bioweapon, that China intentionally released the virus to reduce the world population, and that 5G technology contributed to the fast spread of the pandemic. These beliefs initially led to the hoarding of essential goods as well as racial attacks against Chinese and other Asian people [14,15,25,26].

Other studies examined the role of social media and internet news outlets in generating misinformation, disinformation, fake news, and conspiracy theories about COVID-19 [25,27-29]. These studies tend to leave out the aspect of web searching, such as the use of Google searches, which constitutes a major channel through which the public obtain health-related information [30,31].

Research Objectives

This paper undertakes the first empirical investigation using a web search to learn about SARS-CoV-2 and COVID-19 and people's attitudes toward public health guidelines as expressed in the following research objectives:

1. Determine what people learned about the COVID-19 pandemic through web searching.
2. Examine any association between what people learned about COVID-19 and behavior toward public health guidelines.
3. Analyze the impact of misinformation and conspiracy theories about the COVID-19 pandemic on people's behavior toward public health measures.

These objectives are developed into research hypotheses in the sections that follow.

Theoretical Background

The Connectivism Learning Theory

This section examines the connectivism learning theory, which explains the use of digital platforms to enable learning [32,33]. This study employs this approach to explore how people learned about SARS-CoV-2 and COVID-19 through web searching and the potential behavioral implications toward public health guidelines, which scientists and medical experts recommend as ways to check the spread of COVID-19. For example, the study investigates if learning through web searching helped people acquire accurate knowledge or misinformation and conspiracy theories about the COVID-19 pandemic and its implications. Also, recent studies show that many people have yet to understand the science and the concept of SARS-CoV-2 and the disease, COVID-19, which increases the danger of embracing misinformation [14,28]. Several web platforms, including social media, online news, and other internet channels, contribute significantly to misinformation and conspiracy theories [14,34].

According to Dunaway [33], the connectivism theory developed by George Siemens analyzes the use of digital devices, computer networks, and electronic platforms to learn. The view is considered a pedagogical strategy for the digital age, emphasizing knowledge sharing across an interconnected web and internet network [32,33]. The approach focuses on knowledge acquisition using information technology platforms and learning from multiple sources, developing skills, and disseminating information [32]. The platforms incorporate information on social media, internet websites or blogs, and search engines that users can employ to learn and exchange knowledge, skills, and expertise [28,33].

Implications of the Connectivism Learning Theory

One of the implications of the connectivism learning theory is that learning can occur outside the traditional classroom by using networked systems that enhance connections, interactions, and collaborations among learners [29]. However, some learning theory experts criticize the connectivism theory for not offering any improvement to the actual learning method other than using Web 2.0 and related platforms [32,33]. Hence, it cannot be deemed a substantive learning theory. Instead, it provides a bridge to other pedagogical methods: behaviorism, cognitivism, and constructivism. The core of the Siemens and Downes connectivism idea aims to move away from the traditional classroom learning techniques to a new theory of learning that embraces technology as the learning tool, which can inspire the new generation of learners and educators [32,33]. Thus, the theory draws its strength from web-based activities [29].

The key benefit of the method is its intuitiveness and its ability to captivate learners due to the ubiquitous use of the internet in today's world. The following principles contribute to the popularity of connectivism as a learning theory [33]:

1. Learning and knowledge rest in diversity of opinions, as experienced today.
2. Learning is a process of connecting specialized nodes or information sources.
3. Learning may reside in nonhuman appliances.
4. The capacity to know more is more critical than what is currently known.
5. Nurturing and maintaining connections help to facilitate continual learning.
6. Ability to see connections between fields, ideas, and concepts is a core skill.
7. Currency (accurate, up-to-date knowledge) is the intent of all connectivism learning activities.
8. Decision making is itself a learning process. Choosing what to learn and the meaning of incoming information is seen through a shifting reality.

The connectivism learning theory, as explained above, closely mirrors the use of Google Trends and other internet platforms to learn about the outbreak of SARS-CoV-2 and COVID-19, especially where the masses did not get adequate, timely information about the coronavirus from the public agencies [35,36].

The connectivism learning theory is well suited to personal study and self-regulated learning [37,38]; in this case, how individual members of the public learned about SARS-CoV-2 and COVID-19 using web searches in the first 6 months of the COVID-19 pandemic.

An Overview of the SARS-CoV-2 Outbreak and the COVID-19 Pandemic

The Global Impacts of the SARS-CoV-2 Outbreak

The COVID-19 pandemic has caused severe problems ranging from health crises to psychological, social, business, and economic consequences all over the world [16,18,39]. Meanwhile, there is currently no specific cure for COVID-19. However, there has been significant progress in technological advances leading to substantial breakthroughs in vaccine discovery and development through the pioneering efforts of Pfizer, Moderna, and other companies from the United Kingdom, India, China, and other countries [40]. Administering the COVID-19 vaccines is ongoing worldwide, while several other vaccine discoveries and developments are in progress [41]. In the meantime, ongoing prevention, monitoring, and public health awareness are essential to mitigate the public health and economic burdens. The most important prevention strategy is to understand the disease and how it spreads.

Transmission

Coronavirus transmission primarily occurs through respiratory droplets released from infected persons during coughing, sneezing, or speech. One can also become infected with the virus via contact with contaminated surfaces. The virus can remain infectious in the air for 3 hours and on inanimate surfaces for up to 9 days or longer. This has implications for nosocomial spread and superspreading events [42]. The virus has also been isolated from blood, urine, and stool specimens. It is important to note that asymptomatic infected people may not be aware that they are infected because they do not have the symptoms or may not recognize the symptoms. Infected individuals can be contagious for up to 4 weeks and can unknowingly be spreading the infection during this time [42].

Clinical Presentation and Diagnosis

Symptoms usually appear 2 to 14 days after exposure. Most confirmed cases of SARS-CoV-2 infection are asymptomatic, and they recover without treatment. Common symptoms include fever, cough, shortness of breath, chills, myalgia, headache, sore throat, anosmia, and dysgeusia. Severe cases present with dyspnea, tachypnea, hypoxia (blood oxygen saturation ≤93%), a ratio of the arterial partial pressure of oxygen to the fraction of inspired oxygen of less than 300 mm Hg, and lung infiltration [43]. Some patients present with gastrointestinal symptoms, such as vomiting, diarrhea, and abdominal pain, as well as cardiovascular features, such as arrhythmia, shock, and acute cardiac injury [44]. There have been reports of asymptomatic carriers presenting with symptoms such as loss of smell and taste. In children, the majority present with mild (ie, fever, cough, fatigue, and congestion) or moderate (ie, pneumonia) symptoms [44]. Some may be asymptomatic. Children under 5 years old may present with respiratory organ failure.

Chest computed tomography scans shows a distinct appearance of ground-glass lung opacity, often bilateral, in patients who develop pneumonia [43]. Other radiographic features, such as "crazy-paving sign, multifocal organizing pneumonia, and architectural distortion in a peripheral distribution," may appear with disease progression. Diagnostic testing is performed from respiratory (ie, nose, throat, and saliva) and serum samples, using a real-time reverse transcription–polymerase chain reaction panel or antibody test. Viral RNA has also been detected in stool and blood [5].

Complications

Some hospitalized patients develop thromboembolism, especially deep venous thrombosis, and pulmonary embolism. Other complications include microvascular thrombosis of the toes, clotting of catheters, myocardial injury with ST-segment elevation, and large vessel strokes. This complication may be associated with the release of high levels of inflammatory cytokines and activation of the coagulation pathway caused by hypoxia and systemic inflammation secondary to COVID-19 [45].

Prevention and Control

People must be well-informed. Infected persons must practice respiratory etiquette to avoid infecting others, including covering coughs and sneezes with a tissue and discarding it properly, coughing into the inside of the elbow, and covering the nose and mouth properly with a surgical face mask. Best practices include proper handwashing with soap and water for at least 20 seconds or at least 60% alcohol–based hand rub. Touched surfaces must be cleaned frequently with disinfectants. People must avoid touching the eyes, nose, and mouth with unwashed hands and they must avoid close contact with people who are ill [46]. The US Centers for Disease Control and Prevention (CDC) recommends that infected and exposed individuals must isolate or quarantine themselves, respectively, for at least 14 days. The CDC also recommends social distancing, including avoiding mass gatherings or large community events, shaking hands, or giving "high fives" [41]. In health care settings, standard contact and airborne precautions, as well as eye protection, should be used to mitigate the spread of SARS-CoV-2 [46]. There is no specific cure for COVID-19. Management is mainly supportive care and treatment of secondary infections. Severely ill patients may need advanced organ support.

Methods

Google Trends and Search Keywords About SARS-CoV-2 and COVID-19

This study used an infodemiology approach to evaluate the use of web searching to learn about SARS-CoV-2 and COVID-19. As an area of scientific research, infodemiology is a method or technique designed to measure and track health information "demand" automatically (eg, by analyzing search queries) as well as "supply" on the internet [18,20,47]. The goal is to inform public health policy and practice. This study uses data from Google Trends, a freely available online resource that provides information on what was and is trending based on actual users' Google queries [11,48,49].

Google Trends offers various search options, such as "Trending Searches" (ie, trending queries for daily search trends and real-time searches in a selected region) or "Year in Search" (ie, what was trending in a specific area in a particular year). Another option is to "Explore," which allows an investigation of an area of interest based on keywords over the selected periods and regions. This study uses the "Explore" option, which allowed data to be retrieved directly from the Google Trends "Explore" page in comma-separated values format. It is also important to note that Google Trends data points are normalized to have a maximum value of 100 and a minimum value of 0. We normalized the data set by dividing each data point by total searches of the geography and the time range it represents to compare relative popularity. Note that the value 0 does not necessarily indicate no searches but represents a significantly low search volume that does not warrant inclusion in the results [47].

In this study, we captured the worldwide Google Trends data covering the initial months of the SARS-CoV-2 outbreak from January 1 to June 30, 2020 (ie, 182 daily data points for each search term). Regarding the search terms, this paper employed 25 keywords and phrases used by the public to learn about the COVID-19 pandemic through web searches. We identified the search keywords (Table 1) through a literature survey of

published documents indexed on the Web of Science. Six search terms were related directly to the ongoing pandemic: "nCoV," "2019-nCoV," "SARS-CoV-2," "COVID-19," "pandemic," and "coronavirus." Another six keywords addressed previous viral or coronavirus epidemics: "SARS-CoV," "SARS," "MERS-CoV," "MERS," "virus," and "influenza." The third category of search terms represented public health safety measures that experts recommended as guidelines to limit the spread of COVID-19: "social distancing," "wear a facial mask," and "wash hands." The final category of keywords represented misinformation and conspiracy theories, such as "China virus," "common cold," and "bioweapon" (Table 1 [2,3,6,7,9,11,14,15,25-27,42,46,50-60]).

Table 1. Web search terms used to learn about SARS-CoV-2 and the COVID-19 pandemic, misinformation and conspiracy theories, and public health safety guidelines, based on a literature review.

Category	Search terms
COVID-19 and related epidemics	"2019-nCoV" [3], "nCoV" [3,9], "SARS-CoV-2" [42], "COVID-19" [9,25,50], "pandemic" [6,7], "MERS-CoV" [51,52,60], "MERS[a]" [51,53], "SARS-CoV" [42,52], "SARS[b]" [52,53], "virus" [54], "coronavirus" [2,53,55], "influenza" [46], and "flu" [59]
Misinformation and conspiracy theories	"virus hoax" [14,15], "injecting or ingesting bleach" [56], "5G" or "@5G" technology enhancing the spread of the virus [25,27], "COVID-19 hoax" [14,15,26], "common cold" ("commoncold2020") [11], "China virus" [50], and "bioweapons" created by China [14,25,60]
Public health measures	"social distancing" [57], "wash hands" or "hands wash" [57], "wear a facial mask" [58], "isolation" [57], and "quarantine" [57]

[a]MERS: Middle East respiratory syndrome.

[b]SARS: severe acute respiratory syndrome.

Research Hypotheses

Hypothesis 1

For the purpose of determining what people learned about the COVID-19 pandemic through web searching, we defined the null and alternative hypotheses as follows:

- Hypothesis 1_0: People did not learn about COVID-19 through web searching using the identified keywords.
- Hypothesis 1_1: People learned about COVID-19 through web searching using the identified keywords.

Hypothesis 2

Based on the literature, using a web search to learn about a subject of interest can influence the learner's decision making and actions [61]. On this premise, this study examined any association between what people learned about COVID-19 and people's behavior toward the public health guidelines. We developed two separate hypotheses (Hypotheses 2 and 3). The first aspect related to web searching to learn about COVID-19 (ie, concept, science, and structure of SARS-CoV-2 and COVID-19), while the second aspect evaluated learning about misinformation and conspiracy theories as well as the behavioral response to the public health measures. The null and alternative hypotheses regarding the first aspect are as follows:

- Hypothesis 2_0: There is no association between what people learned about COVID-19 through web searching and behavior toward public health measures.
- Hypothesis 2_1: There is an association between what people learned about COVID-19 through web searching and behavior toward public health measures.

Hypothesis 3

There is a widely held assertion that misinformation and conspiracy theories about the COVID-19 pandemic have had a significant impact on people's behavior toward public health measures. We defined the null and alternative hypotheses for learning about misinformation as follows:

- Hypothesis $3A_0$: There is no association between misinformation learned about COVID-19 and people's behavior toward public health measures.
- Hypothesis $3A_1$: There is an association between misinformation learned about COVID-19 and people's behavior toward public health measures.

Similarly, we defined the null and alternative hypotheses for learning about conspiracy theories as follows:

- Hypothesis $3B_0$: There is no association between conspiracy theories learned about COVID-19 and people's behavior toward public health measures.
- Hypothesis $3B_0$: There is an association between conspiracy theories learned about COVID-19 and people's behavior toward public health measures.

Data Analysis

Data analysis employed statistical trends and graphical visualization, correlation and regression, principal component analysis (PCA), and predictive models [12,56,62-64]. The statistical trends and analyses involved evaluating relationships among the listed variables using the statistical trends, including graphical display, correlation, and PCA, which helped determine the predictiveness of the learning attributes and learners' actions toward public health guidelines. We used the JMP 15 package from SAS software (SAS Institute Inc) [65] for statistical analysis and Microsoft Excel 2019 to create the charts and graphs. We also used SPSS software (version 27; IBM Corp) to compute the correlation matrix and the PCA as well as statistical packages in R (version 4.0.5; The R Foundation) for the linear modeling. The evaluation helped establish the correlation between the study attributes.

Results

Overview

The data analyzed in this study came from the Google Trends worldwide index covering the period from the initial outbreak of COVID-19 on January 1, 2020, up to June 30, 2020, when the pandemic became widely known [2]. The outbreak had been reported to the WHO's office in China on December 31, 2019 [6]. The reason for focusing on the first 6 months of the pandemic was to capture what people learned during the early days of the outbreak as well as the possible impacts of what people learned through web searching on individuals' attitudes toward public health safety measures.

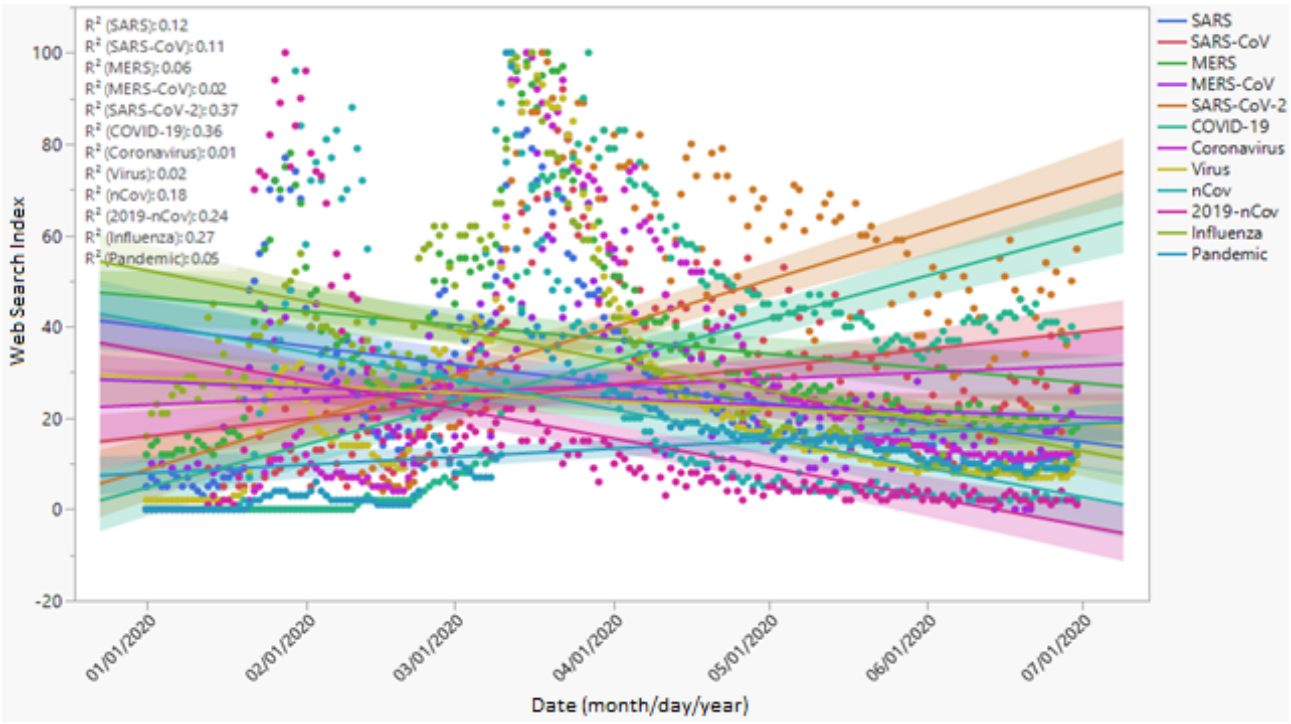
To better understand the characteristics of Google Trends data, we have presented the summary statistics of the daily search index for each of the 25 keywords or search terms and phrases (Table S1 in Multimedia Appendix 1). The average normalized scores for the terms varied from 2.65 ("ingesting bleach") to 39.75 ("SARS-CoV-2"), as shown in Table S1 in Multimedia Appendix 1.

Temporal Trends: Using Web Searches to Learn About SARS-CoV-2

The keywords employed to conduct web searches indicate what people learned about the COVID-19 pandemic [66]. As presented in Table 1, some of the search keywords addressed the novel coronavirus directly, while others examined misinformation and conspiracy theories.

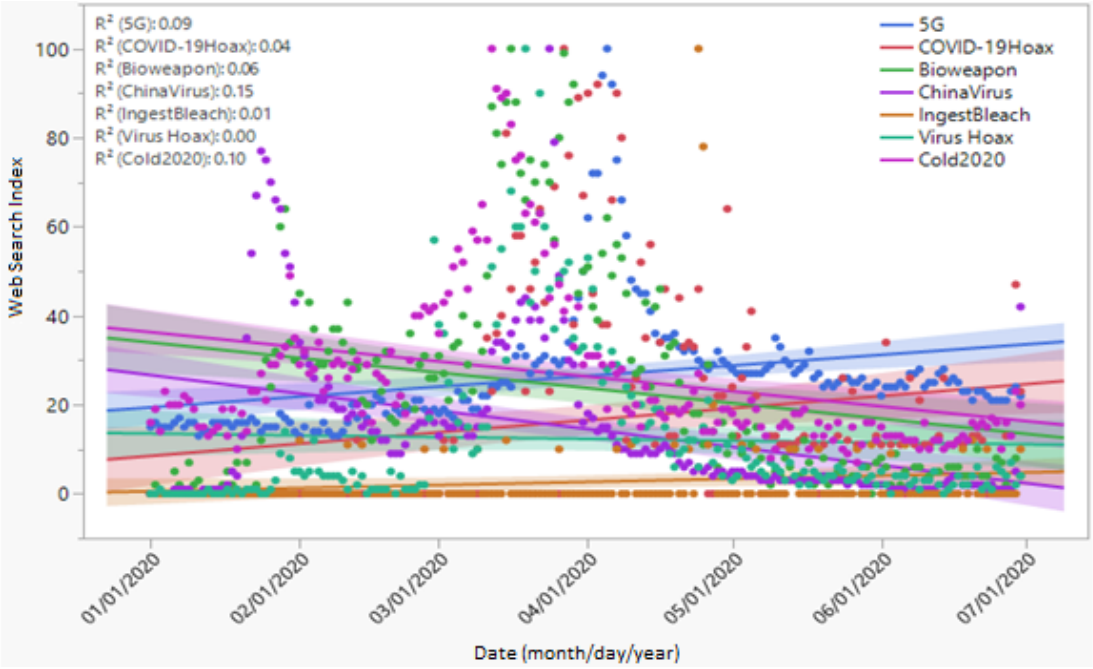
Figure 1 presents the first category of web search terms that people used to learn about the COVID-19 pandemic. In the early period, most people used keywords and phrases that explain previous coronavirus epidemics, including "influenza," "MERS," "MERS-CoV," "SARS," and "virus." Although scientists ruled out the past epidemics, the WHO officials highlighted those terms as examples of past coronavirus outbreaks during press briefings [6]. The use of those keywords in the web searches nosedived after the WHO formally named the novel coronavirus and the disease (ie, "nCoV," "2019-nCoV," "SARS-CoV-2," and "COVID-19"). The coefficients of determination (R^2) of the keywords are 0.37 for "SARS-CoV-2," 0.36 for "COVID-19," 0.27 for "influenza," 0.24 for "2019-nCoV," 0.18 for "nCoV," 0.12 for "SARS," and 0.11 for "SARS-CoV," indicating the proportion of the variation in the search index over the period for the listed keywords. Similarly, the search terms "MERS," "MERS-CoV," "pandemic," "virus," "coronavirus," "nCoV," and "2019-nCoV" had R^2 values of less than 0.1.

Figure 1. The keywords used by people to learn about SARS-CoV-2 and COVID-19 through web searches. MERS: Middle East respiratory syndrome; SARS: severe acute respiratory syndrome.



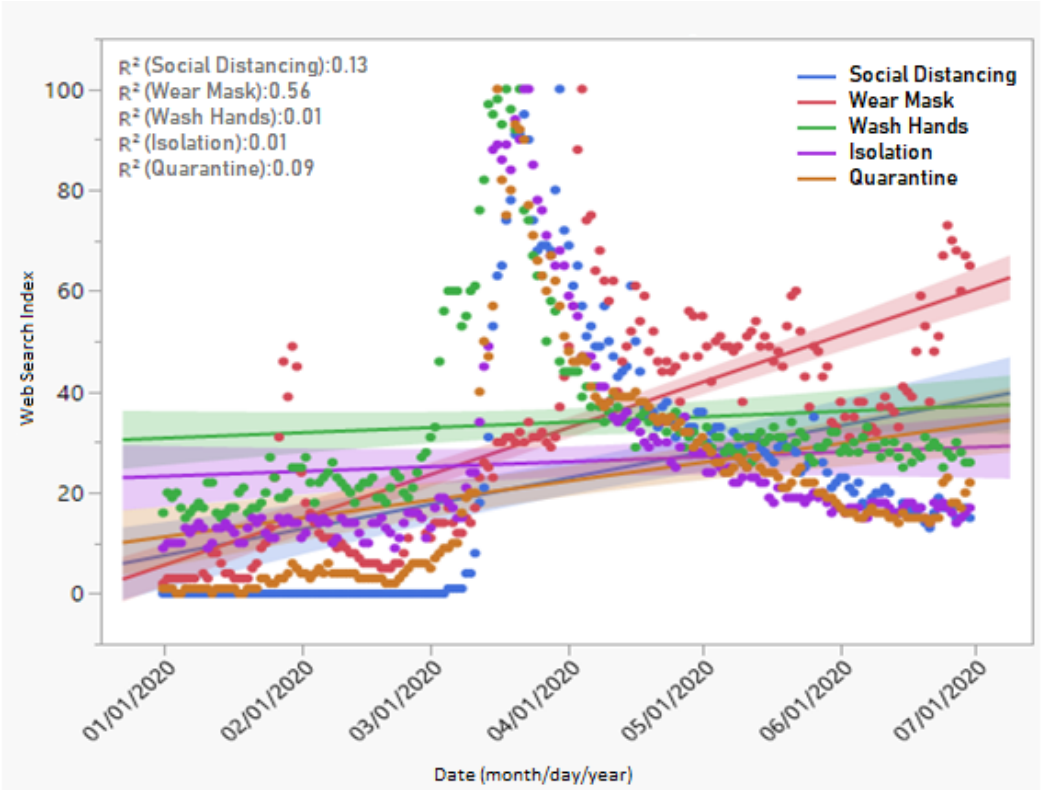
The second category of keywords involved misinformation and conspiracy theories (Figure 2). The variation in the use of the terms was measured using coefficients of determination (R^2), which were 0.15 for "China virus," 0.10 for "common cold," and 0.09 for "5G" or "@5G." Most searches in the initial months of the outbreak used the keywords "common cold" ("cold2020"), "biological weapon," and "China virus," thus encouraging the misconception about SARS-CoV-2 as a "common cold," a "biological weapon," or a "China virus" [50]. Some studies explain that the purpose of releasing the coronavirus was to reduce the world population [14,25]. However, web searches using these terms fell continuously over time to a near-zero search index, while new words ("5G" and "COVID-19 hoax") surfaced and increased significantly (Figure 2).

Figure 2. Worldwide search index showing the learning terms that represent misinformation and conspiracy theories about COVID-19.



The third segment of the trend analysis involved web searches to learn about public health measures (Figure 3). The results show that there was little or no interest in learning about wearing a facial mask ("wear mask") and maintaining social distancing ("social distancing") at the start of the pandemic. But the trends changed quite quickly, recording a dramatic increase from a search index of 0 at the beginning of the outbreak to achieving a maximum search index of 100 in March and April 2020, as the pandemic spread worldwide. The coefficients of determination were as follows: "wear mask" (R^2=0.56), "social distancing" (R^2=0.13), and "quarantine" (R^2=0.09). The increases, especially regarding "wear mask" and "social distancing," were sustained for a long time.

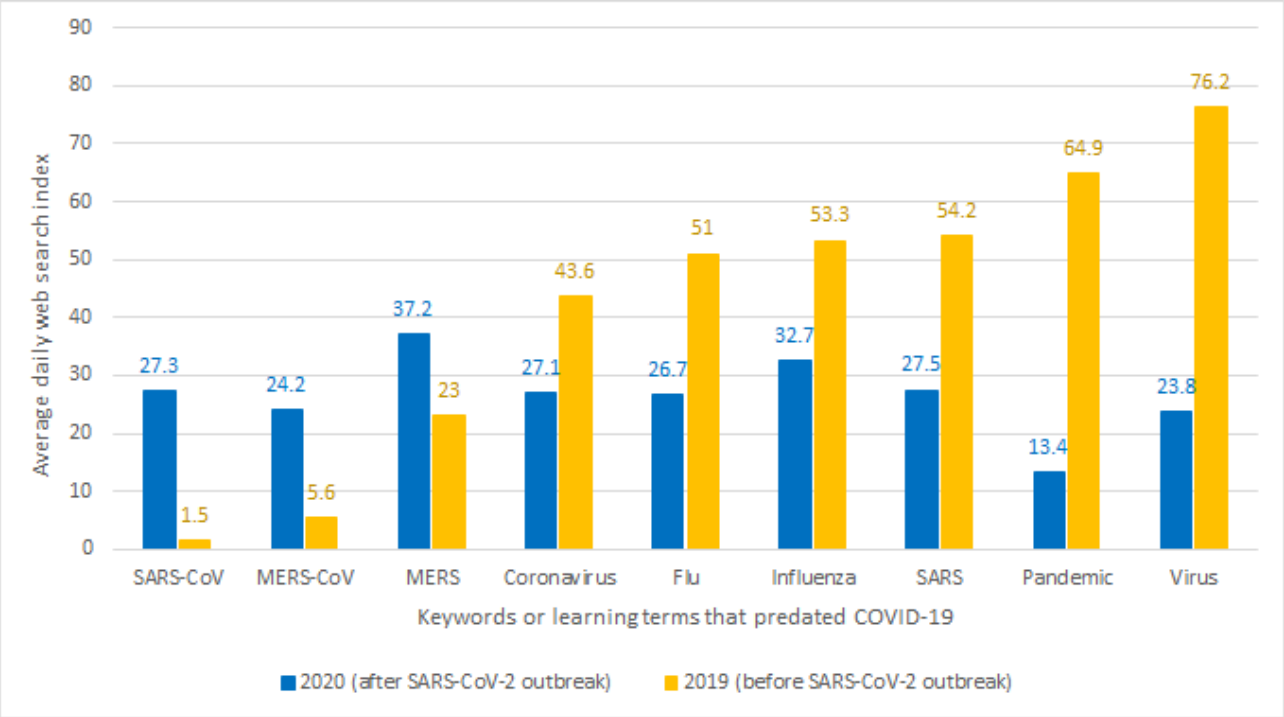
Figure 3. Trend analysis showing the web search index for learning about public health safety measures.



What Did People Learn About COVID-19 Through Web Searches?

The keywords identified above approximate what people learned about SARS-CoV-2 and COVID-19 through web searches. However, some of the search keywords existed before the ongoing pandemic, while some terms referred to previous coronavirus epidemics (eg, “SARS,” “SARS-CoV,” “MERS,” “MERS-CoV,” “influenza,” “virus,” “pandemic,” and “coronavirus”). It is plausible to argue that the search index for the pre-existing keywords listed represent purposes other than learning about COVID-19. Based on this assumption, we conducted a dependent two-sample *t* test to examine the difference in the mean search index of the pre-existing keywords in the previous years before the COVID-19 outbreak and during the ongoing pandemic. Figure 4 compares the mean search index before and after the outbreak for each keyword.

Figure 4. The average search index for pre-existing keywords before and during the COVID-19 pandemic. MERS: Middle East respiratory syndrome; SARS: severe acute respiratory syndrome.

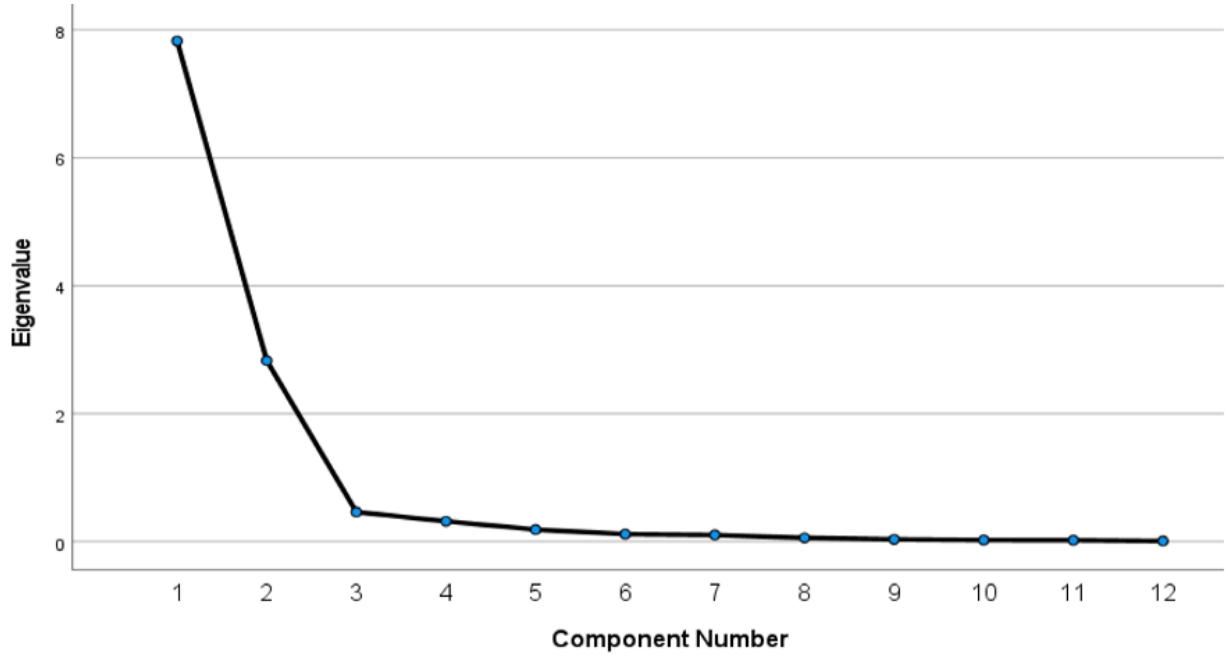


The null hypothesis was that the search indexes of each learning term—“SARS,” “SARS-CoV,” “MERS,” “MERS-CoV,” “influenza,” “virus,” “pandemic,” and “coronavirus”—before and after the outbreak of COVID-19 would be equal, and the alternative hypothesis was that they would be unequal. The results, as seen in Figure 4, show that the differences in the mean search indexes before and during the ongoing pandemic were more than 60% in all cases. Also, the P values were close to zero for all the variables. We rejected the null hypothesis and concluded that the significant differences in the mean search indexes of the variables were due to the ongoing COVID-19 pandemic. Unexpectedly, the mean search index for some pre-existing keywords (eg, “flu,” “influenza,” “SARS,” “pandemic,” and “virus”) declined during the pandemic.

What Search Terms Contributed to Learning About SARS-CoV-2 and COVID-19?

PCA was employed to evaluate the underlying latent variable of the search terms that contributed to learning about COVID-19. Based on the scree plot and the elbow rule, we can limit the factors extracted to the first two principal components (Figure 5): the keywords that address previous epidemics (pastCoVepidemics) and the keywords that explain the ongoing pandemic (presCoVpandemic). A scree plot is a graphical representation of the percentage variability explained by each principal component.

Figure 5. Scree plot of COVID-19 learning terms; two principal components extracted 88.78% of the total variation.



The first two underlying components explained a total of 88.78% of the variation in learning terms, with the first (pastCoVepidemics) and second (presCoVpandemic) components determining about 65.2% and 23.58%, respectively, of the information about COVID-19 from the 12 search keywords (Table 2).

Table 2. Total explained variance for search keywords used to learn about COVID-19.

Measure	PC[a] 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9	PC 10	PC 11	PC 12
Total initial eigenvalue	7.82	2.83	0.46	0.32	0.19	0.12	0.11	0.06	0.04	0.03	0.02	0.01
Variance, %	65.20	23.58	3.86	2.65	1.57	0.98	0.88	0.49	0.29	0.21	0.19	0.09
Cumulative variance, %	65.20	88.78	92.64	95.30	96.87	97.84	98.72	99.21	99.51	99.72	99.91	100

[a]PC: principal component.

Linear combinations of the two components (pastCoVepidemics and presCoVpandemic) are as follows:

- PastCoVepidemics = 0.98 “virus” + 0.934 “coronavirus” + 0.929 “MERS” + 0.923 “flu” + 0.858 “MERS-CoV” + 0.858 “SARS” + 0.791 “SARS-CoV” + 0.799 “pandemic” + 0.814 “influenza”
- PresCoVpandemic = –765 “nCoV” + 0.784 “COVID-19” + 0.766 “SARS-CoV-2.”

Table 3 shows the weights (loadings) of the terms for the two components. Note that we record loadings greater than 0.6 to combine only search keywords that have a high correlation with the component in the linear combinations.

Table 3. Component matrix and weight loadings for search keywords used to learn about COVID-19.

Component[a]	Weight loading for each keyword											
	Virus 2020	Coronavirus 2020	MERS[b] 2020	Flu 2020	MERS-CoV 2020	SARS[c] 2020	Influenza 2020	Pandemic 2020	SARS-CoV 2020	COVID-19	SARS-CoV-2	2019-nCoV
1	0.98	0.934	0.929	0.923	0.858	0.858	0.814	0.799	0.791	0.562	0.601	0.45
2	–0.059	0.278	–0.305	–0.278	–0.121	–0.443	–0.477	0.315	0.505	0.784	0.766	–0.765

[a]Component 1 is pastCoVepidemics (keywords that address previous epidemics) and component 2 is presCoVpandemic (keywords that explain the ongoing pandemic).

[b]MERS: Middle East respiratory syndrome.

[c]SARS: severe acute respiratory syndrome.

What Terms Fueled Misinformation and Conspiracy Theories About COVID-19?

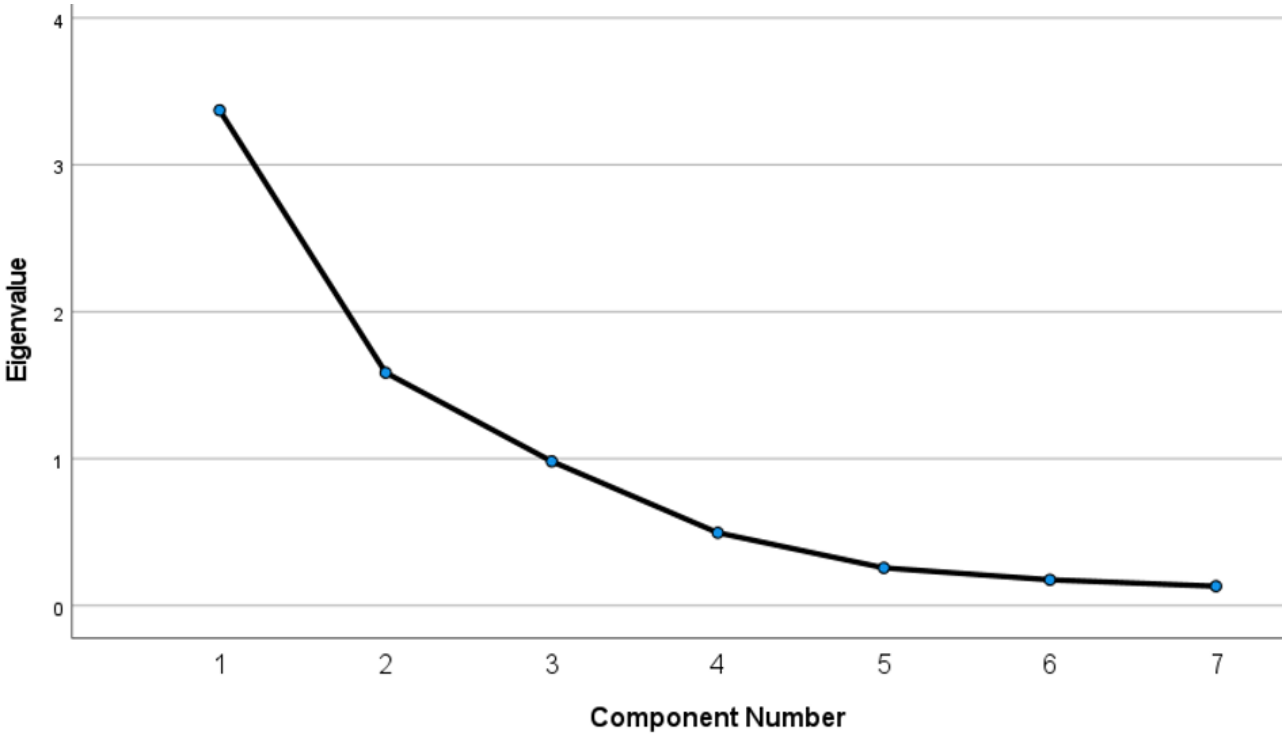
We identified eight search keywords from the literature that denote misinformation and conspiracy theories (Table 1). We also performed a PCA to evaluate the search terms that fueled misinformation and conspiracy theories. The results, as seen in Table 4 and Figure 6, identified three principal components and their variabilities based on the elbow rule: principal component 1 (48.17%), principal component 2 (22.65%), and principal component 3 (14.03%).

Table 4. Total variance explained involving terms that fueled misinformation and conspiracy theories.

Measure	PC[a] 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7
Total initial eigenvalue	3.372	1.586	0.982	0.496	0.257	0.176	0.132
Variance, %	48.171	22.652	14.026	7.085	3.669	2.509	1.888
Cumulative variance, %	48.171	70.823	84.849	91.934	95.603	98.112	100

[a]PC: principal component.

Figure 6. Scree plot of misinformation and conspiracy theory terms.



The three components explained 84.85% of the variation in the search keywords under misinformation and conspiracy theories. The first component represented misinformation. We quantified the daily number of misinformation terms searched using the linear combination as follows:

- Misinformation = 0.789 “common cold” + 0.928 “bioweapon” + 0.908 “virus hoax” + 0.875 “cold 2020” + 0.692 “COVID-19 hoax” + 0.60 “China virus.”

The second and third components addressed two conspiracy theories (ConspTheory1 and ConspTheory2), which speculated that 5G technology contributes to the spreading of COVID-19, that COVID-19 is a “China virus” that was intentionally created and released, and that ingesting or injecting bleach can cure COVID-19 infection or kill the virus. The results present these variables in the separate components (Table 5) as follows:

- ConspTheory1 = 0.786 “@5G”
- ConspTheory2 = 0.97 “ingest bleach.”

Table 5. Component matrix and weight loading for terms that fueled misinformation and conspiracy theories.

Component[a]	Weight loading for each keyword						
	Bioweapon 2020	Virus hoax	Common cold 2020	COVID-19 hoax	China virus	@5G	Ingest bleach
1	0.928	0.908	0.789	0.692	0.601	0.471	–0.045
2	–0.152	0.033	–0.438	0.624	–0.558	0.786	0.226
3	0.03	0.005	0.093	–0.044	0.059	–0.145	0.973

[a]Component 1 is misinformation and is not included, component 2 is the conspiracy theory that 5G technology contributes to the spreading of COVID-19, and component 3 is the conspiracy theory that ingesting or injecting bleach can cure COVID-19 infection or kill the virus.

Public Health Safety Measures

This section investigates the impacts of what people learned through web searches on behaviors toward public health safety measures against COVID-19. Based on the elbow rule, the PCA identified two public health measure components, which we labeled as PubHealthMes1 and PubHealthMes2. The two components accounted for the variability in the search index of keywords used to learn about the public health measures against the spread of COVID-19 (Figure 7). The first and second components explained more than 75.4% and 18.7% of the variability, respectively; that is, the first two components explained about 94% of the variability (Table 6).

Figure 7. Scree plot of public health measures against COVID-19; two components extracted 93.99% of the variability in the search index.

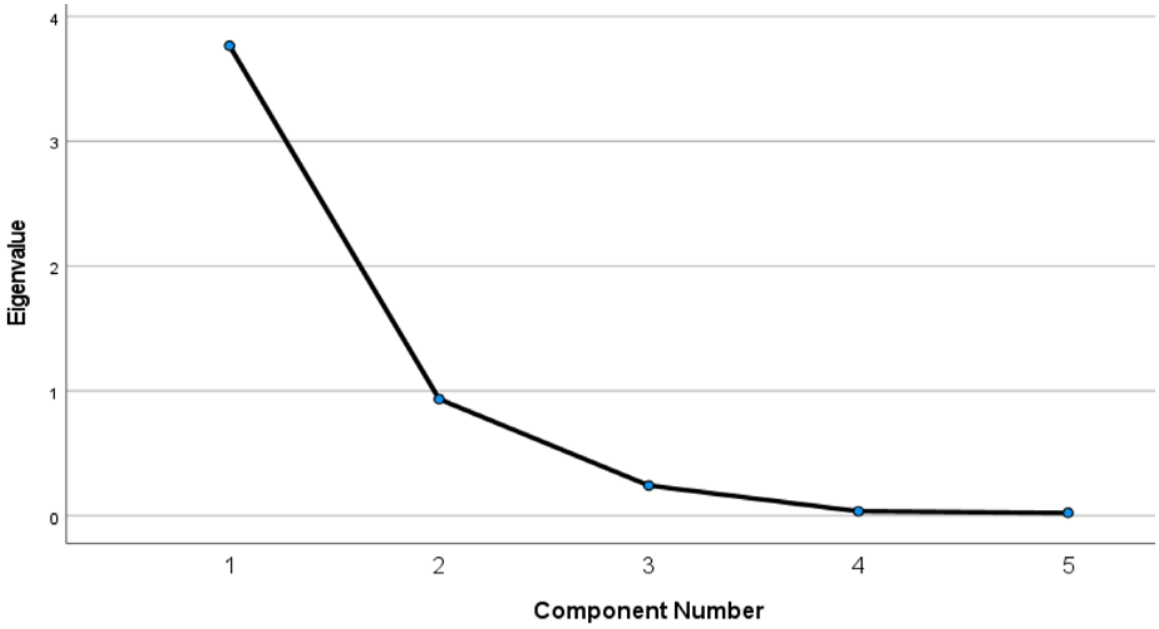


Table 6. Total explained variance for terms that explained the public health measures against COVID-19.

Measure	PC[a] 1	PC 2	PC 3	PC 4	PC 5
Total initial eigenvalue	3.768	0.933	0.24	0.035	0.023
Variance, %	75.359	18.665	4.799	0.71	0.466
Cumulative variance, %	75.359	94.025	98.824	99.534	100

[a]PC: principal component.

The first component, PubHealthMes1, includes the keywords "social distancing," "wash hands," "isolation," and "quarantine." In the second component, PubHealthMes2, the keyword "wear mask" explained 84.7% of the variability (Table 7):

- PubHealthMes1 = 0.953 "social distancing" + 0.847 "wash hands" + 0.953 "isolation" + 0.99 "quarantine"
- PubHealthMes2 = 0.847 "wear mask."

Table 7. Component matrix and weight loading for terms that explained the public health measures.

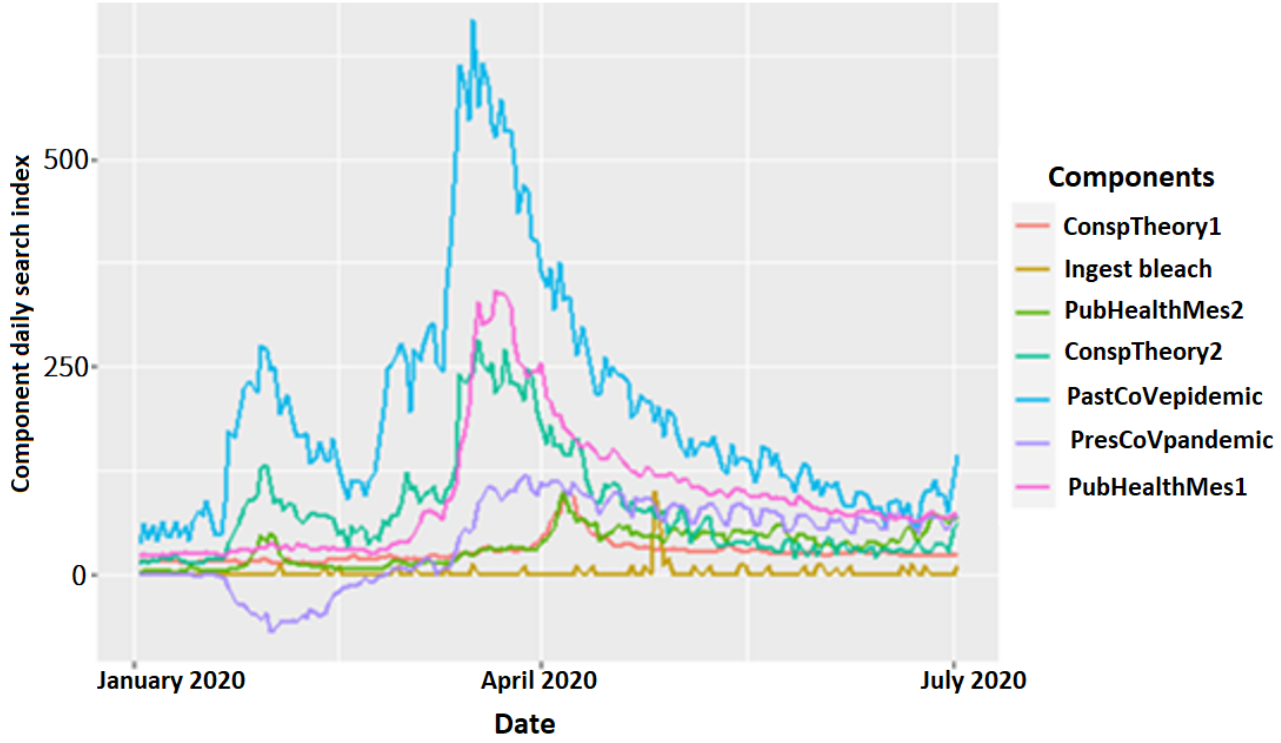
Component[a]	Weight loading for each keyword				
	Quarantine	Social distancing	Isolation	Wash hands	Wear mask
1	0.99	0.953	0.953	0.847	0.503
2	–0.028	0.143	–0.223	–0.38	0.847

[a]Component 1 is public health measures represented by the keywords "social distancing," "wash hands," "isolation," and "quarantine," and component 2 is the public health measure represented the keyword "wear mask."

Analysis of the Relationships Among the Principal Components

This section presents further analysis that tested the hypothesis raised in the earlier section using built predictive models. The variables identified the linear combination of search keywords significantly correlated (loading >0.6) to the principal components discussed in the Results subsections above. Also, Tables 2 to 7 present the underlying latent variables of the 25 search terms used to learn about COVID-19, the misinformation and conspiracy theories, and the public health measures. As stated, Figure 8 shows the underlying variables.

Figure 8. Daily search index of the principal components. ConspTheory1: conspiracy theory 1; ConspTheory2: conspiracy theory 2; pastCoVepidemics: keywords that address previous epidemics; presCoVpandemic: keywords that explain the ongoing pandemic; PubHealthMes1: public health measures 1; PubHealthMes2: public health measures 2.



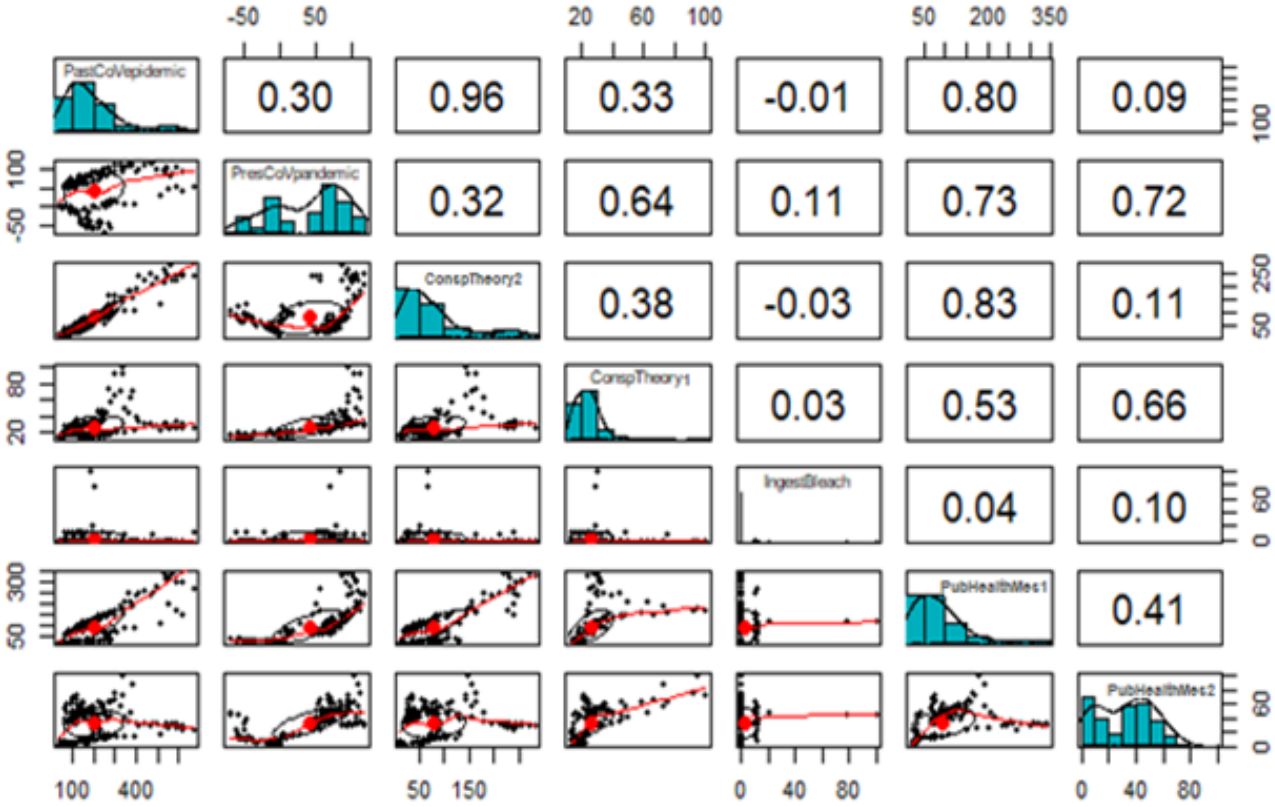
Here, we examine how the underlying variables and the search terms impacted learning and behavior toward the public health measures: learning about COVID-19 (pastCoVepidemics and presCoVpandemic), misinformation and unproven or misleading assertions (misinformation), conspiracy theories (ConspTheory1 and ConspTheory2), and public health safety measures (PubHealthMes1 and PubHealthMes2).

The results show three essential highlights from the daily search index. First, the most popular search terms used at the initial outbreak of the pandemic in early January 2020 were terms representing misinformation and past epidemics. The search keywords that represented conspiracy theories were not used until May 2020. Also, the use of learning terms that directly explained COVID-19 (ie, presCoVpandemic) corresponded with the WHO's naming and renaming of the coronavirus and the disease (ie, "2019-nCoV," "SARS-CoV-2," and "COVID-19") in January, February, and March 2020.

Information Learned Versus Behavior Toward Public Health Measures

We employed correlation analysis among the variables, scatterplots, and their histograms to examine the relationship between what people learned and their attitudes toward public health measures (Figure 9).

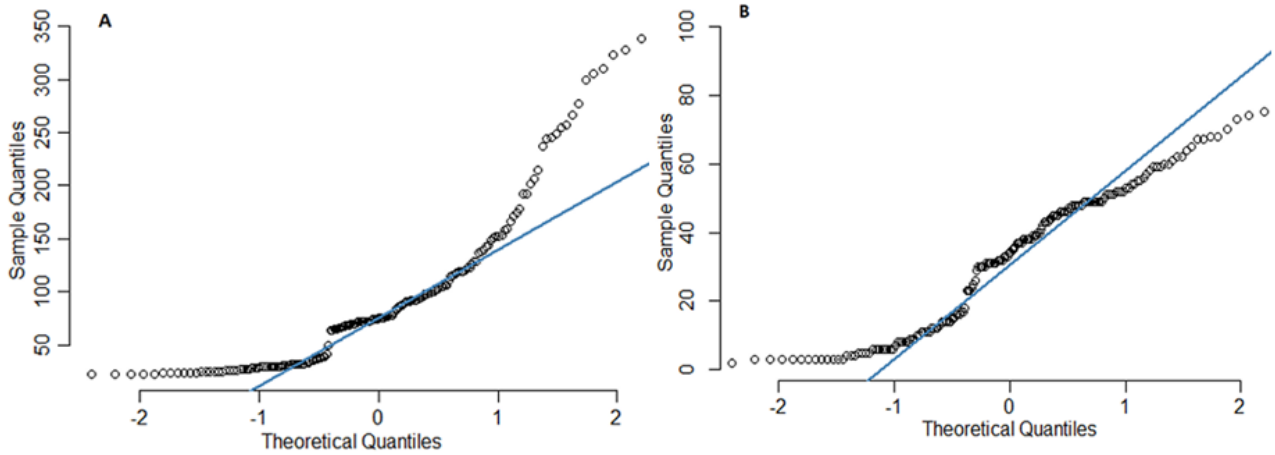
Figure 9. Panel pair plots of linear correlations between misinformation and conspiracy theory keywords. pastCoVepidemics: keywords that address previous epidemics; presCoVpandemic: keywords that explain the ongoing pandemic; PubHealthMes1: public health measure 1; PubHealthMes2: public health measure 2.



PubHealthMes1 has a robust positive relationship with pastCoVepidemics (r=0.80) and a moderate positive relationship with presCoVpandemic (r=0.73), which implies the effectiveness of learning keywords associated with past coronavirus epidemics (eg, "SARS," "SARS-CoV," "MERS," "MERS-CoV," and more) and the ongoing pandemic (eg, "2019-nCoV," "SARS-CoV-2," and "COVID-19"). Similarly, PubHealthMes2 (ie, wearing a facial mask) has a moderate positive relationship with presCoVpandemic (r=0.71). There is a strong association between actions taken and the information learned. Figure 9 shows a correlation matrix.

Given such a strong linear relationship between the search terms and people's behaviors and actions, a multiple linear regression model seems acceptable as a predictive model. But the data failed the assumption of normality, as shown by the quantile-quantile (Q-Q) plots in Figure 10, A and B. A normal Q-Q plot helps to compare two probability distributions, by plotting the residuals against theoretical quantiles. Plots A and B in Figure 10 show that most residuals are not lying on the diagonal line; hence, the data are not normally distributed.

Figure 10. Quantile-quantile (Q-Q) plots of (A) PubHealthMes1 (public health measures 1, represented by the keywords "social distancing," wash hands," "isolation," and "quarantine") and (B) PubHealthMes2 (public health measures 2, represented the keyword "wear mask").



From Figure 9, we can observe that the PubHealthMes1 histogram suggests a right-skewed distribution, while the PubHealthMes2 histogram suggests a bimodal distribution. To describe and explain the relationship between the public health measures and the terms that contributed to learning about COVID-19, misinformation, and conspiracy theories, we used the log-linear predictive model, since the data are based on a number of occurrences or frequency and not normally distributed. Also, the log-linear model does not need to satisfy any assumptions, which we represent as follows:

where x_i is the covariate and β_i is the parameter to be estimated.

The data obtained through the PCA have been fitted to a log-linear regression model using RStudio tools from R (version 4.0.5; The R Foundation). The regression model obtained becomes the following:

$$\text{Log(PubHealthMes1)} = 3.33 + 0.0024\ \text{pastCoVepidemics} + 0.011\ \text{presCoVpandemic}$$

The R^2 of the model is 0.93 and the P value is <.001. All the coefficients in this model are significant at an α level of 0. From the model, we see that there is a 1.10 percentage ($100 \times [\exp^{0.011} - 1]\%$) increase in the public health measure for every 1% increase in presCoVpandemic (ie, "2019-nCoV," "SARS-CoV-2," and "COVID-19"). This may be explained by links contained in recent COVID-19 pandemic articles on public health measures. For every 1% increase in past epidemic searches, there is a 0.25% increase in the searches of public health measures. Similarly, the regression model for PubHealthMes2 obtained becomes the following:

$$\text{Log(PubHealthMes2)} = 0.01 \times \text{presCoVpandemic} + 2.35$$

The R^2 of the new model is 0.50 and the P value is <.001. The two models show that both PubHealthMes1 and PubHealthMes2 are significantly predictive using the presCoVpandemic keywords. It shows that learning about the present pandemic creates an incentive for people to learn about public health measures, which we can approximate to a desired intention to comply with the measures.

Misinformation, Conspiracy Theories, and Public Health Measures

This section examines the relationship between conspiracy theories, misinformation, and public health measures using correlation and predictive analyses. As discussed earlier, we recategorized the search terms representing misinformation and conspiracy theories into two principal components. Hypothesis 3A, as defined in the Research Hypotheses section, focuses on misinformation and public health guidelines. At the same time, Hypothesis 3B addresses the impact of conspiracy theories on people's behavior toward the same safety measures.

Using correlation analysis to evaluate the association between conspiracy theories and public health measures, we observed a moderate positive linear association between ConspTheory2 and PubHealthMes2 (r=0.66) and a moderate linear relationship between ConspTheory1 and PubHealthMes1 (r=0.53). The log-linear analysis showed that conspiracy theories are not significant predictors of PubHealthMes1 (P=.62) but are for PubHealthMes2 (P=.008). Thus, the null hypothesis that there is no association between conspiracy theories (ie, ConspTheory1) and people's responses to public health (ie, wearing facial masks to limit the spread of COVID-19) is rejected. We can conclude that conspiracy theories are predictive of people's behaviors in wearing facial masks.

We also analyzed the relationship between misinformation and behavior toward public health measures. Although the correlation and the predictive analyses showed a moderate positive linear relationship between misinformation and PubHealthMes1 (r=0.83), it is not predictive. The relationship between the two variables is just mathematical but not causal. Despite a negligible negative linear relationship between misinformation and wearing a facial mask (r=0.11), the log-linear model shows that misinformation is not a significant predictor for both PubHealthMes1 (P=.27) and PubHealthMes2 (P=.13). Notwithstanding the strong linear relationship between web searches to learn about misinformation and public health measures, there is no sufficient evidence to reject the null hypothesis. We can conclude that there is no association between learning about COVID-19 misinformation and people's response to public health guidelines.

Discussion

Principal Findings

We found that people used search keywords related to past coronavirus epidemics (pastCoVepidemics) and the ongoing pandemic (presCoVpandemic) to learn about SARS-CoV-2 and COVID-19. However, the attention accorded to the pandemic led to less focus on terms relating to perennial illnesses (eg, "common cold," "flu," and more). These results corroborate studies reporting the unintended positive consequences of COVID-19 leading to declines in cases of influenza, flu, and similar infections (eg, Soo et al [46]). Other learning terms employed were keywords that addressed the pandemic directly. The average search indexes for those keywords were 19.01 for "nCoV" and "2019-nCoV," 39.75 for "SARS-CoV-2," and 32.40 for "COVID-19."

Studies examining learning by web searching emphasized the significance of the search terms or phrases on what the users intended to learn. A trending word on the web indicates what information people are interested in learning [11,61,66]. This study identified the 25 most-used keywords to learn about SARS-CoV-2 and COVID-19 through web searches.

Regarding the impacts of what people learned on their behavior toward public health measures, the PCA identified three latent variables, classified as misinformation, ConspTheory1, and ConspTheory2. Only ConspTheory1 ("@5G") directly and significantly influenced people's behavior toward public health measures (ie, PubHealthMes2 ["wear mask"]). The conspiracy that 5G technology enhances the easy spread of COVID-19 [14] highlights danger, which can cause people to take precautions. A different study [15] identified erroneous beliefs in the 5G

conspiracy theory as leading to the hoarding of essential goods during the initial period of the SAR-CoV-2 outbreak. Although there was a high correlation between misinformation (Tables 4 and 5) and behavior toward public health measures, this was not statistically significant based on the web search index. Also, as the pandemic lingers, thereby causing severe health and social crises, strains in family relations, and economic and business losses, many people are becoming increasingly aware of COVID-19 dangers [21,57,67-69]. Through direct impacts or by experience, this can cause changes in people's behavior irrespective of whether they believed the misinformation or not.

Strengths and Limitations

Internet platforms continue to play a significant role in health communication during the ongoing COVID-19 pandemic. Some studies attribute the increase in misinformation and conspiracy theories about COVID-19 in different countries to web searches, social media use, and online news media platforms that are used to learn about SARS-CoV-2 and COVID-19 [10,11,29]. However, most studies were anecdotal with no empirical evidence. Using Google Trends data, this study provides the first empirical evidence to this discourse. In the era of big data, the analysis of Google queries can be envisioned as a valuable tool for researchers to explore and predict human behavior, especially as studies suggest that online data can correlate with actual health data [70,71].

Infodemiology studies have their limitations too. While Google search keywords are short and easy to classify automatically, interpreting the terms semantically can be challenging. It is not clear why people are searching for these keywords. Furthermore, when using Google Trends, the sample is unknown and may not be representative, and individuals using the internet are not representative of the entire population. They are more likely to be younger, more educated, earn higher incomes, and reside in urban areas [18]. Individuals who are more likely to be severely affected by COVID-19 are not usually represented by this population [72,73]. Despite the identified limitations, previous studies suggest that web-based data provide valuable and valid results in exploring and predicting behavior and highly correlate with actual data [70,71]. Further, there are reports of rapid penetration of internet access and usage in different parts of the world, except for in regions with low internet penetration or countries with low scores in freedom of speech [22,74,75].

Conclusions

The results of this empirical infodemiology study showed that a good portion of the global population learned about the outbreak of SARS-CoV-2 and COVID-19 through web searches, particularly in the early period of the pandemic. The period covers the initial days, weeks, and months from the emergence of the novel coronavirus in January 2020 up to June 30, 2020, when the public became more aware of the pandemic, especially after the first wave [1].

The PCA showed that people used the web to learn about the ongoing COVID-19 pandemic in two ways, namely, using pastCoVepidemics keywords and using presCoVpandemic keywords. The use of pastCoVepidemics keywords in web searches nosedived as the WHO formally named the novel coronavirus and the disease (ie, "nCoV," "2019-nCoV," "SARS-CoV-2," and "COVID-19") and, therefore, as these terms became available. The trends analysis showed that web searches used to learn about COVID-19 followed a similar trend as learning about public health measures, implying that the more that people focused their attention on learning about SARS-CoV-2 and COVID-19, the more they also learned about public health measures, and vice versa. Interestingly, learning about the conspiracy theory (ConspTheory1) that 5G technology contributes to the fast global spread of COVID-19 is a predictor of people's behavior toward public health measures (PubHealthMes2). This erroneous belief makes people take precautionary measures, such as wearing a facial mask, although borne out of fear [14,15]. The same studies using the survey method also identified the same 5G-related conspiracy theory as making people respond out of fear to take precautions. This factor contributed to stockpiling of goods in the early days of the pandemic [15]. This study is the first to examine what people learned through web searches and how these influence people's social behavior toward public health safety guidelines.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary statistics of the normalized daily global Google Trends scores for different keywords used in this study. Data correspond to the time window between January 1, 2020, and June 30, 2020 (N=182 data points).
[DOCX File , 18 KB - jmir_v23i9e28975_app1.docx]

References

1. Lu H, Stratton CW, Tang YW. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. J Med Virol 2020 Apr;92(4):401-402 [FREE Full text] [doi: 10.1002/jmv.25678] [Medline: 31950516]
2. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]

3. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. Lancet 2020 Feb 29;395(10225):689-697 [FREE Full text] [doi: 10.1016/S0140-6736(20)30260-9] [Medline: 32014114]
4. Chan JF, Kok K, Zhu Z, Chu H, To KK, Yuan S, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect 2020 Jan 28;9(1):221-236 [FREE Full text] [doi: 10.1080/22221751.2020.1719902] [Medline: 31987001]
5. Woo J, Lee EY, Lee M, Kim T, Cho Y. An in vivo cell-based assay for investigating the specific interaction between the SARS-CoV N-protein and its viral RNA packaging sequence. Biochem Biophys Res Commun 2019 Dec 10;520(3):499-506 [FREE Full text] [doi: 10.1016/j.bbrc.2019.09.115] [Medline: 31594639]
6. WHO statement regarding cluster of pneumonia cases in Wuhan, China. World Health Organization. 2020 Jan 09. URL: https://www.who.int/china/news/detail/09-01-2020-who-statement-regarding-cluster-of-pneumonia-cases-in-wuhan-china [accessed 2020-12-01]
7. Labrague LJ, Yboa BC, McEnroe-Petitte DM, Lobrino LR, Brennan MGB. Disaster preparedness in Philippine nurses. J Nurs Scholarsh 2016 Jan;48(1):98-105. [doi: 10.1111/jnu.12186] [Medline: 26650188]
8. Atzrodt CL, Maknojia I, McCarthy RDP, Oldfield TM, Po J, Ta KTL, et al. A guide to COVID-19: A global pandemic caused by the novel coronavirus SARS-CoV-2. FEBS J 2020 Sep;287(17):3633-3650 [FREE Full text] [doi: 10.1111/febs.15375] [Medline: 32446285]
9. Events as they happen: Rolling updates on coronavirus disease (COVID-19). World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen [accessed 2020-10-31]
10. Dadaczynski K, Okan O, Messer M, Leung AYM, Rosário R, Darlington E, et al. Digital health literacy and web-based information-seeking behaviors of university students in Germany during the COVID-19 pandemic: Cross-sectional survey study. J Med Internet Res 2021 Jan 15;23(1):e24097 [FREE Full text] [doi: 10.2196/24097] [Medline: 33395396]
11. Rovetta A, Bhagavathula AS. COVID-19-related web search behaviors and infodemic attitudes in Italy: Infodemiological study. JMIR Public Health Surveill 2020 May 05;6(2):e19374 [FREE Full text] [doi: 10.2196/19374] [Medline: 32338613]
12. Sulyok M, Ferenci T, Walker M. Google Trends data and COVID-19 in Europe: Correlations and model enhancement are European wide. Transbound Emerg Dis 2021 Jul;68(4):2610-2615. [doi: 10.1111/tbed.13887] [Medline: 33085851]
13. Akpan IJ, Aguolu OG, Akpan A. An empirical analysis of what people learned about COVID-19 through a web search and the impacts on misinformation and attitude towards public health safety guidelines. medRxiv. Preprint posted online on February 26, 2021. [FREE Full text] [doi: 10.1101/2021.02.23.21252323]
14. Miller JM. Do COVID-19 conspiracy theory beliefs form a monological belief system? Can J Polit Sci 2020 May 21;53(2):319-326. [doi: 10.1017/S0008423920000517]
15. Miller JM. Psychological, political, and situational factors combine to boost COVID-19 conspiracy theory beliefs. Can J Polit Sci 2020 Jun 11;53(2):327-334. [doi: 10.1017/s000842392000058x]
16. Allington D, Duffy B, Wessely S, Dhavan N, Rubin J. Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. Psychol Med 2020 Jun 09;51(10):1763-1769. [doi: 10.1017/s003329172000224x]
17. Hornik R, Kikut A, Jesch E, Woko C, Siegel L, Kim K. Association of COVID-19 misinformation with face mask wearing and social distancing in a nationally representative US sample. Health Commun 2021 Jan;36(1):6-14. [doi: 10.1080/10410236.2020.1847437] [Medline: 33225745]
18. Eysenbach G. Infodemiology and infoveillance: Framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
19. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
20. Mavragani A. Infodemiology and infoveillance: Scoping review. J Med Internet Res 2020 Apr 28;22(4):e16206 [FREE Full text] [doi: 10.2196/16206] [Medline: 32310818]
21. Akpan IJ, Soopramanien D, Kwak DH. Cutting-edge technologies for small business and innovation in the era of COVID-19 global health pandemic. J Small Bus Entrepreneurship 2020 Aug 21:1-11. [doi: 10.1080/08276331.2020.1799294]
22. Akpan IJ, Ibidunni AS. Digitization and technological transformation of small business for sustainable development in the less developed and emerging economies: A research note and call for papers. J Small Bus Entrepreneurship 2021 May 18:1-7. [doi: 10.1080/08276331.2021.1924505]
23. Neely S, Eldredge C, Sanders R. Health information seeking behaviors on social media during the COVID-19 pandemic among American social networking site users: Survey study. J Med Internet Res 2021 Jun 11;23(6):e29802 [FREE Full text] [doi: 10.2196/29802] [Medline: 34043526]
24. Bautista JR, Zhang Y, Gwizdka J. Healthcare professionals' acts of correcting health misinformation on social media. Int J Med Inform 2021 Apr;148:104375. [doi: 10.1016/j.ijmedinf.2021.104375] [Medline: 33461008]
25. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G conspiracy theory: Social network analysis of Twitter data. J Med Internet Res 2020 May 06;22(5):e19458 [FREE Full text] [doi: 10.2196/19458] [Medline: 32352383]

26. Garrett L. COVID-19: The medium is the message. Lancet 2020 Mar 21;395(10228):942-943 [FREE Full text] [doi: 10.1016/S0140-6736(20)30600-0] [Medline: 32171075]
27. Fleming N. Coronavirus misinformation, and how scientists can help to fight it. Nature 2020 Jul;583(7814):155-156. [doi: 10.1038/d41586-020-01834-3] [Medline: 32601491]
28. Lee JJ, Kang K, Wang MP, Zhao SZ, Wong JYH, O'Connor S, et al. Associations between COVID-19 misinformation exposure and belief with COVID-19 knowledge and preventive behaviors: Cross-sectional online study. J Med Internet Res 2020 Nov 13;22(11):e22205 [FREE Full text] [doi: 10.2196/22205] [Medline: 33048825]
29. Kim HK, Ahn J, Atkinson L, Kahlor LA. Effects of COVID-19 misinformation on information seeking, avoidance, and processing: A multicountry comparative study. Sci Commun 2020 Sep 13;42(5):586-615. [doi: 10.1177/1075547020959670]
30. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of Google Trends in health care research: A systematic review. PLoS One 2014;9(10):e109583 [FREE Full text] [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]
31. Chung M, Oden RP, Joyner BL, Sims A, Moon RY. Safe infant sleep recommendations on the internet: Let's Google it. J Pediatr 2012 Dec;161(6):1080-1084 [FREE Full text] [doi: 10.1016/j.jpeds.2012.06.004] [Medline: 22863258]
32. Goldie JGS. Connectivism: A knowledge learning theory for the digital age? Med Teach 2016 Oct;38(10):1064-1069. [doi: 10.3109/0142159X.2016.1173661] [Medline: 27128290]
33. Dunaway MK. Connectivism: Learning theory and pedagogical practice for networked information landscapes. Ref Serv Rev 2011 Nov 15;39(4):675-685. [doi: 10.1108/00907321111186686]
34. Mian A, Khan S. Coronavirus: The spread of misinformation. BMC Med 2020 Mar 18;18(1):89 [FREE Full text] [doi: 10.1186/s12916-020-01556-3] [Medline: 32188445]
35. Lima-Pereira P, Bermúdez-Tamayo C, Jasienska G. Use of the internet as a source of health information amongst participants of antenatal classes. J Clin Nurs 2012 Feb;21(3-4):322-330. [doi: 10.1111/j.1365-2702.2011.03910.x] [Medline: 22093043]
36. Cervellin G, Comelli I, Lippi G. Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. J Epidemiol Glob Health 2017 Sep;7(3):185-189 [FREE Full text] [doi: 10.1016/j.jegh.2017.06.001] [Medline: 28756828]
37. Schunk DH. Self-regulated learning: The educational legacy of Paul R Pintrich. Educ Psychol 2005 Jun;40(2):85-94. [doi: 10.1207/s15326985ep4002_3]
38. Akpan IJ. Formative assessment as a technique for improving student engagement and learning effectiveness in self-regulated e-learning environment. In: Proceedings of the IEICE Information and Communication Technology Forum. 2014 Aug 10 Presented at: IEICE Information and Communication Technology Forum; May 28-30, 2014; Poznań, Poland p. 34 URL: https://www.ieice.org/publications/proceedings/bin/pdf_link.php?fname=COMM1-2.pdf&iconf=ICTF&year=2014&vol=19&number=COMM1-2&lang=E [doi: 10.34385/proc.19.COMM1-2]
39. Donthu N, Gustafsson A. Effects of COVID-19 on business and research. J Bus Res 2020 Sep;117:284-289 [FREE Full text] [doi: 10.1016/j.jbusres.2020.06.008] [Medline: 32536736]
40. Tuite AR, Zhu L, Fisman DN, Salomon JA. Alternative dose allocation strategies to increase benefits from constrained COVID-19 vaccine supply. Ann Intern Med 2021 Apr;174(4):570-572 [FREE Full text] [doi: 10.7326/M20-8137] [Medline: 33395334]
41. Shrivastava S, Shrivastava P. Minimizing the risk of international spread of coronavirus disease 2019 (COVID-19) outbreak by targeting travelers. J Acute Dis 2020;9(2):47. [doi: 10.4103/2221-6189.278654]
42. van Doremalen N, Bushmaker T, Morris DH, Holbrook MG, Gamble A, Williamson BN, et al. Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. N Engl J Med 2020 Apr 16;382(16):1564-1567. [doi: 10.1056/nejmc2004973]
43. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases. Radiology 2020 Aug;296(2):E32-E40 [FREE Full text] [doi: 10.1148/radiol.2020200642] [Medline: 32101510]
44. Interim clinical guidance for management of patients with confirmed coronavirus disease (COVID-19). Centers for Disease Control Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html [accessed 2020-12-02]
45. COVID-19 Treatment Guidelines Panel. Coronavirus Disease 2019 (COVID-19) Treatment Guidelines. Bethesda, MD: National Institutes of Health; 2020. URL: https://files.covid19treatmentguidelines.nih.gov/guidelines/covid19treatmentguidelines.pdf [accessed 2021-08-25]
46. Soo RJJ, Chiew CJ, Ma S, Pung R, Lee V. Decreased influenza incidence under COVID-19 control measures, Singapore. Emerg Infect Dis 2020 Aug;26(8):1933-1935 [FREE Full text] [doi: 10.3201/eid2608.201229] [Medline: 32339092]
47. Mavragani A, Ochoa G. Google Trends in infodemiology and infoveillance: Methodology framework. JMIR Public Health Surveill 2019 May 29;5(2):e13439 [FREE Full text] [doi: 10.2196/13439] [Medline: 31144671]
48. Akpan IJ, Akpan AA. Multiple criteria analysis of the popularity and growth of research and practice of visual analytics, and a forecast of the future trajectory. Int Trans Oper Res 2021 Feb 26;28(5):2275-2298. [doi: 10.1111/itor.12952]
49. Brodeur A, Clark AE, Fleche S, Powdthavee N. COVID-19, lockdowns and well-being: Evidence from Google Trends. J Public Econ 2021 Jan;193:104346 [FREE Full text] [doi: 10.1016/j.jpubeco.2020.104346] [Medline: 33281237]

50. Masters-Waage TC, Jha N, Reb J. COVID-19, coronavirus, Wuhan virus, or China virus? Understanding how to "do no harm" when naming an infectious disease. Front Psychol 2020;11:561270 [FREE Full text] [doi: 10.3389/fpsyg.2020.561270] [Medline: 33362626]
51. Jo S, Kim H, Kim S, Shin DH, Kim M. Characteristics of flavonoids as potent MERS-CoV 3C-like protease inhibitors. Chem Biol Drug Des 2019 Dec;94(6):2023-2030 [FREE Full text] [doi: 10.1111/cbdd.13604] [Medline: 31436895]
52. Xie M, Chen Q. Insight into 2019 novel coronavirus - An updated interim review and lessons from SARS-CoV and MERS-CoV. Int J Infect Dis 2020 May;94:119-124 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.071] [Medline: 32247050]
53. Mahase E. Coronavirus COVID-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. BMJ 2020 Feb 18;368:m641. [doi: 10.1136/bmj.m641] [Medline: 32071063]
54. Naming the coronavirus disease (COVID-19) and the virus that causes it. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it [accessed 2020-12-01]
55. Coronavirus disease (COVID-19) weekly epidemiological update and weekly operational update. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2020-12-01]
56. Chary MA, Overbeek DL, Papadimoulis A, Sheroff A, Burns MM. Geospatial correlation between COVID-19 health misinformation and poisoning with household cleaners in the Greater Boston Area. Clin Toxicol (Phila) 2021 Apr;59(4):320-325. [doi: 10.1080/15563650.2020.1811297] [Medline: 32901533]
57. Matias T, Dominski FH, Marks DF. Human needs in COVID-19 isolation. J Health Psychol 2020 Jun;25(7):871-882. [doi: 10.1177/1359105320925149] [Medline: 32375564]
58. Leung CC, Lam TH, Cheng KK. Mass masking in the COVID-19 epidemic: People need guidance. Lancet 2020 Mar 21;395(10228):945 [FREE Full text] [doi: 10.1016/S0140-6736(20)30520-1] [Medline: 32142626]
59. De Giorgio A. COVID-19 is not just a flu. Learn from Italy and act now. Travel Med Infect Dis 2020;35:101655 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101655] [Medline: 32272199]
60. Akpan IJ, Aguolu OG, Ezeume IC. Overcoming the challenge of communicating the concept and science of SARS-CoV-2 and COVID-19 to non-experts. Challenge 2021 May 04;64(2):117-131. [doi: 10.1080/05775132.2021.1912984]
61. Roscoe RD, Grebitus C, O'Brian J, Johnson AC, Kula I. Online information search and decision making: Effects of web search stance. Comput Human Behav 2016 Mar;56:103-118. [doi: 10.1016/j.chb.2015.11.028]
62. Sanchez-Caballero S, Selles MA, Peydro MA, Perez-Bernabeu E. An efficient COVID-19 prediction model validated with the cases of China, Italy and Spain: Total or partial lockdowns? J Clin Med 2020 May 20;9(5):1547 [FREE Full text] [doi: 10.3390/jcm9051547] [Medline: 32443871]
63. Akpan IJ, Shanker M, Razavi R. Improving the success of simulation projects using 3D visualization and virtual reality. J Oper Res Soc 2019 Sep 12;71(12):1900-1926. [doi: 10.1080/01605682.2019.1641649]
64. Pickles K, Cvejic E, Nickel B, Copp T, Bonner C, Leask J, et al. COVID-19 misinformation trends in Australia: Prospective longitudinal national survey. J Med Internet Res 2021 Jan 07;23(1):e23805 [FREE Full text] [doi: 10.2196/23805] [Medline: 33302250]
65. Krehbiel TC. SAS. In: Ross PW, editor. The Handbook of Software for Engineers and Scientists. Boca Raton, FL: CRC Press; May 04, 2018:1047-1066.
66. Hölscher C, Strube G. Web search behavior of internet experts and newbies. Comput Netw 2000 Jun;33(1-6):337-346. [doi: 10.1016/s1389-1286(00)00031-1]
67. Akpan IJ, Udoh EAP, Adebisi B. Small business awareness and adoption of state-of-the-art technologies in emerging and developing markets, and lessons from the COVID-19 pandemic. J Small Bus Entrepreneurship 2020 Sep 29:1-18. [doi: 10.1080/08276331.2020.1820185]
68. Lebow JL. The challenges of COVID-19 for divorcing and post-divorce families. Fam Process 2020 Sep;59(3):967-973 [FREE Full text] [doi: 10.1111/famp.12574] [Medline: 32594521]
69. Akpan IJ, Ezeume IC. The challenges faced by parents and children from divorce. Challenge 2020 Nov 09;63(6):365-377. [doi: 10.1080/05775132.2020.1842025]
70. Mavragani A, Ochoa G. The internet and the anti-vaccine movement: Tracking the 2017 EU measles outbreak. Big Data Cognit Comput 2018 Jan 16;2(1):2-18. [doi: 10.3390/bdcc2010002]
71. Jun S, Yoo HS, Choi S. Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. Technol Forecast Soc Change 2018 May;130(2018):69-87. [doi: 10.1016/j.techfore.2017.11.009]
72. Ko J, Danielson M, Town M, Derado G, Greenlund K, Kirley P, COVID-NET Surveillance Team. Risk factors for coronavirus disease 2019 (COVID-19)-associated hospitalization: COVID-19-Associated Hospitalization Surveillance Network and Behavioral Risk Factor Surveillance System. Clin Infect Dis 2021 Jun 01;72(11):e695-e703 [FREE Full text] [doi: 10.1093/cid/ciaa1419] [Medline: 32945846]
73. Pijls BG, Jolani S, Atherley A, Derckx RT, Dijkstra JIR, Franssen GHL, et al. Demographic risk factors for COVID-19 infection, severity, ICU admission and death: A meta-analysis of 59 studies. BMJ Open 2021 Jan 11;11(1):e044640 [FREE Full text] [doi: 10.1136/bmjopen-2020-044640] [Medline: 33431495]
74. Akpan IJ, Akpan AA. The impact of internet use on students' learning outcomes in higher education in developing countries. Int J Education Res 2017 Sep;12(1):45-54 [FREE Full text]

75. Mavragani A, Tsagarakis KP. Predicting referendum results in the big data era. J Big Data 2019 Jan 14;6:1-20 [FREE Full text] [doi: 10.1186/s40537-018-0166-z]

Abbreviations

2019-nCoV: 2019 novel coronavirus
CDC: Centers for Disease Control and Prevention
ConspTheory1: conspiracy theory 1
ConspTheory2: conspiracy theory 2
MERS: Middle East respiratory syndrome
pastCoVepidemics: keywords that address previous epidemics
PCA: principal component analysis
presCoVpandemic: keywords that explain the ongoing pandemic
PubHealthMes1: public health measures 1
PubHealthMes2: public health measures 2
Q-Q: quantile-quantile
SARS: severe acute respiratory syndrome
WHO: World Health Organization

Edited by C Basch; submitted 01.05.21; peer-reviewed by A Mavragani, R Poluru; comments to author 13.06.21; revised version received 09.07.21; accepted 09.07.21; published 02.09.21.

Please cite as:
Akpan IJ, Aguolu OG, Kobara YM, Razavi R, Akpan AA, Shanker M
Association Between What People Learned About COVID-19 Using Web Searches and Their Behavior Toward Public Health Guidelines: Empirical Infodemiology Study
J Med Internet Res 2021;23(9):e28975
URL: https://www.jmir.org/2021/9/e28975
doi:10.2196/28975
PMID:34280117

Original Paper

An Evaluation Service for Digital Public Health Interventions: User-Centered Design Approach

Kassandra Karpathakis[1], BSc; Gene Libow[2], BA; Henry W W Potts[3], PhD; Simon Dixon[4], BA; Felix Greaves[5,6*], PhD; Elizabeth Murray[7*], FRCGP, FRCP (Edin), PhD

[1]NHS Artificial Intelligence Laboratory, NHSX, London, United Kingdom

[2]Independent Service Design Consultant, London, United Kingdom

[3]Institute of Health Informatics, University College London, London, United Kingdom

[4]NHSX, London, United Kingdom

[5]Department of Primary Care and Public Health, Imperial College London, London, United Kingdom

[6]Science, Evidence and Analysis, National Institute of Health and Care Excellence, London, United Kingdom

[7]Research Department of Primary Care and Population Health, University College London, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Henry W W Potts, PhD
Institute of Health Informatics
University College London
222 Euston Road
London, NW1 2DA
United Kingdom
Phone: 44 20 3549 5303
Email: h.potts@ucl.ac.uk

Abstract

Background: Digital health interventions (DHIs) have the potential to improve public health by combining effective interventions and population reach. However, what biomedical researchers and digital developers consider an effective intervention differs, thereby creating an ongoing challenge to integrating their respective approaches when evaluating DHIs.

Objective: This study aims to report on the Public Health England (PHE) initiative set out to operationalize an evaluation framework that combines biomedical and digital approaches and demonstrates the impact, cost-effectiveness, and benefit of DHIs on public health.

Methods: We comprised a multidisciplinary project team including service designers, academics, and public health professionals and used user-centered design methods, such as qualitative research, engagement with end users and stakeholders, and iterative learning. The iterative approach enabled the team to sequentially define the problem, understand user needs, identify opportunity areas, develop concepts, test prototypes, and plan service implementation. Stakeholders, senior leaders from PHE, and a working group critiqued the outputs.

Results: We identified 26 themes and 82 user needs from semistructured interviews (N=15), expressed as 46 Jobs To Be Done, which were then validated across the journey of evaluation design for a DHI. We identified seven essential concepts for evaluating DHIs: evaluation thinking, evaluation canvas, contract assistant, testing toolkit, development history, data hub, and publish health outcomes. Of these, three concepts were prioritized for further testing and development, and subsequently refined into the proposed PHE Evaluation Service for public health DHIs. Testing with PHE's Couch-to-5K app digital team confirmed the viability, desirability, and feasibility of both the evaluation approach and the Evaluation Service.

Conclusions: An iterative, user-centered design approach enabled PHE to combine the strengths of academic and biomedical disciplines with the expertise of nonacademic and digital developers for evaluating DHIs. Design-led methodologies can add value to public health settings. The subsequent service, now known as *Evaluating Digital Health Products*, is currently in use by health bodies in the United Kingdom and is available to others for tackling the problem of evaluating DHIs pragmatically and responsively.

(J Med Internet Res 2021;23(9):e28356) doi:10.2196/28356

KEYWORDS

digital health; internet-based interventions; mHealth; evaluation studies; public health; human-centered design; service design; mobile phone

Introduction

For public health interventions to significantly impact healthy life expectancy in a population, reach has to be combined with efficacy [1]. Reach is the proportion of the target population affected by an intervention, and efficacy is the effect of an intervention on the individuals it reaches. There is considerable interest in the potential of digital health interventions (DHIs) to improve public health. DHIs are delivered on digital platforms, such as the web or smartphone apps intended to deliver health care or health promotion [2]. Such DHIs are expected to combine the reach of large-scale population initiatives, such as media advertising, with the efficacy of individual treatments. Furthermore, DHIs are intended to increase both capacity and access to public health initiatives by providing services in areas where face-to-face options are unavailable or unable to meet demand.

Although there is evidence for the efficacy of many DHIs in public health [3-10], few studies indicate that their public health potential is being realized in practice. What constitutes success is viewed differently by biomedical researchers and digital product developers. Biomedical research on digital health is heavily influenced by the pharmaceutical model with a focus on trials and effectiveness, whereas digital developers often focus on usability and acceptability [11]. However, both perspectives are required [11]. A digital product cannot be successful at scale without satisfactory usability. However, user ratings can be insufficient, sometimes bearing either no relation or even an inverse relation to the effectiveness of DHIs [12-14].

As a public health body with a finite budget and responsibility for improving population health, Public Health England (PHE) was at the forefront of considering how best to use DHIs to improve health outcomes and evaluate their value (measured as an improvement to public health). Multiple biomedical and digital approaches to the evaluation of DHIs exist and are described and critiqued elsewhere [2,11,15-18]. This paper reports on a project by PHE to develop and operationalize an evaluation framework that combines these approaches with the goal of demonstrating the impact, cost-effectiveness, and benefit of DHIs on public health.

Methods

Design

User-centered design (UCD), applied by a multidisciplinary team, was used to synthesize the strengths of digital evaluation approaches with those of the biomedical approach. The project objectives were as follows:

1. Identify core audiences and stakeholders for evaluation of DHIs (user groups)
2. Identify the needs of users for evaluating DHIs, including barriers and facilitators
3. Identify the key performance indicators (KPIs) and outcomes that different audiences and stakeholders consider important in evaluating DHIs
4. Identify evaluation methods (ways of conducting an evaluation study), tools (resources that can aid in carrying out an evaluation), and metrics (standard quantitative measures that can be used in an evaluation) applicable to DHIs
5. Prototype and test an evaluation approach for DHIs used in a public health context

User-Centered Design

A modified UCD approach, known as service design, was adopted. The UCD approach bases the design, testing, and development of a product or service on the needs of the users affected by it [13]. This approach, which began as a focus on the end user's interaction with products, has evolved into designing both products and services. It considers the needs of a broad range of users, including organizational users providing the service, users responsible for the finances and direction of the service, and other users in the service ecosystem. This holistic evolution is reflected in service design [19,20], a choreography of the processes, technologies, and interactions needed to set up and drive a service through a UCD perspective [21].

Service design helps reduce risk by investing upfront in validating a service proposition with end users before physically building the product or service (desirability and usability), clarifying what is required of the organization to deliver the service (feasibility), clarifying the potential business impact of the service (viability), having a strong focus on outcomes, and embedding the host organization's team that will eventually run the service in research and design activities [22]. By designing to meet the needs of users, service design methods are intended to ensure that the resultant service is more likely to be desirable, usable, feasible, and viable. Throughout this study, the term UCD refers to the service design variation.

The project team followed the English Government Digital Service's agile approach, comprising discovery, alpha, beta, and live phases [23]. This approach has subsequently been adopted by governments worldwide. This paper reports the discovery and alpha phases.

Setting and Project Team

At the time of the project, PHE was the national public health agency in England responsible for protecting and improving the nation's health and well-being and reducing health inequalities. The discovery phase was delivered from May to June 2018 and the alpha phase was delivered from August 2018 to March 2019. The beta phase commenced in July 2019 and was completed in 2021.

The project team was established using a competitive tender [19], where applicants read a seminal paper [2] about evaluating DHIs and presented an example of how they would integrate

evaluation into the design and development of a DHI. The successful applicant, Livework Studio Ltd, a service design agency, worked alongside the internal PHE team to form the overall project team (the project team).

In response to this challenge, Livework created a visual model showing how different evaluation approaches and metrics could be integrated into the design and experience of a DHI. The model or design object (Figure S1 in Multimedia Appendix 1) tangibly represents the problem of evaluating DHIs for various stakeholders. The model was iterated as the project progressed.

PHE formed a working group with representatives across the English health system, including the Department of Health and Social Care, National Institute of Health and Care Excellence (NICE), University College London (UCL), and the World Health Organization. The working group provided knowledge on the wider digital health landscape and sense-checked research and deliverables.

The project team conducted 14 *show and tells* in the discovery and alpha phases, wherein the project's progress was shared with a wider group of stakeholders (*Show and tells* are opportunities for team members to showcase their work to their team and wider stakeholders; it is a project management and engagement technique commonly used by digital professionals). This included sharing project plans, early research findings, and design outputs from different research sessions. Stakeholder feedback helped validate or raise concerns with the findings and linked the project team to initiatives in the wider digital health ecosystem.

Ethics

This project complied with the code of conduct of the Market Research Society as required by PHE's Research Ethics and Governance Group, and informed consent was obtained from the interview participants. The project's methodology and results were presented at two government-mandated service assessments [23] at the end of the discovery and alpha phases to ensure adherence to the government Service Standard [24] for digital projects.

Recruitment

Three categories of professionals involved in the design, development, and commissioning of public health DHIs were identified: academics, digital product developers (hereafter referred to as *digital developers*), and public health professionals. Research participants (N=15) for the interviews were selected to reflect these user groups. Additional participants were recruited in the following stages.

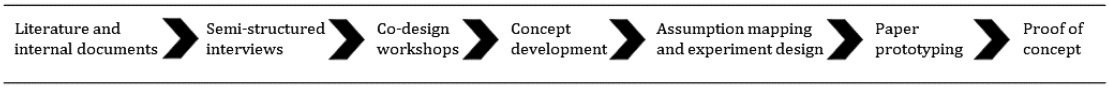
Academics were recruited for their expertise in developing and evaluating DHIs, as reflected in their publications and national reputation. Digital professionals with a track record of successful development and wide-scale deployment of DHIs were selected, including those working for PHE, National Health Service England and Improvement (NHSE/I), and the private sector. Public health professionals were selected for their experience of commissioning or delivering public health services via digital means and their strategic overview of the challenges involved. Within digital and public health professionals, those with and without experience of evaluating DHIs were sampled.

Results

Iterative Data Collection Methods Presented With Results

We used seven data collection methods: (1) review of the literature and internal documents, (2) semistructured interviews, (3) co-design workshops, (4) concept development, (5) assumption mapping and experiment design, (6) paper prototyping, and (7) proof of concept. The outputs of each stage were the inputs for the next stage (Figure 1). Through a series of iterative developments, the project team progressively evolved its understanding of the problem, user needs, and potential solutions. The seven data collection and analysis methods, along with the results of each stage are presented in the following sections.

Figure 1. The seven data collection methods with the output of each stage becoming the input for the next stage.



Review of the Literature and Internal Documents

Methodology

Internal PHE documents and sample publications on the evaluation of DHIs provided by PHE staff, including the paper by Murray et al [2] used in the supplier challenge, were reviewed. The aims of the review were to (1) develop an understanding of the problem space, (2) identify themes pertaining to evaluating DHIs—including barriers and facilitators—for exploration in semistructured interviews (project objective 2), and (3) identify evaluation methods, tools, and metrics applicable to DHIs (project objective 4). The evaluation methods identified were categorized into academic, health, economic, and digital methods. For each method, a simple description, scenarios of use, pros and cons, cost, time, and existing tools and guidance were captured.

The output of stage 1 formed the basis of a discussion guide (Multimedia Appendix 2) for the semistructured interviews (stage 2). Additional publications were collected and synthesized throughout the study.

Corresponding Results

The barriers identified were organizational and disciplinary silos associated with public health commissioning, design and development of digital services, and academic evaluation; unclear expectations and roles; different disciplinary expectations on what tools and methods to use for evaluation and how to measure success or failure; limited time, capacity, and funding; lack of evaluation experience; and rapidly evolving DHIs (inherent tension with resource- and time-intensive evaluations).

The facilitators identified were clear organizational goals (outcomes) and decision-making processes, prioritizing evaluation from the outset, ring-fenced funding, team capacity and resources, guidance and case studies, communities of practice, interdisciplinary collaboration, data standards, and constructive feedback.

The resultant discussion guide (Multimedia Appendix 2) was structured around the process of conducting an evaluation of a DHI, as identified from the review of the literature and internal PHE documents. Discussion guide sections included unawareness and initiation of an evaluation; setup, information, guidance, and tools; data, decisions, and outputs; and wrap up, outputs, and outcomes. Questions within each section were created to further explore themes pertaining to evaluating DHIs, including decision-making; intervention mechanisms of action; organizational obligations (to evaluate); metrics; success criteria; and feedback loops. For the different user groups, additional questions exploring the barriers to and facilitators of evaluating DHIs were added. In total, 46 evaluation methods were identified and categorized (Textbox 1; additional information captured is available upon request).

Textbox 1. Evaluation methods identified from the review of literature and internal documents.

Academic

- Cluster randomized controlled trial (RCT)
- Consultative or deliberative methods
- Documentary analysis
- Ethnography
- Factorial RCT
- Feasibility RCT
- Focus group
- Individual and group interviews
- Interrupted time series
- Natural experiment
- Nonrandomized controlled before and after
- Process evaluation
- Propensity score-matched control study
- RCT
- Sequential multiple assignment randomized trial
- Stepped-wedge RCT
- Survey of digital health intervention users
- Surveys of target population for digital health intervention
- Uncontrolled before and after
- Qualitative comparative analysis
- Qualitative evaluation
- Quantitative observation studies

Digital

- A/B testing
- Advertising data
- App Store data
- Cohort analysis
- Digital uptake
- Email tracking
- Funnel analysis
- Guerrilla user research
- In app analytics
- Likes or shares
- Net promoter score
- Microrandomized trials
- Notifications
- Web-based survey or polling
- Search traffic
- Self-reporting
- Simulation modeling
- Social media research

- Usability testing (laboratory based)
- Wearable device data
- Web analytics

Health Economic

- Cost-benefit analysis
- Cost consequence
- Cost-effectiveness analysis
- Cost utility

Semistructured Interviews

Methodology

Semistructured interviews were conducted to identify user needs (project objectives 1 and 2). Interviews explored participants' roles in relation to commissioning, designing, or developing DHIs; their understanding of evaluation; and their previous evaluation experience. Subsequent questions focused on the process of evaluation and its evolution as a DHI are developed, along with the key themes, barriers, and facilitators identified in stage 1.

Interviews were conducted by service designers with expertise in design research, lasted 60-90 minutes, and were audiorecorded. Detailed field notes were also provided. Both field notes and transcribed interviews formed a part of the data set.

An interview capture template was used to highlight and thematically organize the interview data. After each interview, the researcher reviewed their field notes and transcripts, highlighting the points of interest. These points were coded using a common set of themes based on the stages of the evaluation process (before, beginning, during, and after) and known areas of interest, such as collaboration, clinical risk, support and guidance, technology, and data, as identified in stage 1. If an observation did not obviously fit within an existing theme, a new theme was created to code it, or it was highlighted for further review by the wider project team.

Key user statements and interview observations generated the first iteration of user needs [18], including needs being met by current means of evaluating DHIs and needs that remained unmet or were causing difficulty and frustration in evaluation efforts. Identified user needs were articulated using an accepted UCD template called Jobs To Be Done (JTBD), which identifies the type of user, their need, when the need arises, and the outcome they are trying to achieve [25]. This structured articulation clearly communicates the user's needs to stakeholders and the project team.

After the initial round of analysis, the project team collectively reviewed all interviews; shared points of interest, observations, and user needs; and clustered them according to the common themes. User needs were mapped against a common user journey—a way of showing a user's journey through a service across time [26]. This produced an overview of where and when user needs occurred throughout the process of designing a DHI evaluation.

Corresponding Results

A total of 15 semistructured interviews were completed, five per target user group (Table 1). Some participants were invited to take part in multiple data collection stages, with 6 interviewees participating in the co-design workshops (stages 3 and 4), 4 co-design workshop attendees participating in paper prototyping (stage 7), and 1 interviewee participating in the proof of concept (stage 8).

In total, 26 themes about evaluating DHIs were identified and validated (Textbox S1, Multimedia Appendix 1). These themes were organized by a user group (academic, digital, and public health) and an evaluation stage (overview; unawareness and initiating an evaluation; setup, information, guidance, and tools; data, decisions, and outputs; outcomes, feedback, and decisions; and incidents and changes).

A total of 82 JTBD were generated (18/82, 21% academic; 37/82, 45% digital; and 27/82, 32% public health), derived from the outputs of stages 1 and 2. Table 2 shows an example of the semistructured interview analysis and the resulting JTBD.

The initial set of JTBDs was refined and augmented through feedback from stakeholders at the show and tells. The project team then reviewed the outputs and distilled the superset of research findings and user needs into a smaller subset of user needs or JTBD: (1) representing the essential stages of the journey, (2) most essential stages (ie, if the need was not met then subsequent needs could not be met), and (3) those stages most strongly indicated by research stages 1 and 2 (semistructured interviews and publications). The refined 46 JTBDs (13/46, 28% academic; 20/46, 43% digital; and 13/46, 28% public health) and the outputs of stage 1 formed the input for the first co-design workshop.

Table 1. Overview of participant characteristics at data collection stages 2, 3, 4, 7, and 8.

Participants	Semistructured interviews (stage 2; n=15)	Co-design workshop 1 (stage 3; n=10)	Co-design workshop 2 (stage 4; n=10)	Paper prototypes (stage 7; n=11)	Proof of concept (stage 8; n=6)
User type, n (%)					
Academic	5 (33)	1 (10)	3 (30)	2 (18)	N/A[a]
Digital	5 (33)	5 (50)	4 (40)	6 (55)	6 (100)
Public health	5 (33)	4 (40)	3 (30)	3 (27)	N/A
Organizations	• Be Mindful • Fresh Egg • Lancashire County Council • Mindtech • NHSE/I[b] • PHE[c] • SH:24 • The University of Edinburgh • UCL[d] • WHO[e]	• Mindwave • PHE • SH:24 • UCL	• DHSC[f] • King's College London • NICE[g] • PHE • Unboxed • WHO	• DHSC • PHE	• PHE

[a]N/A: not applicable.

[b]NHS E/I: National Health Service England and Improvement.

[c]PHE: Public Health England.

[d]UCL: University College London.

[e]WHO: World Health Organization.

[f]DHSC: Department of Health and Social Care.

[g]NICE: National Institute of Health and Care Excellence.

Table 2. For each user group, we provide an example aligned theme, illustrative quote, and Jobs To Be Done mapped to the relevant evaluation stage.

User group	Evaluation stage	Theme	Illustrative quote	Jobs To Be Done
Academic	Stage 2: setup, information, guidance, and tools	Evaluation methods	"You have a set of tools and paradigms that are going to be suitable for different types of problems. When you're investigating different types of problems, for particular types of stakeholders, which particular types of objectives, or even particular types of technology, or particular stages of innovation, you have to be careful about finding the right thing and not trying to squeeze a round peg into a square hole."	• As an evaluator • When I design/set up an evaluation • I need access to a range of evaluation tools and paradigms • So that I can choose the one fit for the problem
Digital	Stage 4: outcomes, feedback, and decisions	Design process	"It's fundamental to my role in a non-academic non-traditional sense, as far as a non-empirical sense, because I'm the one who manages [metric company name] and given what it takes to insure and also the research we conduct with users to define and validate services prior to committing resources developing them. But also to maximize them later, so we use informal or design led evaluation means to validate, to research, to prove assumptions prior to designing things."	• As a digital professional • When deciding what to design and how to design it • I need to validate service propositions by proving assumptions • So that I can be confident in committing resources developing them
Public health	Stage 0: overview	Funding and costs	"First, we're using it to channel shift. So, if we can get people to use digital interventions rather than face-to-face interventions which are much more expensive, then we're developing a range of products and services around that. On sexual health, we have a new program on contraceptive access for women to help them to make choices online and to get their contraceptive online rather than going in through GP[a] services."	• As a Director of Public Health and Well-being • When I am planning service changes • I need to know if people will use a digital approach • So that we can save money by shifting service from more expensive face-to face-services

[a]GP: general practitioner.

Co-design Workshop 1

Overview

Two co-design workshops were held to (1) validate the findings from semistructured interviews (project objectives 1 and 2), (2) help create a user journey that reflected the user requirements for all three user groups (project objective 2), (3) identify evaluation methods, tools, and metrics applicable to DHIs (project objectives 2, 3, and 4), and (4) consider the structure of DHIs (ie, intervention component and delivery component; project objectives 4 and 5).

Methodology

Findings from stages 1 and 2 were articulated as JTBDs or user needs for each of the three user groups. These were mapped onto a user journey reflecting all stages for meeting the goal of evaluating a DHI (Figure S2; Multimedia Appendix 1). This mapping was used to reflect the timings and interdependencies of each step and the user needs preceding the evaluation of a DHI.

Workshop participants were divided into groups of three: one person from each profession (digital, public health, and academic). These trios worked through the draft user journey from the perspective of a specified user group, discussing reactions, validating the user needs that they recognized, editing where necessary, and adding missing user needs. The exercise was repeated twice until each trio reviewed the draft journey from the perspective of all three user groups. Reviewing all three sets of user needs promoted the understanding of other perspectives among the workshop participants and the project team.

Participants then reviewed and edited a catalog of 46 evaluation methods and tools collated by the project team in stages 1 and 2. Participants considered different ways of categorizing them according to the timing in the journey of planning and conducting an evaluation.

Corresponding Results

Workshop participants validated, refined, or rearticulated the 46 JTBDs (13/46, 28% academic; 20/46, 43% digital; and 13/46, 28% public health) across the stages of the evaluation journey. A further 27 JTBDs (8/27, 29% academic; 1/27, 3% digital; and 18/27, 66% public health) were added by workshop participants. The project team had a total of 73 JTBDs (21/73, 28% academic; 21/73, 28% digital; and 31/73, 42% public health) at the end of the first workshop.

The first co-design workshop structure also exposed participants representing the three user groups to the aims, perspectives, and corresponding needs of the other user groups. Participants were observed sharing points of view, articulating their needs that were not self-evident to other user groups, and learning where their respective needs coincided. Furthermore, the different user groups learned about the scope and responsibilities of each other's roles and where interdependencies between their needs and evaluation activities were.

Workshop participants reviewed, validated, and differentiated the evaluation catalog into evaluation methods, tools, or metrics. Of the initial 46 evaluation methods, tools, and metrics, six were amended, and the participants added three methods, three tools, and 14 metrics.

Participants were observed learning about evaluation methods, tools, and metrics they had not previously used, including their benefits and potential scenarios of use. Participants from different user groups shared with each other how they used the evaluation methods, tools, and metrics in their own role and how they chose the appropriate one, for example, based on the potential health risk of a DHI.

After the first co-design workshop, the project team collectively reviewed the outputs and distilled the 73 JTBDs (21/73, 28% academic; 21/73, 28% digital; and 31/73, 42% public health) into a smaller subset to be used as design stimulus for the second co-design workshop. Following the same process for selecting a subset of user needs as performed in stage 2, the subset of JTBDs was selected by (1) representing the essential stages of the journey, (2) most essential need (ie, if the need was not met then subsequent needs could not be met), (3) those needs most strongly indicated by research stages 1-3, and (4) those needs that were actionable and useful as stimuli for a design workshop.

The project team prioritized 9 JTBDs (3/9, 33% academic; 3/9, 33% digital; and 3/9, 33% public health; Table 3) for input into the second co-design workshop. The resultant catalog of evaluation methods, tools, and metrics from the first co-design workshop formed the basis of an evaluation method catalog (project objective 4) used in stage 8 (proof of concept: prototype of full-service experience).

Table 3. The nine Jobs To Be Done prioritized for co-design workshop 2, with three prioritized per user group.

User group	JTBD[a] 1	JTBD 2	JTBD 3
Public health	• As a public health professional • When reviewing the outcomes of an evaluation • I need the flexibility within the project and team to make decisions that change the direction of the project • So that learnings can be used to change the service for the better	• As a public health professional • When evaluating a service • I need to know if the service is having an impact on key metrics • So that I can make an investment case for rolling the service out at scale	• As a commissioner • When writing a brief • I need to set expectations around metrics, methods, and implementation as well as provider skills and capabilities • So that providers can build these into their solution design and implementation plan
Academic	• As an evaluator • When evaluating a health product or service • I need access to clean, accessible and linked data from across the health system • So that I can do my evaluation	• As an evaluator • When designing or setting up an evaluation • I need access to a range of evaluation tools and paradigms • So that I can choose the one fit for the problem	• As an academic evaluator • When doing evaluation • I need an in-depth understanding of the intervention and the pathway of action • So that I can properly evaluate it
Digital	• As a digital professional • When evaluating a service qualitatively • I need to be able to observe users • So that I can understand why they are having difficulties	• As a digital professional • When planning a DHI[b] • I need to know what sort of evaluations might need to take place • So that I can be prepared to participate	• As a digital professional • When deciding what to design and how to design it • I need to validate service propositions by testing assumptions • So that I can be confident in committing resources to develop them

[a]JTBD: Jobs To Be Done.

[b]DHI: digital health intervention.

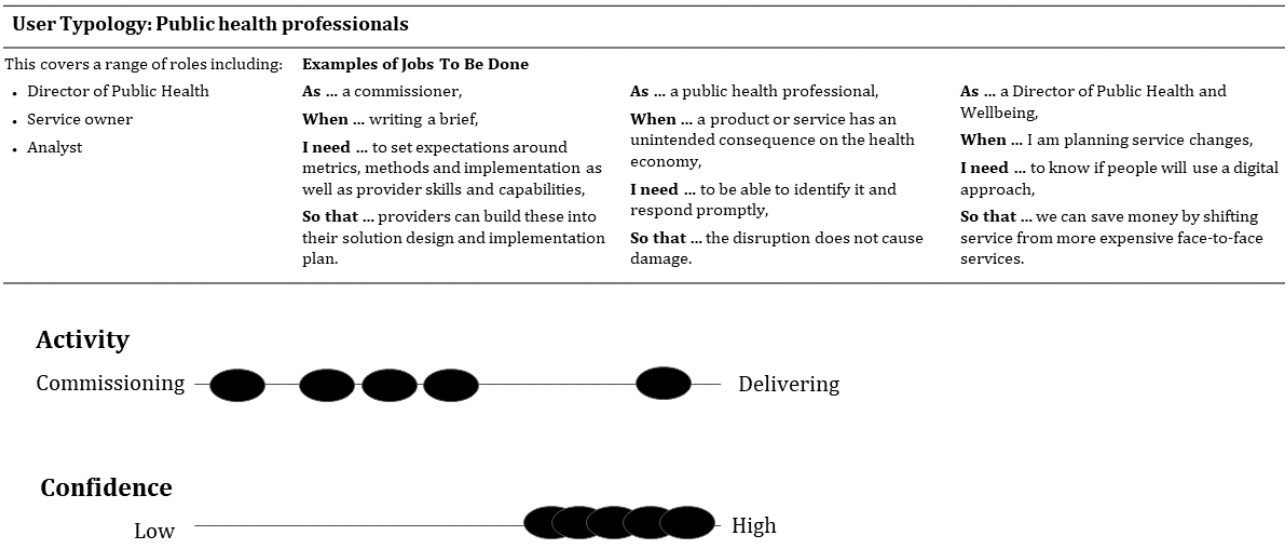
Co-design Workshop 2

Methodology

The second co-design workshop was built on the first and aimed at designing concepts for an evaluation framework for DHIs in a public health context (project objective 5). Attendees worked in groups of three, as before. Using a DHI case study, the attendees created a concept for an Evaluation Service by imagining how a user would conduct an evaluation in each case. Stimulus materials, comprising interactive design artifacts and JTBD from stages 1 to 3 and those specifically created for the workshop (Figure S3 in Multimedia Appendix 1), provided some structure for creating the concept. Attendees synthesized these inputs into a service concept that they named, described, and visualized using a concept template (Figure S3 in Multimedia Appendix 1). Each concept was shared with all workshop attendees for feedback.

Workshop outputs included a series of eight conceptual drawings and descriptions. Using the outputs from stages 1 to 4, the project team further articulated the different user groups as user typologies [27] based on key dimensions of difference (Figure 2).

Figure 2. User typology for public health professionals evaluating a digital health intervention (black circles indicate the distribution of participants in stage 2 semistructured interviews).



Corresponding Results

The workshop produced eight raw concepts considered essential by the participants for the DHI evaluation framework:

1. Integrated evaluation: builds in evaluation thinking, skills, and tools at the beginning of the DHI development process.
2. Parallel playbook: a series of experiments to validate DHI propositions and iterate these to validate the impact.
3. Contract guidance: a person to help write contract terms, thereby ensuring providers build evaluation into the design and agree to cease or change course if the desired outcomes are not indicated.
4. Measurements aggregator: a way to discover and integrate health outcome measures into one place. This would enable digital teams to explore health outcome data and make informed decisions when developing their DHI.
5. Standardized way of collecting or structuring data: a visible, accessible history of the development process for a DHI and how it was evaluated.
6. Conventional medical model: linking user data from a DHI to how it affects end users' health outcomes.
7. Access shortcuts: an NHS data owner who has an agreement with the local authority and clinical commissioning group for sharing and accessing data to inform DHI development and evaluation.
8. Innovation framework: a PHE open-source innovation framework for understanding the DHI context and sharing data.

Figure 2 shows the user group typologies designed using the dimensions of activity in evaluation (commissioning or delivering) and confidence in evaluation (low or high).

Concept Development and Prioritization

Methodology

Stages 1 to 4 outputs formed the basis of the concepts the project team further developed, prioritized, and carried forward. We synthesized seven distinct concepts for evaluating DHIs with clear propositions to support prioritization by the project team and working group (Table 4; Multimedia Appendix 3, containing the image of project team member synthesizing stages 1-4 outputs into the seven distinct concepts for evaluating DHIs). How each concept was linked to meet user needs via three key interconnected *scenarios of use* was visualized (Figure 3).

The concepts were scored and prioritized according to the following criteria: meeting user needs, delivering on project objectives, and feasibility for implementation within the PHE. The highest scoring concepts were taken forward into the alpha phase of the project.

Table 4. Final seven concepts for evaluating digital health interventions and underlying Jobs To Be Done.

Concept	Concept proposition	Example underlying JTBD[a]	Sample user quote[b]
Evaluation thinking	• When doing a digital health project at PHE[c], evaluation thinking, skills, and tools should be integrated into the project from the start. • Evaluation needs to be a central part of the design process and iterative delivery of any PHE service or product.	• As a Director of Public Health • When commissioning services • I need evaluation to be aligned closely with service delivery • So that it is formative and not a separate piece of work	"You have to get them at the beginning. And it's really hard when you don't get them at the beginning, 'cause then you've got to try and do a retrospective evaluation. You never have the right data. It's never as good of quality."
Evaluation canvas	• The evaluation canvas is PHE's validated and accepted portfolio of metrics, tools, and methods for evaluating DHIs[d]. • This canvas is the first step to creating a knowledge base on the effectiveness of digital health in meeting health outcomes and will support decisions on policy, practice, and research.	• As an evaluator • When designing or setting up an evaluation • I need access to a range of evaluation tools and paradigms • So that I can choose the one fit for the problem	"You have a set of tools and paradigms that are going to be suitable for different types of problems. When you're investigating different types of problems, for particular types of stakeholders, which particular types of objectives, or even particular types of technology, or particular stages of innovation, you have to be careful about finding the right thing and not trying to squeeze a round peg into a square hole."
Contract assistant	• A way for PHE teams to create strategic relationships with suppliers, supported by forward-thinking contracts. • A core requirement when working for PHE will be embedding evaluation into the design and development of the DHI and allowing for flexibility in contracted deliverables as the DHI progresses.	• As a public health professional • When setting up evaluations • I need to get the data-sharing agreements in place as simply as possible • So that I can collect, collate, and evaluate the data	"It's setting a level of expectation...as part of their bids, they need to articulate how they're going to capture data and how they're going to evaluate. And we needed some commonality so that if we have three sites, we can compare across the three sites... aggregate all our data together."
Testing toolkit	• Simple tools and methods to enable PHE teams delivering a DHI to test all aspects of the service or product throughout the development journey. • The toolkit could include a guide for face-to-face research, approaches to and models for planning and prototyping, functionality for randomizing users, and digital solutions for validating propositions in the market and/or against existing services.	• As a digital professional • When deciding what to design and how to design it • I need to validate service propositions by testing assumptions • So that I can be confident in committing resources to develop them	"It's fundamental to my role...I'm the one who manages [metric company name] and...also the research we conduct with users to define and validate services prior to committing resources developing them. But also to maximize them later, so to use informal or design-led evaluation means to validate, to research, to prove assumptions prior to designing things."
Development history	• A tool for PHE to record the full development history of their DHI project. This will support decision-making, facilitate process evaluation, and enable successful assurance in line with the Government Digital Service pipeline approach. • This record would include user needs, decisions and rationale, testing methods and results, information about the underlying technology, and stakeholder mapping.	• As an academic evaluator • When doing evaluation • I need an in-depth understanding of the intervention and the pathway of action • So that I can properly evaluate it	"You need a really good understanding of how and why these things are working and what you're changing...which is often much more complex than people gather."

Concept	Concept proposition	Example underlying JTBD[a]	Sample user quote[b]
Data hub	• PHE's data hub provides access to high-quality, accessible, and understandable public health data for providers, academia, and the wider market. • Similar to Transport for London's Open Data, PHE will encourage developers to use PHE's data feed to solve public health problems in innovative ways. • By setting the standard for the metrics in the data hub and promoting collaboration, this data hub may, in the future, allow for longitudinal analysis of DHIs.	• As an evaluator • When evaluating a health product or service • I need access to clean, accessible and linked data from across the health system • So that I can do my evaluation	"A lot of their datasets are in silos, they're maybe not using data the most effectively."
Publish health outcomes	• A place for stakeholders to publish and share how DHIs have met or not met desired health outcomes. This promotes collaboration among teams working in similar areas and enables sharing of best practices. • Collaboration among PHE, public health professionals, academia, and suppliers working in digital public health aligns with the Government's Industrial Strategy and the NHS[e] Innovation Strategy.	• As a digital professional • When working with PHE • I need to understand clearly what clinical data is required • So that I can be clear what success or impact looks like and can provide clinical impact	"We moved away from a lot of input and process measures, and balanced it with output and outcome measures, so we know now what impact they're having with individuals that they're working on, particularly their mental well-being, the escalation of demand into adult social care, and how they're embedding into a neighbourhood context more than they used to before."

[a]JTBD: Jobs To Be Done.

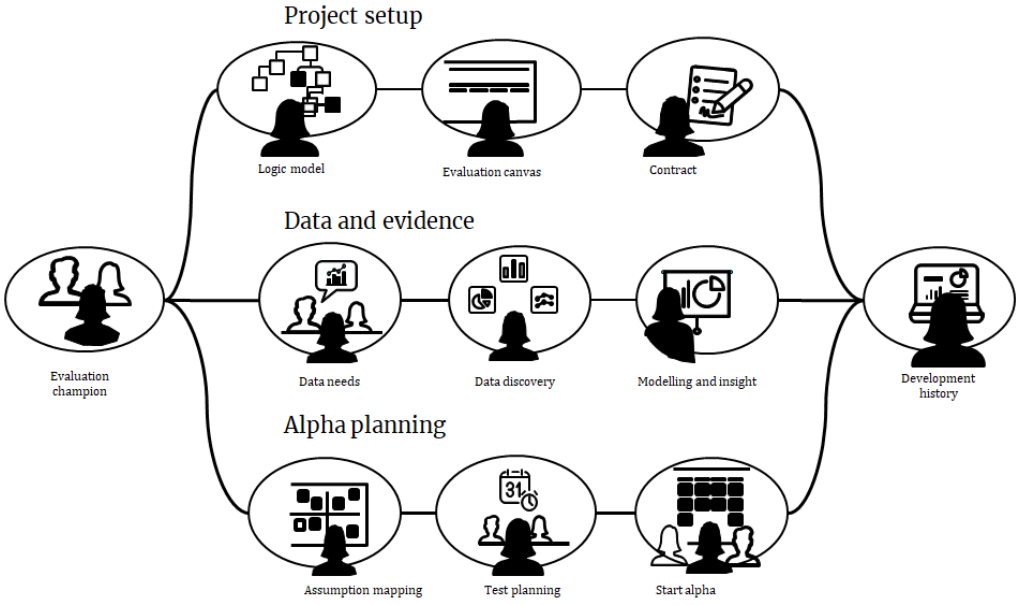
[b]Data from initial semistructured interviews with representatives of target user groups (academic, digital, and public health).

[c]PHE: Public Health England.

[d]DHI: digital health intervention.

[e]NHS: National Health Service.

Figure 3. Scenario of use illustrating how the final concepts fit together to facilitate setting up digital health intervention evaluations from conception (project setup) through collection of data and evidence to planning testing of minimum viable products (evaluation planning).



Corresponding Results

See Table 4 for the final seven concepts created to evaluate the DHIs. The concepts integrated perspectives and approaches from all three user groups and were combined into a *scenario of use* for PHE stakeholders. Figure 3 illustrates how the concepts fit together into a proposed Evaluation Service for helping teams evaluate their public health DHIs. Three concepts

were prioritized to take forward to the next phase of work: evaluation thinking, evaluation canvas, and testing toolkit. The evaluation thinking concept was taken forward in a separate work stream and is, therefore, not discussed in this paper. Textbox 2 illustrates how one of the final concepts, evaluation canvas, was derived and iterated through the methodology outlined in stages 1 to 5. This is an iterative process with multiple feedback loops rather than a linear one. See Tables S1 and S2 in Multimedia Appendix 1 for how the prioritized concepts, testing toolkit, and evaluating thinking were derived.

Textbox 2. Sample Jobs To Be Done life cycle from primary research (stages 1 and 2) to co-design workshops (stages 3 and 4) to refined concepts prioritized by Public Health England for further development (stage 5) in evaluating digital health interventions.

Evaluation Stage

- Stage 2: setup, information, guidance, and tools

Theme

- Evaluation methods

User Quote

- "You have a set of tools and paradigms that are going to be suitable for different types of problems. When you're investigating different types of problems, for particular types of stakeholders, which particular types of objectives, or even particular types of technology, or particular stages of innovation, you have to be careful about finding the right thing and not trying to squeeze a round peg into a square hole."

Jobs To Be Done

- As an evaluator
- When I design/set up an evaluation
- I need access to a range of evaluation tools and paradigms
- So that I can choose the one fit for the problem

Raw Concept From the Co-design Workshop 2

- Parallel playbook:

  a series of experiments to validate digital health intervention propositions and iterate these to validate impact

Refined Concept for Further Development

- Evaluation canvas:

  the evaluation canvas is Public Health England's validated and accepted portfolio of metrics, tools and methods for evaluating digital health interventions. This canvas is the first step to creating a knowledge base on the effectiveness of digital health in meeting health outcomes and will support decisions on policy, practice, and research.

Assumption Mapping and Experiment Design

Methodology

Assumption mapping and experiment design were performed to understand what needed to be true for each of the prioritized concepts to work in practice within PHE (project objective 5). This means, for example, that end users find the concept understandable, usable, desirable, and feasible within their workflow (project objectives 2 and 3).

For each concept, the project team identified assumptions and articulated a hypothesis in the form of "We believe [insert user] will [action]." This activity helped identify the circumstances in which a hypothesis would be considered true or false and therefore whether the assumptions were correct or erroneous. Experiments were conducted to test each hypothesis, including design artifacts for stimulating feedback. These experiments were a combination of thought experiments and practical tasks.

The outputs (prioritized concepts) from stages 1 to 5 and the subsequently articulated assumptions, hypotheses, and potential experiments (stage 6) formed the input for the paper prototypes (stage 7), which made the experiments tangible for the participants and project team.

Corresponding Results

The underlying assumptions and associated hypotheses were mapped for two concepts, evaluation canvas and testing toolkit, and for the proposed Evaluation Service overall. Ten assumptions were mapped with ten hypotheses (Table S3, Multimedia Appendix 1).

Paper Prototypes

Methodology

Testing sessions were held with both individual users and teams (Table 1) running a live DHI. These sessions differed from the semistructured interviews as they tested whole or partial solutions for the service using paper prototypes (project objective 5). These sessions were used to understand (1) all of the requisite steps, (2) the depth of each individual step, and (3) how the steps should be ordered and interrelated. They enabled the project team to iteratively test their understanding and clarify

the evaluation approach. For example, to build a project team's high-level understanding of the potential service (project objective 5), paper cards representing the stages of the emerging Evaluation Service journey were created (see Multimedia Appendix 3, item 2A showing the image of paper cards representing the stages of the emerging Evaluation Service journey used in stage 7, paper prototyping). Teams put these cards in the order that made the most sense to them and added any missing steps (see Multimedia Appendix 3, showing the image of paper cards representing the stages of the emerging Evaluation Service placed in order by user research participant in stage 7, paper prototyping).

Interactive design artifacts were also used to learn more about the individual steps of the Evaluation Service journey. This included testing logic models (conceptual models representing how a product is expected to work) and service blueprints [28] (diagrammatic representations of how the front- and backstage of a service function and dependencies) to learn how DHI teams align on the desired health outcomes and how they assess whether their DHI meets user needs.

The project team turned each of the aforementioned steps into interactive workshop activities to participate in DHI teams. First, they printed the logic model template and put it on the wall, with descriptions for each step (see Multimedia Appendix 3 showing the image of interactive workshop testing the logic model with a DHI team in Stage 7, paper prototyping). The DHI team then went through each step together. The project team observed the DHI team's conversations and interactions, noting whether the activity made sense to them, the questions they asked, and the group discussion.

The project team conducted a similar interactive workshop with the service blueprint (see Multimedia Appendix 3 showing the image of interactive workshop testing the service blueprint with a DHI team in stage 7, paper prototyping). In addition to describing the user's needs and actions, the service blueprint can include layers reflecting the service touchpoints and features, as well as the organizational delivery point of view, such as the people, processes, policies, and systems [29] required to bring each stage of the user experience to life.

The service experience tested had both analog (eg, in-person workshops) and digital components (eg, a digital platform). The proposed digital platform, which includes evaluation guidance and printable workshop materials, is referred to as the Evaluation Service. The Evaluation Service idea was tested with users using wireframes to represent the digital experience of a landing page and web pages for each step of the Evaluation Service. This enabled the project team to learn what users expect from a digital platform, as well as their feedback on content and design. The project team turned the outputs of stage 7 into a prototype of a full Evaluation Service, which was the input for stage 8 (proof of concept).

Corresponding Results

Seven partial service experience paper prototype sessions were conducted with a range of participants (Table 1). The project team collated insights from each session, presented below, which corresponded to the concepts and hypotheses tested (Table S3, Multimedia Appendix 1).

The outputs of the paper card sorting exercise resulted in ordered steps for evaluating a DHI: building an interdisciplinary team, creating a conceptual model (logic model), creating a detailed service blueprint, selecting relevant indicators against health outcomes, selecting relevant evaluation methods (testing toolkit), and learning how to use selected methods and indicators to evaluate a DHI (testing toolkit and evaluation canvas).

By testing different paper versions of conceptual models (logic models and/or theory of change), the project team learned that the logic model was the simplest tool. Participants expressed the benefits of a logic model: looking at the bigger picture, choosing KPIs, prioritizing important outcomes to aid decision-making, capturing and updating project elements such as objectives, explaining how a DHI worked, and cocreating DHI project value with stakeholders. These sessions also revealed that the more a DHI team presents the bigger picture of what they are trying to achieve, the better the resultant understanding of the team and stakeholders.

These testing sessions demonstrated that participants focused more on measuring the uptake of a DHI than the impact on health outcomes. Teams did not know about national resources, such as PHE's Public Health Profiles (Fingertips), and so did not link their DHI to nationally recognized health measures. The project team observed that participants would benefit from signposting and guidance in the proposed Evaluation Service to link nationally recognized and validated measures.

Participants preferred the printed workshop format of the logic model rather than a web-based version, as it was easier to collaborate. Participants requested more accessible language and the use of arrows to indicate causality. Participants liked referring to previous or example logic models to see whether their version made sense by comparison. It was suggested that example logic models and guidance should be part of the proposed Evaluation Service offering.

The result of testing the evaluation canvas was breaking it down into its constituent parts and merging it into the Evaluation Service. For example, the program elements were absorbed into the context section of a logic model.

From testing paper and digital versions of the service blueprint, the project team learned that participants responded well to the dimension of time it introduced and the ability to reflect on user experience throughout time as a result of a DHI (not) meeting a user's needs. By mapping a DHI to its place or impact on a user's journey (via the service blueprint), participants articulated that they could see the gaps in their understanding of user needs and the features of their DHI. Adding the backstage detail (ie, organizational elements: people, processes, policies, and systems) to the service blueprint also gave participants a better understanding of what their DHI did well and where it could be improved.

Overall, the paper prototype testing revealed that the proposed Evaluation Service should contain information about (refer to Table S3, Multimedia Appendix 1 for details on each hypothesis) the following aspects:

- How to determine the aim of a DHI evaluation with health and service outcomes as core elements of building an evaluation strategy (hypotheses G1, G3, T1, C1, C2, and C3).
- Set up an evaluation at the beginning of DHI development (hypotheses G2.1, G2.2, and T2).
- Process and steps for DHI evaluation (hypothesis T2).
- Constant revision and integration of evaluation strategy into DHI development (hypotheses T1, T2, and C1).
- Who is involved in DHI evaluation (hypotheses T3, C1, and C3).
- Different types of evaluation (impact, process, and health economics; hypothesis C1).
- Selection of evaluation methods and tools (hypotheses T1, T3, C1, and C2).
- Required capabilities and resources (hypotheses G2.1, T1, T3, and C3).
- Barriers that may be encountered during an evaluation (hypotheses C1, C2, and T1).
- External policy and stakeholder factors influencing an evaluation (hypotheses G1).
- Creating a common language and understanding among different disciplines involved in DHI evaluation (hypotheses G3, T1, T3, C1, and C3).

Proof of Concept: Prototype of the Full-Service Experience

Methodology

Stages 1 to 7 culminated in the project team conducting a proof-of-concept test in which a digital team followed the proposed Evaluation Service to build an evaluation approach for their DHI (project objective 5). The proof of concept was a series of workshops run with PHE's Couch-to-5K app in-house digital team (see Multimedia Appendix 3 showing the illustration of stage 8 proof of concept: prototype of the full-service experience conducted with PHE's Couch-to-5K app in-house team showing the paper prototyping interactive workshops and associated digital materials). For example, the Couch-to-5K team used the logic model workshop template to define the problem they were addressing and clarify the desired health and service outcomes. They used a service blueprint template to describe users' needs, experiences, and actions concerning service touchpoints and features, organizational and operational capabilities, and actions, as well as the desired health outcomes.

In addition to testing these sequential design artifacts, an evaluation facilitator role was tested in each workshop with the Couch-to-5K team. The Couch-to-5K team was provided with relevant materials and a brief introduction to the workshop aim and then invited to move through the activity as independently as possible. The Couch-to-5K team was observed to see what made sense to them and what they understood correctly. When the Couch-to-5K team needed help, the evaluation facilitator provided guidance; this helped identify when the guidance or materials were unclear and when a DHI team would need and value facilitator support.

Corresponding Results

The logic model workshop brought the Couch-to-5K team together around a shared view of the health outcomes they were trying to deliver for Couch-to-5K app users. By mapping user needs to the service blueprint and aligning them with service features, the Couch-to-5K team obtained a shared view of the user needs and their relation to the desired health outcomes. The Couch-to-5K team identified unmet basic needs (eg, incidents) and areas for improvement in the evaluation journey. The Couch-to-5K digital team was alerted to different evaluation methods and tools relevant to their product and its context (including maturity, budget, time constraints, and risk profile). Spending time reflecting on KPIs highlighted to the Couch-to-5K team additional pertinent indicators worth beyond what they already collected (eg, KPIs related to the broader physical activity market). The Evaluation Service experience subsequently informed the design of the Couch-to-5K behavioral database (incorporating newly identified KPIs aligned to desired health outcomes).

Discussion

Principal Findings

PHE collaboratively developed an Evaluation Service with digital developers, academics, and public health professionals involved in the commissioning, design and development, and evaluation of DHIs. Following an iterative UCD approach, which was novel for PHE, the project team worked across and synthesized each discipline's areas of importance in relation to evaluating DHIs. An in-depth collective understanding of how biomedical and digital evaluation activities map to and can be used during the design and development of a DHI resulted. Testing with PHE's Couch-to-5K app in-house digital team demonstrated the desirability, viability, and feasibility of the Evaluation Service and led to further investment by PHE.

The primary strength of the work was in PHE's organizational openness to UCD and service design methodologies that were, at the time, not commonly used by the national body. Although this sparked some cultural tensions, doing so led to knowledge sharing between the external supplier and internal PHE project team and stakeholders, facilitated the synthesis of biomedical and digital evaluation approaches, and grew PHE's understanding of the benefits of user-centered approaches for designing and evaluating DHIs. Through extensive user and stakeholder engagement throughout the project, we demonstrated to PHE senior leaders and other health system organizations that *design is your friend* and there is a place for design disciplines in public health.

As PHE adapted to the UCD project approach, the project team unearthed tensions among participants such as discomfort with the *fast pace* of the 2-week sprints, divergent expectations on what was considered *good enough* evidence for justifying a user need or concept, and hesitance to try new methods and work in the open. Although some of the PHE stakeholders were initially uncomfortable with the pace and *roughness* of concepts and prototypes, they appreciated that the quick, iterative UCD approach enabled more people to interact, provide feedback, and contribute to the development of the Evaluation Service.

The Evaluation Service was thereby informed by, and socialized with, a wider range of professionals involved in DHI development and evaluation. PHE's senior stakeholders also acquired substantial evidence of the user's need for the Evaluation Service before requesting further investment from the organization.

Limitations

We identified three key limitations of this study. First, the sample size in qualitative design research is often small compared with traditional academic qualitative research. This typical design research practice is based on the finding [30,31] that speaking with a representative sample of 5 to 10 people at a time is sufficient to uncover common challenges, understand underlying causes, and inform decisions. In design research and digital development, as outlined by the Government Digital Service [32], the limitation of small sample size is usually overcome by conducting iterative research and increasing the number of users testing and feeding back on the service as it progresses through the initial phases (discovery and alpha) to later phases (beta and live). In this way, the team's understanding continues to grow as the sample size becomes larger and more diverse with time. The sample size of participants was increased in the later stages of this project; however, this is not within the scope of this paper. Second, PHE's organizational remit and limited financial and human resources resulted in the underutilization of concepts, with only three of the final seven concepts (Table 4) prioritized for further development by PHE. As shown in the *scenario of use* (Figure 3), the seven concepts complemented each other, and, through the prioritization exercise, the potential of the scenario was not fully tested. Third, substantial time and input were drawn down from members of the working group and relevant PHE senior leaders to upskill service design project team members in evaluating DHIs and share tacit knowledge of evaluating DHIs. This was a strength of the project, as the project team was able to bring a fresh perspective to drawing insight from experts and users as they designed the Evaluation Service; however, the personnel cost must be recognized.

Conclusions

The potential of DHIs to combine the reach of large-scale population initiatives with the efficacy of individual treatments is yet to be fully realized. It will continue to be unrealized if how we evaluate and use evaluations to inform the iterative design and development of DHIs do not use the perspectives of both biomedical research and digital development. DHIs are an interdisciplinary endeavor, bringing together clinical or population health interventions, digital product development, product and service design, and communication and health promotion. Hence, the evaluation of DHIs is best informed by interdisciplinary approaches to evaluation to understand both the efficacy of a DHI and its usability and desirability, with measures of success that reflect the different stakeholders involved in the commissioning, design, and development of a DHI and its end users [11].

We have outlined the work undertaken by PHE, a national body, to ensure that DHIs contribute to the improvement of population health and that taxpayers receive the most value (improvement to public health) from investment in digital health. The project is a worked example of using UCD, particularly service design, methods to iteratively understand, synthesize, and embed needs and evaluation approaches of both biomedical researchers and digital product developers. Public health's traditional approach is complemented by the UCD approach, which in turn is made safer and more robust through its interaction with public health. The resultant Evaluation Service enables digital developers or nonacademics to apply evaluation thinking and techniques during the design, development, and implementation of their DHI. By doing so, it demystifies evaluation, traditionally the realm of academia, and harnesses people's motivations to ensure that their DHI is as good as it can be and improves end users' health and well-being. PHE has subsequently built a digital version of the Evaluation Service (named *Evaluating Digital Health Products*), which is openly available on the web [33].

Acknowledgments

The authors acknowledge and thank the PHE team who delivered the discovery and alpha phases of *Evaluating Digital Health Products* (in alphabetical order): Iain Cooper, Simon Dixon, Charlotte Fountaine, Felix Greaves, Bernie Hannigan, Kassandra Karpathakis, Vicki Litherland, Ros-Mari Mitova, Claire Rackstraw, and Joseph Reynolds. The authors acknowledge and thank the Livework Studio Ltd team who delivered the discovery and alpha phases (in alphabetical order): Anton Drachuk, Lilli Graf, Claire Henderson, Gene Libow, Nick Manson, Ben Reason, and Paul Sims. The authors acknowledge and thank the project's working group (in alphabetical order, affiliations at the time the project was delivered): Jamie Brown, UCL; Tim Chadborn, PHE; Darren Curry, NHS Business Service Authority; Kenny Deighton, PHE; Pall Johnson, NICE; Indra Joshi, NHSE/I; Scott Mahony, PHE; Elizabeth Murray, UCL; Felix Naughton, University of East Anglia; Claudia Pagliari, University of Edinburgh; Colin Pattinson, Department of Health and Social Care; John Powell, NICE; Susannah Robinson, World Health Organization; Mark Salmon, NICE; Tero Väänänen, NHS Digital; and Matthew Walmsley, PHE. The authors acknowledge the financial support received from PHE.

Authors' Contributions

KK, SD, and FG were involved in the conception of the project and funding acquisition. KK and GL led the methodology, data collection and curation, formal analysis, and project administration. KK, GL, SD, FG, and EM were involved in the validation of the project with SD and FG, which also contributed to the formal analysis. FG and SD provided the senior institutional

leadership. EM provided senior academic leadership. KK, GL, and EM wrote the first draft, and all authors reviewed and edited the subsequent drafts. HP and KK edited the final drafts. HP is the corresponding author. KK is the guarantor. All authors read and approved the final version of the manuscript.

Conflicts of Interest

KK was employed by PHE and at the time of writing was employed by NHSX. GL was contracted by PHE as an employee of Livework Studio Ltd to complete this project. HWWP was employed by PHE in an honorary academic contract. He received funding support from NHSE/I, and has done or is doing consultancy for Babylon Health and IpsosMORI and has had a recent project grant with Six to Start. He has PhD students funded by Patients Know Best and by AstraZeneca. SD was employed by PHE and is currently employed by NHSX. FG was employed by PHE and is currently employed by NICE. EM has no conflicts of interest to declare.

Multimedia Appendix 1
Additional figures and tables with examples of artefacts used in the development of the service.
[DOCX File , 122 KB - jmir_v23i9e28356_app1.docx]

Multimedia Appendix 2
Discussion guide for semistructured interviews.
[DOCX File , 31 KB - jmir_v23i9e28356_app2.docx]

Multimedia Appendix 3
Images taken during project delivery.
[ZIP File (Zip Archive), 9909 KB - jmir_v23i9e28356_app3.zip]

References

1. Abrams DB, Orleans CT, Niaura RS, Goldstein MG, Prochaska JO, Velicer W. Integrating individual and public health perspectives for treatment of tobacco dependence under managed health care: a combined stepped-care and matching model. Ann Behav Med 1996 Dec;18(4):290-304 [FREE Full text] [doi: 10.1007/bf02895291]
2. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
3. Scott-Sheldon LA, Lantini R, Jennings EG, Thind H, Rosen RK, Salmoirago-Blotcher E, et al. Text messaging-based interventions for smoking cessation: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016;4(2):e49 [FREE Full text] [doi: 10.2196/mhealth.5436] [Medline: 27207211]
4. Griffiths SE, Parsons J, Naughton F, Fulton EA, Tombor I, Brown KE. Are digital interventions for smoking cessation in pregnancy effective? A systematic review and meta-analysis. Health Psychol Rev 2018 Dec 28;12(4):333-356. [doi: 10.1080/17437199.2018.1488602] [Medline: 29912621]
5. Meredith SE, Alessi SM, Petry NM. Smartphone applications to reduce alcohol consumption and help patients with alcohol use disorder: a state-of-the-art review. Adv Health Care Technol 2015;1:47-54 [FREE Full text] [doi: 10.2147/AHCT.S65791] [Medline: 27478863]
6. Ryan K, Dockray S, Linehan C. A systematic review of tailored eHealth interventions for weight loss. Digit Health 2019;5:2055207619826685 [FREE Full text] [doi: 10.1177/2055207619826685] [Medline: 30783535]
7. Stockwell S, Schofield P, Fisher A, Firth J, Jackson SE, Stubbs B, et al. Digital behavior change interventions to promote physical activity and/or reduce sedentary behavior in older adults: a systematic review and meta-analysis. Exp Gerontol 2019 Jun;120:68-87. [doi: 10.1016/j.exger.2019.02.020] [Medline: 30836130]
8. Rose T, Barker M, Maria Jacob C, Morrison L, Lawrence W, Strömmer S, et al. A systematic review of digital interventions for improving the diet and physical activity behaviors of adolescents. J Adolesc Health 2017 Dec;61(6):669-677 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.05.024] [Medline: 28822682]
9. Roberts AL, Fisher A, Smith L, Heinrich M, Potts HW. Digital health behaviour change interventions targeting physical activity and diet in cancer survivors: a systematic review and meta-analysis. J Cancer Surviv 2017 Dec;11(6):704-719 [FREE Full text] [doi: 10.1007/s11764-017-0632-1] [Medline: 28779220]
10. Guse K, Levine D, Martins S, Lira A, Gaarde J, Westmorland W, et al. Interventions using new digital media to improve adolescent sexual health: a systematic review. J Adolesc Health 2012 Dec;51(6):535-543. [doi: 10.1016/j.jadohealth.2012.03.014] [Medline: 23174462]
11. Blandford A, Gibbs J, Newhouse N, Perski O, Singh A, Murray E. Seven lessons for interdisciplinary research on interactive digital health interventions. Digit Health 2018;4:2055207618770325 [FREE Full text] [doi: 10.1177/2055207618770325] [Medline: 29942629]

12. Bondaronek P, Slee A, Hamilton FL, Murray E. Relationship between popularity and the likely efficacy: an observational study based on a random selection on top-ranked physical activity apps. BMJ Open 2019 Nov 14;9(11):e027536 [FREE Full text] [doi: 10.1136/bmjopen-2018-027536] [Medline: 31727641]
13. Kling R. The organizational context of user-centered software designs. MIS Q 1977 Dec;1(4):41. [doi: 10.2307/249021]
14. Crane D, Garnett C, Brown J, West R, Michie S. Behavior change techniques in popular alcohol reduction apps: content analysis. J Med Internet Res 2015 May 14;17(5):e118 [FREE Full text] [doi: 10.2196/jmir.4060] [Medline: 25977135]
15. Preece J, Sharp H, Rogers Y. Interaction Design: Beyond Human-Computer Interaction. Hoboken, New Jersey, United States: Wiley; 2015:1-584.
16. Blandford A, Furniss D, Makri S. Qualitative HCI Research: Going behind the scenes. Synth Lect Hum Cent Inform 2016 Apr 07;9(1):1-115. [doi: 10.2200/S00706ED1V01Y201602HCI034]
17. Hall CL, Sanderson C, Brown BJ, Andrén P, Bennett S, Chamberlain LR, et al. Opportunities and challenges of delivering digital clinical trials: lessons learned from a randomised controlled trial of an online behavioural intervention for children and young people. Trials 2020 Dec 09;21(1):1011 [FREE Full text] [doi: 10.1186/s13063-020-04902-1] [Medline: 33298127]
18. Robinson H, Appelbe D, Dodd S, Flowers S, Johnson S, Jones SH, et al. Methodological challenges in web-based trials: update and insights from the relatives education and coping toolkit trial. JMIR Ment Health 2020 Jul 17;7(7):e15878 [FREE Full text] [doi: 10.2196/15878] [Medline: 32497018]
19. Polaine A, Løvlie L, Reason B. Service Design: From Insight to Inspiration. New York: Rosenfeld Media; 2013:1-216.
20. Stickdorn M, Hormess M, Lawrence A, Schneider J. This Is Service Design Doing : Applying Service Design Thinking in the Real World. Sebastopol, CA: O'Reilly Media; 2017.
21. About service design and the SDN. Service Design Network. URL: https://www.service-design-network.org/about-service-design [accessed 2021-05-30]
22. Benjamin K, Potts HW. Digital transformation in government: lessons for digital health? Digit Health 2018 Feb 27;4:2055207618759168 [FREE Full text] [doi: 10.1177/2055207618759168] [Medline: 29942624]
23. GOV.UK Service Manual. Government Digital Service. URL: https://www.gov.uk/service-manual [accessed 2021-05-30]
24. GOV.UK Service Manual: Service standard. Government Digital Service. URL: https://www.gov.uk/service-manual/service-standard [accessed 2021-05-30]
25. Jobs to be done: a case study in the NHS. NHS Digital. 2017. URL: https://digital.nhs.uk/blog/transformation-blog/2017/jobs-to-be-done-a-case-study-in-the-nhs [accessed 2021-05-30]
26. Ivey-Williams K. Design in government: Why we use user journey maps in government. GOV.UK. 2016. URL: https://designnotes.blog.gov.uk/2016/03/30/why-we-use-user-journey-maps-in-government/ [accessed 2020-12-29]
27. Schuurman D, De Moor K, De Marez L, Evens T. Investigating user typologies and their relevance within a living lab-research approach for ICT-innovation. In: Proceedings of the 43rd Hawaii International Conference on System Sciences. 2010 Presented at: 43rd Hawaii International Conference on System Sciences; Jan. 5-8, 2010; Honolulu, HI, USA p. 1-10. [doi: 10.1109/hicss.2010.242]
28. Shostack G. Designing services that deliver. Harv Bus Rev 1984;62(1):133-139 [FREE Full text]
29. Reason B, Løvlie L, Flu MB. Service Design for Business: A Practical Guide to Optimizing the Customer Experience. Hoboken, New Jersey, United States: Wiley; 2016:1-208.
30. Turner C, Lewis J, Nielsen J. Determining usability test sample size. In: International Encyclopedia of Ergonomics and Human Factors. Vol 3. Boca Raton, Florida, United States: CRC Press; 2006.
31. Nielsen J, Landauer T. A mathematical model of the finding of usability problems. In: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. 1993 Presented at: INTERCHI93: Conference on Human Factors in Computing; April 24 - 29, 1993; Amsterdam The Netherlands p. 206-213. [doi: 10.1145/169059.169166]
32. GOV.UK Service Manual: Learning about users and their needs. Government Digital Service. URL: https://www.gov.uk/service-manual/user-research/start-by-learning-user-needs [accessed 2021-05-30]
33. Potts H, Death F, Bondaronek P, Gomes M, Raftery J. Evaluating digital health products. Public Health England. 2020. URL: https://www.gov.uk/government/collections/evaluating-digital-health-products [accessed 2021-05-30]

Abbreviations

DHI: digital health intervention
JTBD: Jobs To Be Done
KPI: key performance indicator
NHSE/I: National Health Service England and Improvement
NICE: National Institute of Health and Care Excellence
PHE: Public Health England
UCD: user-centered design
UCL: University College London

Edited by M Focsa; submitted 02.03.21; peer-reviewed by S Stockwell, P Williams; comments to author 09.04.21; revised version received 31.05.21; accepted 19.06.21; published 08.09.21.

Please cite as:
Karpathakis K, Libow G, Potts HWW, Dixon S, Greaves F, Murray E
An Evaluation Service for Digital Public Health Interventions: User-Centered Design Approach
J Med Internet Res 2021;23(9):e28356
URL: https://www.jmir.org/2021/9/e28356
doi:10.2196/28356
PMID:34494965

Original Paper

Virtual Reality Tour to Reduce Perioperative Anxiety in an Operating Setting Before Anesthesia: Randomized Clinical Trial

Lina Vogt[1,2], Dr med; Martin Klasen[1], Dr rer medic, Dipl-Psych; Rolf Rossaint[2], Univ-Prof Dr med; Ute Goeretz[1]; Peter Ebus[3], Dr ir; Sasa Sopka[1,2], PD Dr med

[1]AIXTRA Competence Center for Training and Patient Safety, Medical Faculty, RWTH Aachen University, Aachen, Germany

[2]Department of Anesthesiology, Medical Faculty, RWTH Aachen University Hospital, Aachen, Germany

[3]Faculty of Educational Science, Open University, Heerlen, Netherlands

Corresponding Author:
Lina Vogt, Dr med
AIXTRA Competence Center for Training and Patient Safety
Medical Faculty
RWTH Aachen University
Pauwelsstraße 30
Aachen, 52074
Germany
Phone: 49 2418089974
Email: lvogt@ukaachen.de

Abstract

Background: Perioperative anxiety is a major burden to patients undergoing surgeries with general anesthesia.

Objective: This study investigated whether a virtual operating room tour (VORT) before surgery can be used to ameliorate perioperative anxiety.

Methods: We employed a randomized parallel-group design with 2 study arms to compare VORT to the standard operation preparation procedure. The study included 84 patients. A validated inventory (state-trait operation anxiety-state) was used to assess perioperative state anxiety before (T1) and after (T2) surgery. In addition, trait operation anxiety was evaluated with an additional validated inventory (state-trait operation anxiety-trait). Moreover, user ratings on the usefulness of VORT were assessed with an evaluation questionnaire. Study arms were compared for perioperative state anxiety with two-tailed independent samples *t* tests. Subjective ratings were correlated with STOA-Trait values to investigate possible associations between perioperative anxiety with perceived usefulness.

Results: There were no significant differences in perioperative state anxiety between VORT and standard operation preparation procedures before and after the surgery. Nonetheless, patients' ratings of VORT overall were positive. The tour was perceived as useful and, therefore, showed acceptance for VR use. These ratings were unrelated to the degree of perioperative anxiety.

Conclusions: The subjective benefit of VORT could not be explained by a reduction of perioperative anxiety. Instead, VORT appears to serve the need for information and reduce uncertainty. In addition, VORT is perceived as beneficial regardless of the age of the patients. Considering this effect and the manageable organizational and financial effort toward implementation, the general use of VORT can be recommended.

Trial Registration: ClinicalTrials.gov NCT04579354; https://clinicaltrials.gov/ct2/show/NCT04579354

(J Med Internet Res 2021;23(9):e28018) doi:10.2196/28018

KEYWORDS

anesthesia; anxiety; exposure; operating theater; patient empowerment; periperative trait anxiety; STOA; STAI; virtual reality; VR

Introduction

Fear and anxiety are very common in patients prior to surgery with anesthesia. This perioperative anxiety can lead to additional discomfort, increased pain sensation, and increased stress symptoms [1,2]. It is a significant burden for patients and a potential reason for decreased compliance.

Virtual stimulus exposure is successful in exposure therapy for the treatment of anxiety and tension [3]. Patients are not confronted with real anxiety-inducing stimuli but with virtual representations of them [3]. The effectiveness of virtual stimulus exposure is well documented, especially for specific phobias [4-8], where it is not inferior to in vivo exposition [4]. Therefore, virtual stimulus exposure could also be suitable for minimizing operation-associated fears.

However, much less is known about the application of virtual reality (VR) exposure in reducing situational fear in people without anxiety disorders. The potential of VR exposure to reduce anxiety before surgery with general anesthesia has recently been explored with respect to perioperative anxiety. Ryu et al [9] applied a virtual operating room tour for 69 children 1 hour before surgery with general anesthesia, leading to lower preoperative anxiety and stress levels and higher compliance. However, data on the issue is sparse and inconsistent, especially when evaluating adults.

This study aimed to evaluate the use of VR exposure via a virtual operating room tour (VORT) on perioperative anxiety in adult patients undergoing surgeries with general anesthesia. Previous research has shown that perioperative anxiety exists both as a state (ie, the acute emotional or cognitive reaction of a patient before or after the operation) and as a trait (ie, the patient's disposition to be more or less fearful about operations) [2]. We assumed that VORT reduced perioperative state anxiety in patients before and after surgery with general anesthesia. In addition to the exposure effect, information about anesthesia and treatment can be conveyed using VR at low cost and effort, thus strengthening individual health competence.

Methods

Ethics Approval

The study protocol was approved by the local Ethics Committee (EK 039/20). The study was designed and conducted in accordance with the latest version of the Declaration of Helsinki [10]. Moreover, the study was registered by clinicaltrials.gov on September 28, 2020 (NCT04579354).

Participants

All participants were patients of the Clinic for Anesthesiology, RWTH Aachen University Hospital. The inclusion criteria included age over 18 years, elective surgery with general anesthesia, and a good knowledge of German. Exclusion criteria included emergency surgery, thoracic surgery, neurosurgery, and tumor surgery, as these can be associated with increased anxiety or inability to provide consent. All participants signed an informed consent form before participating in the study.

Sample Size Planning

Sample size planning was based on a meta-analysis by Carl [4] on the effect of virtual exposure on anxiety. For comparisons of VR exposure with participants on a waiting list that receive treatment after the treatment groups, this study reported a mean effect size (Hedges's g) of 0.9, analogous to Cohen's *d*. Sample size planning using g multiplied by power for this effect size resulted in a sample size of 68 patients (34 per group) for a two-sided null hypothesis test (*t* test for independent samples), with a power of 0.95 and an α error probability of .05. To compensate for an expected dropout of 10%-15%, we targeted a total sample size of 80 patients (40 per group).

Study Design

The study employed a randomized parallel-group design to compare VORT to the standard operation preparation procedure (STOPP). We employed the state and trait anxiety scales from the state-trait operation anxiety inventory (STOA) [2] as outcome variables. STOA is a validated inventory with good psychometric properties which assess operation-related trait and state anxiety on separate scales. Trait anxiety (STOA-T, 20 items) is the relatively stable disposition of a person to be anxious about surgeries, whereas state anxiety (STOA-S, 10 items) assesses acute fear reactions in situations shortly before or after surgery on 2 dimensions (cognitive anxiety and affective anxiety, 5 items each). Measuring a relatively stable trait, STOA-T was rated only before the surgery, whereas STOA-S was assessed before (T1) and after the surgery (T2). As an additional outcome, we employed an in-house designed questionnaire to assess the participants' perception and evaluation of VORT after surgery. The questionnaire assessed different statements about VORT with 8 items on a 5-point Likert scale from 1 (do not agree) to 5 (strongly agree). Moreover, in 1 item, patients were asked to give VORT a school grade (according to the German grade system) between 1 (excellent) and 6 (very poor).

Participants were randomly assigned to one of the study arms (VORT vs STOPP) by a person not involved in data acquisition using the following procedure: for all patients (N=80), we created opaque assignment envelopes containing a folded paper with a study arm assignment (VORT or STOPP), with a total of 40 participants in each arm. The order of the envelopes was randomized before the beginning of the study. The assignment envelope for each patient was opened after these steps to avoid any bias in providing study information to the patient or the standardized explanatory interview. Thus, the assignment to the study arms was double-blind (to the study nurse or anesthetist and the patient) until after the standardized explanatory anesthesia interview. The procedural steps of the study can be described as follows:

Providing Study Information and Obtaining Informed Consent

Information was provided in written form, and the study nurse assisted the consent process by providing additional explanations and answered questions. Patients were informed about the study background, the study arms, and the content of the video. All patients received a paper copy of the information sheet. Informed consent was obtained in written form. This step was identical for all patients in both study arms (VORT and STOPP).

Standardized Explanatory Anaesthesia Interview

The interview consisted of a detailed explanation by the educating physician (anesthetist). A standardized information sheet was further used to explain the procedure and the anesthesia used, corresponding to the hospital's standard

operating procedure, and it was identical for all patients in both study arms (VORT and STOPP).

Unblinding of the Study Arm Assignment

Unblinding entailed opening the assignment envelope and according assignment of the patient to one of the study arms.

Arm 1: Virtual Operation Room Tour (VORT)

Patients in this study arm subsequently watched a virtual tour of the operation room using Oculus Go Standalone VR (product number 301-00105-01; Facebook). The tour began with the evening conversation between the patient and a nurse about the approximate time of the operation and the desired sleeping medication. The next scene showed the morning conversation about premedication and drug administration. Furthermore, the route through the hospital from the normal ward to the holding area was illustrated. In the holding area, aspects of the "safe-surgery saves lives" checklist, such as the operation side (left or right), were simulated in the conversation between the patient and a nurse. Finally, a scene from the operating room was shown in which the actual preparations for anesthesia (ie, breathing tube and intravenous access) up to the induction of anesthesia were simulated.

The total duration of VORT was 6 minutes 28 seconds. VR videos were recorded with an Insta360Pro 3D camera (Arashi Vision Inc), with 4k resolution and an internal microphone. The video can be found online [11]. VR videos were recorded from a third-person perspective with a static camera mounted on a tripod. Early in the study, we alternatively considered using a video from a first-person perspective. The decision against the latter was based on three observations: (1) 360° recordings led to strange camera images when shot from the perspective of a person lying in bed, (2) first-person perspective might evoke a strange feeling of alienation (eg, in women), given the fact that the male actor's voice was still audible, and (3) several pilot subjects reported simulation sickness when watching the first-person perspective video.

Arm 2: Standard Operation Preparation Procedure (STOPP)

Following the standard procedure of the hospital, patients in this study arm underwent no additional preparation.

Questionnaires T1 (Before Operation)

Assessment of STOA-T and STOA-S (T1). This step was identical for all patients in both study arms (VORT and STOPP).

Questionnaires T2 (After Operation)

The assessment of STOA-S (T2) for all patients was completed in both study arms within 48 hours after surgery. Simultaneously, patients in the VORT study arm completed the evaluation questionnaire.

Research Hypothesis

1. H1: VORT reduces perioperative state anxiety before the surgery with general anesthesia compared to the standard procedure.
2. H2: VORT reduces perioperative state anxiety after the surgery with general anesthesia compared to the standard procedure.

Moreover, concerning the application of VORT, we were interested in its overall patient evaluation. Also, we wanted to explore the possibility of perceived usefulness, especially in patients with higher trait anxiety, in other words, if the VORT rating was associated with perioperative trait anxiety. Finally, we wanted to evaluate whether the evaluation of VORT depended on the patient's age (eg, older patients' limited experience with VR). Therefore, we additionally investigated the following explorative research questions:

1. ER1: How is the use of VORT prior to surgery evaluated and accepted by patients?
2. ER2: Is there a possible relationship between the evaluation of VORT and perioperative trait anxiety?
3. ER3: Is there a possible relationship between the evaluation of VORT and patients' age?
4. ER4: Does VORT influence anesthesia-related STOA items?

Data Analysis

Data were analyzed with SPSS Statistics (version 25; IBM). STOA-S sum scores for the two dimensions (cognitive and affective anxiety) were compared between study arms with two-tailed independent samples *t* tests separately for T1 (H1) and T2 (H2). To quantify agreement to the Likert scale items in the evaluation questionnaire (ER1), we calculated the median of the responses. Furthermore, we recoded the responses to disagreement (values 1 and 2), neutral (value 3), and agreement (values 4 and 5) and descriptively quantified the percentage of agreement. To evaluate a possible relationship between trait operation anxiety and the evaluation of VORT (ER2), we calculated Pearson correlations between the items from the evaluation questionnaire and STOA-T values. Finally, we calculated correlations between the items from the evaluation questionnaire and the patients' age. A threshold of an α error probability of .05 (two-sided testing) was applied to all analyses.

Results

Overview

In total, 84 patients were included in the study. Dropouts (n=12; 14.3%) were in the expected range; however, since all dropouts occurred during the enrolment of the first 62 patients, we decided at this time point to create and randomize 20 additional assignment envelopes. The primary reason for dropouts was the early discharge of the patient. Other reasons included postponed or additional surgeries and postoperative complications. VORT-associated concerns or worries were not reported; therefore, VORT did not contribute to participation dropouts. One patient withdrew from study participation. Data from the remaining 72 patients (mean age 54.19, SD 15.94, with a range of 20 to 81 years), with 35 female and 37 male subjects, was analyzed (VORT: n=35 vs STOPP: n=37). Due to single missing data points, the numbers of included patients vary slightly for the different analyses. Results of the study are presented according to our hypotheses.

H1: VORT Reduces Perioperative State Anxiety Before the Surgery as Compared to the Standard Procedure

No significant differences in state anxiety emerged between VORT and STOPP before surgery. This was the case for both the affective (n=71; VORT: mean 10.49, SD 4.61 vs STOPP: mean 10.83, SD 4.09; t_{69}=–0.37; *P*=.74) and the cognitive dimension (n=67; VORT: mean 10.88, SD 4.32 vs STOPP: mean 10.70, SD 4.38; t_{65}=–0.18; *P*=.86).

H2: VORT Reduces Perioperative State Anxiety After the Surgery as Compared to the Standard Procedure

Similarly, no significant differences in state anxiety emerged between VORT and STOPP after the surgery. This was the case both for the affective (n=68; VORT: mean 8.53, SD 4.95 vs STOPP: mean 7.91, SD 3.26; t_{66}=0.61; *P*=.55) and the cognitive dimension (n=71; VORT: mean 9.11, SD 4.19 vs STOPP: mean 9.17, SD 4.00; t_{69}=–0.05; *P*=.96).

ER1: How is the Use of VORT Prior To Surgery Evaluated by Patients?

The median values of the responses and proportion agreement to the statements are listed in Table 1. Altogether, the patients' ratings of the VORT are positive. The results indicate that most patients perceived the VORT as helpful for preparation and can recommend it to other patients. A median school grade of 2 (corresponding to "good" in the German grade system) corresponds well with these ratings.

Table 1. Evaluation of the virtual operation tour.

Item	Median rating (1=do not agree; 5=strongly agree)	Agreement to statement (%)
The virtual operation room tour helped me to prepare for the surgery.	4	71.4
The information the operation room tour gave me was helpful for my preparation.	4	61.8
I found it helpful for my preparation to see the surgical environment before the surgery.	4	74.3
The virtual surgery tour calmed me down.	4	58.6
The virtual surgery tour was too detailed.	2	19.4
The virtual surgery tour was not detailed enough.	2	8.8
Should I be operated on again, I would like to do the virtual surgery tour again.	4	54.3
I can recommend the virtual surgery tour to other patients.	4	91.4

ER2: Is There a Possible Relationship Between the Evaluation of VORT and Perioperative Trait Anxiety?

There were no significant correlations between any of the items from the evaluation of the VORT (Table 1) and STOA-T values of the patients (all *P*>.19).

ER3: Is There a Possible Relationship Between the Evaluation of VORT and Patients' Age?

There were no significant correlations between any of the items from the evaluation of the VORT (Table 1) and the age of the patients (all *P*>.10).

ER4: Does VORT Have an Influence on Anesthesia-Related STOA Items?

Comparing VORT and STOPP with a two-tailed independent samples *t* test for all 7 items related to anesthesia treatment, the difference was only found in favor of VORT slightly below the predefined significance threshold (*P*=.047). Comparisons for all other items were not significant, resulting in no evidence that VORT addresses anesthesia-specific fear components.

Moreover, we calculated comparisons between men and women for all single VORT rating items in two ways to assure that men and women felt equally addressed by our video using a male protagonist. First, we calculated independent samples *t* tests for the Likert scale ratings. Second, we compared male and female participants using X^2 tests to assess possible sex-dependent differences in the agreement to the statements (recoded ratings). For the *t* tests, no significant sex differences emerged concerning the ratings of VORT (all *P*>.17); for the X^2 tests, we found one significant difference for item 6 ("the virtual surgery tour was not detailed enough"), with higher agreement among female participants (21.4%) compared to males (0%). Moreover, we compared male and female participants with respect to their overall rating of VORT (school grade) in an independent samples *t* test. Men and women did not differ in their school grades (*P*=.25), with women tending towards better grades (mean 1.86, SD 0.66) compared to men (mean 2.10, SD 0.54).

Discussion

VORT Effects

Although VORT did not affect perioperative anxiety, the tour received good patient ratings and was perceived as helpful for surgery preparation. Remarkably, the positive evaluation of the virtual surgery tour was independent of the fear of surgery (ie, the personality trait or trait anxiety). In other words, the subjective benefit of VORT for the patients could not be attributed to anxiety reduction. A possible explanation for this finding may be the beneficial effect of preoperative education and information. The assumption of uncertainty being a central element in the emergence of hospitalization stress was first

raised by Mishel [12] and later specifically adapted to operation situations by Kagan and Bar-Tal [13], who reported that preoperative uncertainty negatively affected mental and physical recovery after surgery. This aligns with the findings of Moerman et al [14], showing that more than 80% of surgical patients had a positive attitude towards receiving information, with more than half of them having high information requirements. Receiving information about health care interventions positively affects several aspects of patients' mental and physical health, such as recovery, pain levels, and psychological distress [15-17]. Thus, in our study, the positive ratings of VORT may be attributed to reducing uncertainty and satisfying the need for information. Similar to Moerman et al [14], we conclude that surgical fear and need for information are related but distinguishable factors and addressing either of them can reduce patients' distress. For future studies, physiological measures such as heart rate, blood pressure, and skin conductivity may be relevant indicators for assessing the arousal of the patients. Including these physiological measures might help disentangle physiological processes and their cognitive appraisal [18]. For example, it seems conceivable that VORT evokes physiological arousal due to the operation-associated settings and situations but reduces subjective stress via cognitive control.

Surprisingly, VORT is perceived as beneficial irrespective of the patient's age, indicating 2 interesting aspects. First, the need for information and the wish to reduce uncertainty seem independent of the patient's age. Second, the acceptance of VORT was equally high among elder and younger patients. Therefore, there is no evidence that experience with VR may play any role in the evaluation or that elderly patients may feel uncomfortable with the VR setup. Given the high number of elderly patients undergoing surgeries with general anesthesia, these findings favor the general use of VORT in preoperative preparations.

Furthermore, there was no evidence that VORT addresses anesthesia-specific fear components. Our experience shows us that there are two patient groups: those who fear the anesthesia and its consequences in isolation and those who are only afraid of the procedure, the surgical part. As far as we know, no questionnaire provides information on fear of anesthesia in isolation.

VR and Perioperative Anxiety

There is mixed evidence concerning the potential of VR in reducing perioperative anxiety. Some studies show no reduction in anxiety [19-21] by interventions as in our study, and others offer a detectable influence [22-26]. However, anxiety reduction has been applied preoperatively with good results, especially in children [9,26,27]. These inconsistencies may be partly attributed to the use of different anxiety inventories. The Spielberger State and Trait Anxiety Inventory (STAI) is frequently used [28]. However, the suitability of the STAI has been questioned for operation situations since it primarily measures anxiety in assessment situations [2]. Accordingly, more specific inventories have been developed, such as the STOA [2], the Amsterdam Perioperative Anxiety and Information Scale [14], and the Yale Preoperative Anxiety Scale [29], which was designed especially for children. The content of the items, length of the questionnaire, and other aspects vary considerably between these inventories, influencing study outcomes. For instance, Krohne et al [2] demonstrated that trait scores of STAI and STOA were correlated at r=.40, indicating that both inventories reflected a common factor of general anxiety but measured distinguishable facets of anxiety. Therefore, when comparing studies on VR interventions for perioperative anxiety, one must consider which inventories were used.

Limitations

One possible limitation of our study is that all patients received the VR tour after the "educating interview." Regarding patient compliance, it would be relevant to know whether there are fewer canceled surgeries when patients receive a VR education beforehand. Moreover, with the changed order, the VR tour could prepare patients for the anesthesia interview, enabling the patient to ask more detailed and specific questions. The second important limitation concerns the limited generalizability of our results to other types of anesthesia (eg, regional procedures). This is relevant in terms of distinguishing separate aspects of preoperative anxiety. Specifically, we assume that a distinction between fear of anesthesia and fear of the operation procedure by itself can be much more precise since several aspects of general anesthesia are not provided here (eg, fear of not waking up, lack of trust, or losing control while being unconscious).

Another limitation concerns whether the patients already had previous surgery with general anesthesia. Our study did not evaluate this aspect, and we cannot exclude the possibility it played a role in the VORT assessment. However, after careful consideration, we conclude that it is rather unlikely that prior surgical experience with anesthesia affected the VORT assessment. The probability of undergoing surgery increases with life age, suggesting that there was most likely a substantial correlation between life age and operations in the past in our study sample as well [30]. Our data found no correlation between life age and VORT ratings. We are aware the latter cannot distinguish between people who have experienced an operation and those who have not; however, we conclude that there is at least no indication from our data that having undergone an operation or not plays a role in the VORT rating.

Conclusion

VORT is beneficial for a patient's preparation for surgery with general anesthesia. Remarkably, this effect did not result in reduced perioperative anxiety. Instead, VORT seems to serve the need for information and to reduce uncertainty. Moreover, it is both accepted and recommended by patients. Considering the manageable organizational and financial effort and minimal time required in its implementation and the amount of time spent waiting in the perioperative outpatient clinic, the general use of VORT can be recommended.

Acknowledgments

The authors thank all participating patients and the premedication ambulance staff for their valuable support. This study was funded by the EU Fund, Interreg EMR, Clinic for Anesthesiology, and University Hospital RWTH Aachen and medical faculty in Aachen, Germany.

Authors' Contributions

LV, MK, and UG developed the study concept and study design. LV and MK had the primary idea for this study. Data were collected by UG. MK completed the ethical approval and performed statistical data analyses. LV drafted the article. SS and RR provided substantial intellectual input, and SS served as the senior supervisor. MK, SS, and RR critically revised the manuscript for important intellectual content. PE provided substantial media support and was responsible for the implementation of the virtual reality. All authors read and approved the final version of the article. All authors agreed to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT checklist.
[PDF File (Adobe PDF File), 146 KB - jmir_v23i9e28018_app1.pdf]

References

1. Ip HYV, Abrishami A, Peng PWH, Wong J, Chung F. Predictors of postoperative pain and analgesic consumption: a qualitative systematic review. Anesthesiology 2009 Sep;111(3):657-677 [FREE Full text] [doi: 10.1097/ALN.0b013e3181aae87a] [Medline: 19672167]
2. Krohne HW, Schmukle SC, de Bruin J. The Inventory "State-Trait Operation Anxiety" (STOA): construction and empirical findings]. Psychother Psychosom Med Psychol 2005;55(3-4):209-220. [doi: 10.1055/s-2004-834604] [Medline: 15800815]
3. Teismann T, Margraf J. Exposition und Konfrontation. Verhaltenstherapiemanual 2017:104-115. [doi: 10.1026/02825-000]
4. Carl E, Stein AT, Levihn-Coon A, Pogue JR, Rothbaum B, Emmelkamp P, et al. Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials. J Anxiety Disord 2019 Jan;61:27-36. [doi: 10.1016/j.janxdis.2018.08.003] [Medline: 30287083]
5. Garcia-Palacios A, Botella C, Hoffman H, Fabregat S. Comparing acceptance and refusal rates of virtual reality exposure vs. in vivo exposure by patients with specific phobias. Cyberpsychol Behav 2007 Oct;10(5):722-724. [doi: 10.1089/cpb.2007.9962] [Medline: 17927544]
6. Garcia-Palacios A, Hoffman H, Carlin A, Furness T, Botella C. Virtual reality in the treatment of spider phobia: a controlled study. Behaviour Research and Therapy 2002 Sep;40(9):983-993. [doi: 10.1016/s0005-7967(01)00068-7]
7. Miloff A, Lindner P, Hamilton W, Reuterskiöld L, Andersson G, Carlbring P. Single-session gamified virtual reality exposure therapy for spider phobia vs. traditional exposure therapy: study protocol for a randomized controlled non-inferiority trial. Trials 2016 Feb 02;17:60 [FREE Full text] [doi: 10.1186/s13063-016-1171-1] [Medline: 26833396]
8. Parsons TD, Rizzo AA. Initial validation of a virtual environment for assessment of memory functioning: virtual reality cognitive performance assessment test. Cyberpsychol Behav 2008 Feb 07;11(1):17-25. [doi: 10.1089/cpb.2007.9934] [Medline: 18275308]
9. Ryu J, Park S, Park J, Kim J, Yoo H, Kim T, et al. Randomized clinical trial of immersive virtual reality tour of the operating theatre in children before anaesthesia. Br J Surg 2017 Nov 04;104(12):1628-1633. [doi: 10.1002/bjs.10684] [Medline: 28975600]
10. WMA Declaration Of Helsinki: Ethical Principles For Medical Research Involving Human Subjects. World Medical Association. URL: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ [accessed 2021-08-16]
11. https://www.youtube.com/watch?v=twpTilf4xAA. URL: https://www.youtube.com/watch?v=twpTilf4xAA [accessed 2021-05-19]
12. Mishel MH. Perceived uncertainty and stress in illness. Res Nurs Health 1984 Sep;7(3):163-171. [doi: 10.1002/nur.4770070304] [Medline: 6567948]
13. Kagan I, Bar-Tal Y. The effect of preoperative uncertainty and anxiety on short-term recovery after elective arthroplasty. J Clin Nurs 2008 Mar;17(5):576-583. [doi: 10.1111/j.1365-2702.2007.01968.x] [Medline: 18093119]
14. Moerman N, van Dam FS, Muller MJ, Oosting H. The Amsterdam Preoperative Anxiety and Information Scale (APAIS). Anesth Analg 1996 Mar;82(3):445-451. [doi: 10.1097/00000539-199603000-00002] [Medline: 8623940]

15. Lai YL, Van Heuven A, Borire A, Kandula T, Colebatch JG, Krishnan AV, et al. The provision of written information and its effect on levels of pain and anxiety during electrodiagnostic studies: A randomised controlled trial. PLoS One 2018;13(5):e0196917 [FREE Full text] [doi: 10.1371/journal.pone.0196917] [Medline: 29758078]
16. Sugai DY, Deptula PL, Parsa AA, Don Parsa F. The importance of communication in the management of postoperative pain. Hawaii J Med Public Health 2013 Jun;72(6):180-184 [FREE Full text] [Medline: 23795326]
17. Devine E. Effects of psychoeducational care for adult surgical patients: A meta-analysis of 191 studies. Patient Education and Counseling 1992 Apr;19(2):129-142. [doi: 10.1016/0738-3991(92)90193-m]
18. Damasio A. The somatic marker hypothesis and the possible functions of the prefrontal cortex. Philos Trans R Soc Lond B Biol Sci 1996 Oct 29;351(1346):1413-1420. [doi: 10.1098/rstb.1996.0125] [Medline: 8941953]
19. Salzwedel C, Petersen C, Blanc I, Koch U, Goetz AE, Schuster M. The effect of detailed, video-assisted anesthesia risk education on patient anxiety and the duration of the preanesthetic interview: a randomized controlled trial. Anesth Analg 2008 Jan;106(1):202-209. [doi: 10.1213/01.ane.0000287665.96156.72] [Medline: 18165579]
20. Habibzadeh H, Milan ZD, Radfar M, Alilu L, Cund A. Effects of Peer-Facilitated, Video-Based and Combined Peer-and-Video Education on Anxiety Among Patients Undergoing Coronary Angiography: Randomised controlled trial. Sultan Qaboos Univ Med J 2018 Feb;18(1):e61-e67 [FREE Full text] [doi: 10.18295/squmj.2018.18.01.010] [Medline: 29666683]
21. Li Z, Geng W, Yin J, Zhang J. Effect of one comprehensive education course to lower anxiety and depression among Chinese breast cancer patients during the postoperative radiotherapy period - one randomized clinical trial. Radiat Oncol 2018 Jun 14;13(1):111 [FREE Full text] [doi: 10.1186/s13014-018-1054-6] [Medline: 29898748]
22. Garcia S. The effects of education on anxiety levels in patients receiving chemotherapy for the first time: an integrative review. Clin J Oncol Nurs 2014 Oct;18(5):516-521 [FREE Full text] [doi: 10.1188/14.CJON.18-05AP] [Medline: 25164233]
23. Cakmak M, Kose I, Zinzircioglu C, Karaman Y, Tekgul ZT, Pektas S, et al. [Effect of video-based education on anxiety and satisfaction of patients undergoing spinal anesthesia]. Braz J Anesthesiol 2018;68(3):274-279 [FREE Full text] [doi: 10.1016/j.bjan.2018.01.001] [Medline: 29636177]
24. Che Y, Gao Y, Jing J, Kuang Y, Zhang M. Effects of an Informational Video About Anesthesia on Pre- and Post-Elective Cesarean Section Anxiety and Recovery: A Randomized Controlled Trial. Med Sci Monit 2020 Apr 08;26:e920428 [FREE Full text] [doi: 10.12659/MSM.920428] [Medline: 32265432]
25. Lin S, Huang H, Lin S, Huang Y, Wang K, Shi H. The effect of an anaesthetic patient information video on perioperative anxiety: A randomised study. Eur J Anaesthesiol 2016 Feb;33(2):134-139. [doi: 10.1097/EJA.0000000000000307] [Medline: 26196527]
26. Liu PP, Sun Y, Wu C, Xu WH, Zhang RD, Zheng JJ, et al. The effectiveness of transport in a toy car for reducing preoperative anxiety in preschool children: a randomised controlled prospective trial. Br J Anaesth 2018 Aug;121(2):438-444 [FREE Full text] [doi: 10.1016/j.bja.2018.02.067] [Medline: 30032883]
27. Kain ZN, Mayes LC, Cicchetti DV, Bagnall AL, Finley JD, Hofstadter MB. The Yale Preoperative Anxiety Scale: how does it compare with a "gold standard"? Anesth Analg 1997 Oct;85(4):783-788. [doi: 10.1097/00000539-199710000-00012] [Medline: 9322455]
28. Spielberger CD, Gorsuch RL, Lushene R, Vagg PR, Jacobs GA. Manual for the State-Trait Anxiety Inventory. Palo Alto: Consulting Psychologists Press; 1983.
29. Kain ZN, Mayes LC, Cicchetti DV, Caramico LA, Spieker M, Nygren MM, et al. Measurement tool for preoperative anxiety in young children: The yale preoperative anxiety scale. Child Neuropsychology 1995 Dec;1(3):203-210. [doi: 10.1080/09297049508400225]
30. Anzahl der Operationen vollstationärer Patienten in deutschen Krankenhäusern nach Altersgruppe im Jahr 2019. Federal Statistical Office of Germany. URL: https://de.statista.com/statistik/daten/studie/29220/umfrage/krankenhaeuser-operationen-vollstationaerer-patienten/#statisticContainer [accessed 2021-05-10]

Abbreviations

STAI: Spielberger State and Trait Anxiety Inventory
STOA: state-trait operation anxiety
STOPP: standard operation preparation procedure
VORT: virtual operating room tour
VR: virtual reality

Edited by R Kukafka; submitted 18.02.21; peer-reviewed by S Said Yekta-Michael, HY Shi; comments to author 23.04.21; revised version received 19.05.21; accepted 14.06.21; published 01.09.21.

Please cite as:
Vogt L, Klasen M, Rossaint R, Goeretz U, Ebus P, Sopka S
Virtual Reality Tour to Reduce Perioperative Anxiety in an Operating Setting Before Anesthesia: Randomized Clinical Trial
J Med Internet Res 2021;23(9):e28018
URL: https://www.jmir.org/2021/9/e28018
doi:10.2196/28018
PMID:34252034

Review

Virtual Reality Technology Use in Cigarette Craving and Smoking Interventions (I “Virtually” Quit): Systematic Review

Merel Keijsers[1,2,3], BSc, MSc, PhD; Maria Cecilia Vega-Corredor[4], MSc, DVM; Melanie Tomintz[4], DI, PCTT, PhD; Simon Hoermann[1,2], BSc, MSc, PhD

[1]School of Product Design, College of Engineering, University of Canterbury, Christchurch, New Zealand

[2]HIT Lab NZ, College of Engineering, University of Canterbury, Christchurch, New Zealand

[3]John Cabot University, Rome, Italy

[4]Geospatial Research Institute, University of Canterbury, Christchurch, New Zealand

Corresponding Author:
Simon Hoermann, BSc, MSc, PhD
School of Product Design
College of Engineering
University of Canterbury
9 Engineering Road
Christchurch, 8041
New Zealand
Phone: 64 3 369 2457
Email: simon.hoermann@canterbury.ac.nz

Abstract

Background: Over the last 2 decades, virtual reality technologies (VRTs) have been proposed as a way to enhance and improve smoking cessation therapy.

Objective: This systematic review aims to evaluate and summarize the current knowledge on the application of VRT in various smoking cessation therapies, as well as to explore potential directions for future research and intervention development.

Methods: A literature review of smoking interventions using VRT was conducted.

Results: Not all intervention studies included an alternative therapy or a placebo condition against which the effectiveness of the intervention could be benchmarked, or a follow-up measure to ensure that the effects were lasting. Virtual reality (VR) cue exposure therapy was the most extensively studied intervention, but its effect on long-term smoking behavior was inconsistent. Behavioral therapies such as a VR approach-avoidance task or gamified interventions were less common but reported positive results. Notably, only 1 study combined Electronic Nicotine Delivery Devices with VRT.

Conclusions: The inclusion of a behavioral component, as is done in the VR approach-avoidance task and gamified interventions, may be an interesting avenue for future research on smoking interventions. As Electronic Nicotine Delivery Devices are still the subject of much controversy, their potential to support smoking cessation remains unclear. For future research, behavioral or multicomponent interventions are promising avenues of exploration. Future studies should improve their validity by comparing their intervention group with at least 1 alternative or placebo control group, as well as incorporating follow-up measures.

(J Med Internet Res 2021;23(9):e24307) doi:10.2196/24307

KEYWORDS

virtual reality technology; nicotine dependency; nicotine addiction; smoking addiction; smoking intervention; smoking therapy; Electronic Nicotine Delivery Systems

Introduction

Background

Smoking addiction is a worldwide [1] and costly [2] problem. Tobacco smoking has been linked to cardiovascular disease [3], various types of cancer [4-6] and respiratory problems [7], among other diseases, reducing the consumers’ quality of life [8], and lowering life expectancy by up to 4 years [9]. Approximately 30% of all cancer deaths in high-income countries are caused by cigarette smoking [10].

The Diagnostic and Statistical Manual of Mental Disorders lists nicotine dependency (rephrased as tobacco use disorder in a more recent edition) as a psychological disorder [11]; 55% of US smokers try to quit every year [12], but it has been estimated that between 60% and 90% of them relapse within the first year [12-14]. Smoking addiction has a strong psychological component [15,16] in addition to physiological dependence. Physiological craving may only take a few weeks to wane [17], but the habit and psychological link between smoking, socializing, relaxing, and rewarding are much harder to abandon [16]. Thus, different therapies and interventions have been developed to overcome this link.

Interventions and Therapies

Therapies and interventions are both care strategies, but their main difference is the scope. Interventions usually aim to motivate someone to commit to a specific action, such as a teenager saying *no* when their peers offer a cigarette or a smoker entering treatment. Therapies, on the other hand, address more comprehensive goals, such as maintaining abstinence and altering habits or attitudes. In practice, there is often no clear distinction between the two and they can be thought of as lying on a continuum rather than reflecting 2 dichotomous categories [18].

Cue exposure therapy (CET) uses a classical conditioning approach to unlearn a response (craving) to stimuli (smoking-related items and situations). During the therapy sessions, craving is elicited by exposing participants to smoking-related cues such as cigarette packages and images of situations in which the participant usually smoked. In daily life, participants would relieve cravings through smoking, which reinforces the craving response. During CET, typically no nicotine reward is provided [19] (however, see the study by Kotlyar et al [20] and De La Garza et al [21] that combines CET with nicotine replacement therapy [NRT]). Once the craving has receded, the participant is exposed to smoking cues again, with the aim of repeating this procedure until the association between smoking cues and reward is weakened and eventually extinguished [19,22].

The approach-avoidance task (AAT) adopts an operant conditioning approach. This task can be used to measure and influence subconscious bias [23,24]. During the AAT, participants have to either pull or push away a lever in response to visual cues. People tend to be quicker to pull when presented with cues that elicit a positive (approach) bias than when cues have a neutral or negative association, whereas they are quicker to push away (avoid) negative cues than positive or neutral ones [23]. Thus, AAT can be used to measure existing subconscious biases, as smokers will be quicker to pull when presented with smoking-related cues than nonsmokers. In addition, AAT can be used as an instrument to break or reverse subconscious biases by instructing participants to respond to smoking cues with avoidance behaviors (eg, pushing away) over a number of trials or sessions until smoking cues are automatically associated with avoidance and negative affect [24,25].

Cognitive behavioral therapy (CBT)–based interventions often take a multicomponent approach: teaching smokers to recognize the thought patterns they engage in before smoking and to identify triggering factors in the environment. Smokers are then trained in alternative strategies to cope with craving and temptation [26].

NRT aims to reduce cravings and by extension smoking behavior by replacing the source of nicotine [27]. It is an accessible standalone intervention for smokers who do not have the resources or motivation to commit to a more substantial intervention but will not target the psychological component of the smoking habit [27]. To this effect, NRT is often combined with CBT [28,29].

One particular form of NRT includes the use of Electronic Nicotine Delivery Systems (ENDS), more commonly known as electronic cigarettes, e-cigarettes, or vapes. Invented at the start of the 2000s, these devices heat a solution usually containing nicotine and flavoring agents and deliver the vapor (aerosol) to the user to be inhaled [30], although the amount of nicotine varies and nonnicotine liquids are available as well [31,32]. ENDS were originally introduced as a device to help reduce the number of cigarettes smoked or even quit (conventional) smoking altogether [30,33,34]. However, there has been an ongoing debate in the scientific and medical community on the health risks associated with ENDS [35,36] as well as the potential of ENDS to form an introduction to smoking for previous nonsmokers (the gateway hypothesis; [37]). Recent meta-analyses suggest that although far from harmless, ENDS may still pose a lower health risk than smoking cigarettes [37,38] and that they may be more effective than NRT for smoking cessation [39]. The question of whether this makes ENDS an eligible form of NRT has yet to be resolved [40,41].

Virtual reality technologies (VRTs) have been recognized as potentially helpful in increasing the effects of these and other interventions. These technologies provide an immersive interface that can be used to enhance (augmented reality [AR]) or even replace (virtual reality [VR]) reality with computer-generated simulations. AR is commonly used on a screen that combines the display of the real world with some added virtual features; a well-known example is the game Pokémon Go (Niantic Inc), which displayed the camera view on screen but added virtual Pokémon creatures to the scene, which the user could interact with. In contrast, in VR apps, users often wear a VR headset such as a stereoscopic head-mounted display that projects video images in 3D. Although the focus in VR and AR generally lies in visual simulation, the experience can be enhanced through haptic, olfactory, and audio feedback.

There are multiple potential benefits of using the VRT. First, training smokers to respond to a potentially tempting situation will be more effective if the therapy can mimic those situations more closely [42,43], and VR has been shown to create a stronger feeling of being immersed than 2D images [44-46]. Second, VR offers a safe environment for coping skills [47]. Finally, it can be easily tailored to the specific needs of individual smokers [22,42]. Therefore, over the last 15 years, the potential of VR for smoking cessation therapy has been extensively studied. Although AR has not yet been widely embraced, it was still included in the current search strategy.

This Research

In this paper, we systematically review and summarize the findings from the literature on the adoption of VRT in smoking cessation therapy.

This paper is centered around three main research questions:

- Has VRT been used satisfactorily to elicit smoking cravings?
- What VRT interventions exist and how do they compare with regular interventions in terms of smoking cessation outcomes?
- What are the potential future directions for VRT in smoking cessation therapy?

The review will focus on the adult population (people aged ≥18 years) of smokers with no comorbidities. Randomized controlled trials, controlled trials, single group pre- and posttest studies, and case studies were all included, as well as protocols (to give an indication of future directions of research), meta-analyses (as those can detect effects with greater power than individual studies), and reviews (for the reference list search and to refer to as an overview for the interested reader). Intervention studies were included if they incorporated VRT in their intervention and measured either smoking cue reactivity or intervention effectiveness. With regard to the data extracted, the comparators were treatment, placebo, pre- and postcomparison, and waiting list. The outcomes were craving or smoking urge, nicotine dependence, number of cigarettes smoked, abstinence rates, and quit rates.

Some systematic review papers have been published [48-52], but their scopes only partially overlapped with that of this one; for example, some systematic reviews included nonadult subjects [50], focused exclusively on cue reactivity or CET [48-50,52], limited themselves to head-mounted display VR [51], studied a variety of addictions [48,50,51], or were simply published over 5 years ago, thus missing more recent publications [49,50]. This paper includes 15 papers that were not covered by the previous reviews and thus expands on previous findings by presenting a wider and updated overview of the potential of VRT in smoking cessation.

Methods

Search Criteria

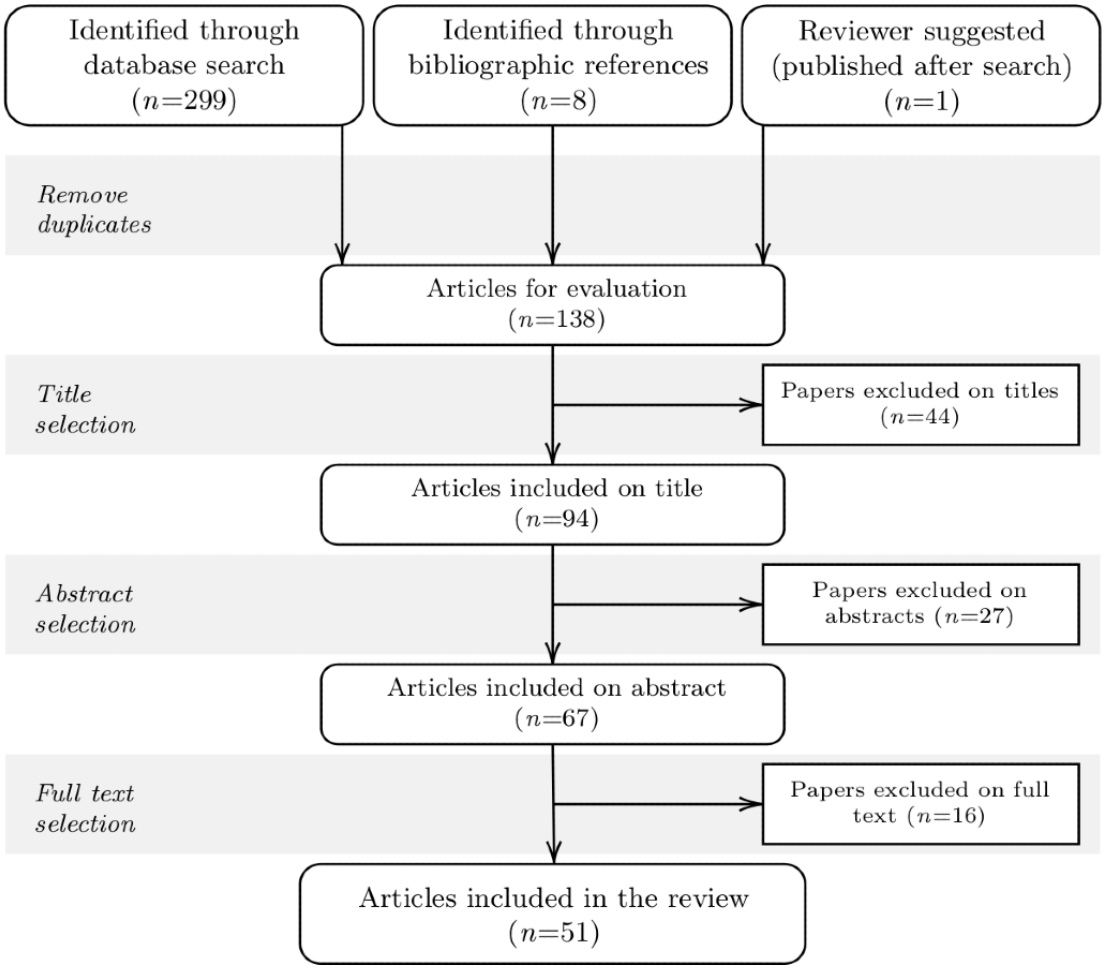
The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for review [53] were followed for the current review. The database search strategy was developed under the supervision of a trained subject librarian. After deliberation with the librarian, 5 databases (MEDLINE, Embase, Scopus, Cochrane, and EbscoHost) were identified to be searched using the following search terms: (*Virtual reality*, OR *Augmented reality*, OR *Mixed Reality*, OR *Augmented Virtuality*) AND (*smoking* OR *vaping* OR *tobacco* OR *cigarette* OR *nicotine* OR *vape* OR *e-cigarette*) AND (*craving* OR *crave* OR *withdrawal* OR *cue exposure* OR *cue reactivity* OR *urge* OR *cessation* OR *desire*). The reference lists of selected papers were also scanned for other relevant papers.

Literature Selection

The search was completed in July 2020, resulting in a total of 299 papers: MEDLINE (n=51), Embase (n=51), Scopus (n=88), Cochrane (n=52), and EbscoHost (n=57). After removing duplicates, 137 papers remained. Three rounds of selection were completed: exclusion based on title, exclusion based on the abstract, and exclusion based on the entirety of the paper. Inclusion and exclusion criteria were determined beforehand [54], with the inclusion criteria being as follows: the paper had to discuss the (potential) use of VRT for smoking cessation treatment; the target population had to be adults, with no pre-existing psychological or physiological conditions; only tobacco smoking in the form of cigarettes, cigars, or pipe and the vaping of nicotine were included (ie, no smoking of crack, marijuana, or any other substance; smoking a nargile or bong; or vaping of tetrahydrocannabinol S containing liquids). After 3 rounds of selection, 42 papers remained. Furthermore, 8 papers that had been identified through the reference list search and met the inclusion and exclusion criteria were added to this selection, as well as 1 meta-review that was published after the search but recommended by the paper reviewer, resulting in a final set of 51 papers. However, because the 2 studies by Pericot-Valverde et al [42,55] are based on the same data set (Pericot-Valverde, personal communication, January 30, 2021), we counted them as 1 study, reducing the total number to 50. Paper selection and data extraction were completed by one of the authors. Figure 1 shows a flowchart overview of the review rounds (see Multimedia Appendix 1 [20,21,42-44,46,48-52,55-95] for a list of the selected papers).

The timestamped literature selection plan, including search terms and inclusion and exclusion criteria, can be found [54], and the search terms, as well as the raw results from the individual databases, can be found [96] as well as in the Multimedia Appendix 1.

Figure 1. Overview of the screening and paper selection process.

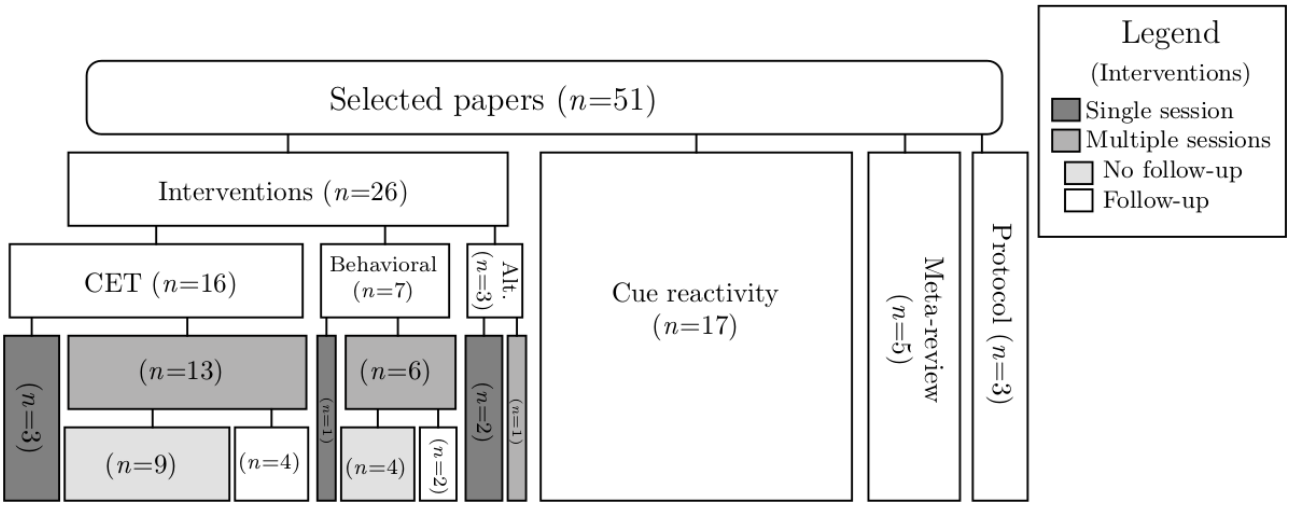


Results

Preliminary Analysis

A few general observations were made using the data set of papers. Of the 51 selected papers, 26 (51%) introduced or tested an intervention, 17 (33%) studied cue reactivity, 5 (10%) were meta-reviews, and 3 (6%) were protocols for studies yet to be conducted. In addition, 77% (22/26) of the intervention papers described a multisession intervention, stretching out from 3 to 10 weekly sessions. Furthermore, 27% (6/22) of these multisession interventions reported at least 1 follow-up measure, ranging from 1 week to 12 months after the final session. Follow-up measurements are a great asset to intervention studies as initial effects might fade over time (for example, the results in the study by Pericot-Valverde et al [42]). Figure 2 shows a flowchart indicating the division of the papers.

Figure 2. Overview of included studies. Alt: alternative (interventions that are neither based on cue exposure therapy nor behavioral therapy); CET: cue exposure therapy.



In addition, 54% (14/26) of the interventions were compared with an alternative therapy; however, only 15% (4/26) included a placebo condition. Although not including an alternative intervention or placebo condition does not invalidate the overall findings of a study, it also does not provide any information on the efficacy of one therapy over the other or indeed the efficacy of one therapy over an optimistic mindset.

Furthermore, of the 42 papers examining the effect of an intervention on smoking craving or behavior, 7 (17%) used smokers unmotivated to quit as participants; the rest of the papers was roughly equally divided between smokers motivated to quit (17/42, 40%) and smokers with undisclosed motivation to stop (18/42, 43%). All but 3 papers [44,56,57] used VR in their interventions.

Has VRT Been Used Satisfactorily to Elicit Smoking Cravings?

The prerequisite for VRT to have any (added) value in smoking interventions is whether or not these techniques can reliably induce craving through activation of smoking and smoking-related associations. Once these cravings have been elicited, any subsequent therapy can target the cravings. Cravings in non-VRT settings are usually elicited by using smoking-related paraphernalia [97], 2D images of cigarettes and scenes with people smoking [19], and imaginary procedures where participants are instructed to imagine themselves in a (personalized) high-risk situation [98]. VR can create a more immersive experience [44-46], which may add to the effectiveness of the therapy, and being virtual, it theoretically allows for tailoring specifically to the user's needs [42,58].

A plethora of studies (n=17) set out to test whether VR could induce nicotine craving and confirmed that craving can be successfully induced in VR by exposing participants to smoking-related cues [46,58-72] (see the study by Pericot-Valverde [49] for a meta-analysis). Moreover, exposing nonsmokers to VR smoking cues did not elicit cravings [46].

Furthermore, craving can be elicited using minimal cues. No explicit mention of smoking is needed; simply presenting an environment where cigarettes are usually handled (eg, a bar or the checkout counter of a newspaper kiosk) or providing smoking-related cues in the background (such as ashtrays) was sufficient to elicit cravings. The addition of olfactory cues [60,73] or branded cigarette packages [66,67] did not increase the craving response. Ferrer-García et al [74] tested different potential moderators of cue-induced craving and found that the (sense of) presence in VR was positively linked to craving.

Whether VR-cue exposure (CE) elicits craving to a greater extent than the more traditional methods has been less extensively studied. Bordnick et al [70] found that a VR environment elicited greater craving in participants than exposure to a 2D image of the same environment. In contrast, Karelitz [52] conducted a meta-analysis on different factors that may influence cue reactivity and found that although VR cues increased craving, they did not do so to a greater degree than 2D pictorial images or scripted imagery.

It is noteworthy that the number of years that participants had been smoking correlated negatively with cue reactivity [52]. The majority (11/17, 65%) of the cue reactivity studies did not report measuring the number of years smoking; of the 6 that did, 5 did not use it as a control variable or to assess whether the experimental groups were similar before commencing the experiment, thus potentially allowing for bias. Related to this, only 6 of the papers on cue reactivity reported participants' intention to quit. However, CE supposedly elicits craving in both types of cigarette smokers to an equal extent [62], so this lacuna in the data may not be problematic.

What VRT Interventions Exist and How Do They Compare to Regular Interventions in Terms of Smoking Outcomes?

Overview

Different VRT adaptations of smoking interventions were found. CET in VR (VR-CET) was the most frequently reported (n=16). In addition, VR has been combined with different behavioral interventions (eg, the AAT and gamified behavior training; n=7) and antismoking campaigns (n=3). Figure 2 shows a flowchart detailing the breakdown of the topics.

VR-CET Intervention

A total of 16 papers on VR-CET or a variation were included in the review. VR-CET has been successful in reducing smoking cravings over repeated exposures in most experiments [42,58,60,64,68,70,75,76] experiments. One study in the selected papers did not find a significant effect of VR-CET on long-term cravings [77], but their sample size was rather limited at n=8. The studies by Hone-Blanchet et al [50] and Segawa et al [51] present a meta-review on the effectiveness of VR-CET on craving and smoking behavior. A lower amount of smoking behavior after VR-CET treatment compared with the preintervention baseline was reported in the studies by Choi et al [67] and Pericot-Valverde et al [75]. Choi et al [67] measured a reduction from approximately 14 to 4 cigarettes per day based on self-reports, which was substantiated by a significant reduction in exhaled carbon monoxide (CO; 12.5 to 2.5 ppm). Pericot-Valverde et al [75] reported a reduction from 18 cigarettes per day before the intervention to a little under 4 cigarettes per day after the intervention (CO change from 13.2 to 5.0 ppm).

Furthermore, 4 studies investigated the added value of VR-CET to CBT [42,58,61,78]. Culbertson et al [58] compared craving and smoking behaviors between a group of smokers who received CBT and VR-CET and a group of smokers who received CBT and a placebo-VR training. Participants in the VR-CET group not only reported less craving and smoked less after the treatment but also had a significantly higher dropout rate. Thus, it is possible that the results could be biased, as only the participants for whom the VR-CET treatment worked well stayed in the program. Moreover, the intervention effects were only measured directly after intervention.

In addition, 3 studies [42,61,78] did not add a placebo-VR condition but instead compared the craving and smoking behaviors of participants receiving CBT. In the study by Papini et al [78], a group of smokers receiving VR-CET was compared with a group receiving CBT; in the other studies [42,61], both groups received CBT, but only 1 additionally received VR-CET

[42,61]. No added value of VR-CET to CBT was found in either of the 3 studies. Furthermore, similar data were found on reduction in dependence and craving [61], abstinence rates at the end of the treatment period [42,78], and decrease in the mean number of cigarettes smoked [78]. In order to further substantiate these findings, Pericot-Valverde et al [42] and Park et al [78] used follow-up measures that confirmed no added benefit of VR-CET to CBT. Relapse at 2 months postintervention was similar between the groups (47% for CBT and 53% for VR-CET) [78]. Pericot-Valverde et al [42] found a significantly higher relapse rate at 12 months postintervention for VR-CET compared with CBT alone (64.3% vs 37%). Relapse rates between 60% and 90% have been reported in the general literature for the first year of quitting. Thus, VR-CET may only marginally improve the odds of smokers successfully giving up their habits.

Three papers described a combination or adaptation of VR-CET with other interventions. First, Kotlyar et al [20] combined VR-CET with NRT by having a third of the participants ingesting a nicotine lozenge before the VR-CET session, thereby artificially reducing the body's craving for nicotine; the other two-thirds received either a placebo lozenge or none at all. They found that VR-CET combined with both lozenges that contained nicotine and the placebo lozenges reduced self-reported craving compared with no lozenge.

Second, in a similar design, De La Garza et al [21] combined VR-CET with ENDS, where the e-liquid contained different doses of nicotine (0 mg, 8 mg, and 16 mg). Compared with smoking a conventional cigarette, ENDS did not reduce elicited cravings as much, even if they contained a high dose of nicotine. However, the authors noted that the ENDS they used have been known to deliver inferior amounts of nicotine, even with high-nicotine liquids. Moreover, the study did not report whether it controlled for previous ENDS use by the participants.

These results show an interesting pattern in combination with the findings from Moon and Lee [77], who found that even as brain areas associated with addiction and craving became less active over time during VR-CET, participants still reported feeling similar levels of craving. Overall, these studies further underline that bodily craving and subjective feeling of craving only partially overlap, and that nicotine addiction has a strong psychological component as well that a therapeutic intervention should seek to address. This intricate set of motivations for smoking may partially explain why VR-CET as a smoking cessation therapy has resulted in such mixed results. However, more research on the effects of VR-CET in combination with NRT is needed to confirm and nuance this hypothesis. It will be interesting to see the results from one of the protocol papers, which outlined a planned clinical trial that combined VR-CET with NRT (eg, see the study by Papini et al [79]). This study will have a larger number of participants (n=102 vs n=41 [20] and n=7 [21]), and although it does not propose a multisession VR-CET intervention or a follow-up measure, it will still be interesting to see if the results from this intervention are in line with previous findings.

In the third study [80], VR-CET was combined with mindfulness training. Participants were taught mindfulness as a coping mechanism to regulate cravings after CE. Thus, the CET aspect in this intervention moved away from pure extinction in favor of a more cognitive response training component; however, as the authors presented their work as an adaptation of VR-CET, it is listed here as such [80]. This adaptation of VR-CET obtained good results: significantly more (23%) of the mindfulness VR-CET trained participants quit smoking at the end of the intervention compared with the control group (5%), who had received a smoking cessation manual for self-assisted quitting. Furthermore, 3 months after intervention, this significant difference remained, as the percentage of smokers who quit had grown to 33% for the intervention group and remained the same for the control group.

With regard to the entire group of papers on VR-CET, a little over half (n=9) of the papers did not report on participants' years of smoking. Of the 7 that did, 1 (14%) used it as a covariate and 3 (43%) used it to ensure that the experimental groups were similar. Given how smoking history may have moderated the cue response [74], the 12 papers that did not control for it may have been at risk of bias. About two-thirds of the VR-CET studies (n=11) reported using participants who were motivated to stop smoking; only 1 study targeted smokers who were unmotivated to quit. Although this may not have been problematic for eliciting craving [62], it is unknown whether the motivation to quit affects response extinguishing.

Behavioral Interventions

A total of 7 VRT-based smoking interventions that relied on a behavioral component were identified: 1 VR adaptation of the AAT (VR-AAT), 4 papers on gamified behavioral training, and 2 on skill training.

The AAT can be used to measure and break and reverse subconscious biases [23,24]. In a smoking intervention context, this is done by instructing smokers to carry out an avoidance behavior (eg, pushing or swatting away) when confronted with smoking cues, thus reconditioning smoking cues to be associated with avoidance rather than approaching behavior [24]. VR-AAT is a relative newcomer in the list of VR smoking interventions. The 2 papers (supposedly using the same sample, and therefore counted as a single experiment in this review) were published in 2019. As a result, the studies merely functioned as proof of concept for a VR adaptation of the AAT. Indeed, VR-AAT was found to be equally useful for measuring cognitive bias in smokers as regular AAT, and smokers displayed a stronger positive bias toward smoking cues than nonsmokers [81,82].

The 4 gamified interventions all appear to be inspired by the AAT: participants kicked and slapped away cigarette-related cues [56] and crushed cigarettes [83,84] or an alternative *cue*/*reaction* element was incorporated [85].

In a study [56], participants completed 3 weekly sessions in VR and reported on their smoking behavior a week after intervention. No statistical testing of the data was reported; hence, it is not clear if the results were statistically significant.

Participants in the studies by Girard et al [83,84] completed 4 weekly sessions in which they either crushed virtual cigarettes or played a VR placebo game, in addition to smoking cessation counseling. The game significantly reduced smoking behavior

as measured directly after the intervention, both in terms of the number of participants who quit and the number of cigarettes smoked by the still-smoking participants [84]. Moreover, 6 months after the intervention, participants who had played the VR crushing cigarette game reported significantly lower craving, as well as smoking fewer cigarettes, whereas a greater number had quit altogether [83]. In addition, fewer participants dropped out from the gamified intervention, and the participants who dropped out did so at a later stage in the program compared with the placebo game [83]. Given that dropout rates in smoking interventions can reach as high as 70% [99] and dropouts are more likely to have higher nicotine dependence and be heavy smokers, especially for early dropouts [99,100], lower dropout rates are a major advantage.

The third gamified VR intervention consisted of 9 weekly sessions [85]. The game is an adaptation of the game proposed by Girard et al [83,84] and has been explicitly designed around the psychological factors underlying motivation. It incorporates elements from both self-determination theory and cue or reaction therapy, thus combining operant (behavior) conditioning with psychological coaching. Although the only data available is a pilot sample (n=8) and no statistical tests have been performed because of the small sample size, the initial results are promising, with lower self-reported dependence as well as a reduction in smoking behavior.

Although these interventions were not phrased as an AAT type of intervention, their results suggest that avoidance behavior training (whether it is pushing away or crushing) could be a promising addition to smoking cessation therapy. What is particularly interesting is the multicomponent approach of the last intervention, which makes it similar to CBT. The behavioral component in CBT, however, is more specific: participants get taught a different response to situations in which they usually give in and smoke, instead of training an automatic, subconscious avoidance response. Although the gamified studies [83,84] included a control group, an aspect that remains unclear is whether VR, the gamification elements, or an interaction between the two were the efficacious components of the interventions.

The 2 VR skill training programs appear to be based more on CBT than on AAT. Bordnick et al [86] combined a 10-week VR coping skill training with NRT and compared this to a standalone NRT intervention. During the VR training sessions, participants were immersed in craving-inducing situations in VR, while the therapist assisted them in identifying high-risk triggers and training coping responses. The addition of VR training reduced smoking rates to a greater degree than NRT alone, both when measured directly after treatment and at the 6-month follow-up. However, similar to the gamified interventions, whether VR was necessary (ie, if it was an improvement over regular CBT) was not tested.

Pericot-Valverde et al [87] described a single case study that used a different CBT-inspired VR intervention, Virtual Stop Smoking therapy. This multicomponent intervention involves self-monitoring, information sessions about smoking, stimulus control procedures, strategies for relapse prevention, problem-solving procedures, strategies to cope with withdrawal syndrome, physiological CO exhalation feedback, and VR exposure, spread out over 6 weekly sessions. The participant had successfully quit smoking by the end of the intervention, but a replication with a larger sample as well as control conditions will be needed before any recommendations can be made.

Of the 7 behavioral interventions described, 2 [83,86] used mean years of smoking as a randomization control variable; none of the other studies reported on the statistic. It is unclear if this factor has an influence on behavioral interventions as it does on cue reactivity. Furthermore, none of the papers reported on using a sample of smokers unmotivated to quit; about half (n=4) used participants motivated to stop smoking.

Antismoking Campaigns

Finally, there is the use of VRT in information-based antismoking campaigns that warn about the consequences of smoking. VRT can create a deeper impression than 2D images or movies [88], elicit a stronger negative response, and enhance behavioral intentions to not smoke [44]. However, whether this impact translated to an actual change in behavior was not tested.

Future Directions

Given the underwhelming results that were achieved by VR-CET in terms of smoking cessation, future studies might want to explore the potential of alternative interventions. However, at present, the interest in VR-CET does not appear to have waned yet, as shown by the number of recent studies that attempt to obtain results by combining VR-CET with other therapies, such as CBT [42,89] and NRT [20,79]. It is possible that combining VR-CET with other interventions will result in a more successful intervention, as multiple aspects of nicotine addiction are addressed.

In contrast to VR-CET, interventions that used some form of behavioral intervention showed promising results. More work will be needed to solidify these findings, especially as a number of the studies reported were merely intended as pilot studies or proof of concept. One protocol study [43] laid out the design for a VR-AAT study, which would be timely as thus far only proof of concept has been reported for VR-AAT interventions. The results from the gamified interventions are promising and could lead to a lower attrition rate.

Possibly because of the ongoing controversy [38,40] surrounding ENDS, few studies to date have explored the potential use of ENDS in combination with VRT for smoking cessation therapy. However, ENDS and VRT both aim to mimic real life and allow for control over the surroundings as well as the amount of nicotine ingested, which could make them an alternative for therapy settings. De La Garza et al [21] reported a pilot study on the combination of VR-CET and ENDS as a form of NRT. The participants indicated that ENDS were not as *satisfying* as the conventional cigarette, even if it contained equal amounts of nicotine; at the same time, the placebo nicotine delivered in the study by Kotlyar et al [20] reduced craving in the same way as the actual nicotine treatment. These findings suggest that craving is a subjective experience and can be tweaked by participants' beliefs of what relief they will feel. Capitalizing

on this make-believe could be an interesting avenue for ENDS-based NRT.

Discussion

Principal Findings

The use of VRT could offer an alternative for, or addition to, smoking cessation interventions. VR can be used to recreate triggering situations in a more life-like and persuasive way than traditional methods. This may allow for the creation of an experience where both the environment and the triggers presented are tailored to the user's circumstances, which may further enhance the effectiveness of the intervention.

Although VR-CET has been the most extensively researched of all interventions, the results have been mixed at best. VR CE reliably elicited cravings, and most VR-CET interventions found that by the end of the therapy, craving in response to smoking-related cues was reduced. However, the effectiveness of therapy above and beyond alternative interventions such as CBT is debatable. These findings echo the conclusions from earlier meta-reviews [48,50,51] and analysis [49]. This lack of results is disappointing but may not be unpredictable as regular CET has been shown to have very limited behavioral effectiveness in battling addictions [22,101]. In fact, this lack of results in conventional CET has been one of the larger motivations for the adoption of VR; extinction of cue reactivity has been shown to be highly context dependent [102], and it has been suggested [46,61] that the discrepancy between the controlled laboratory environment where people were conditioned and the real-world situations where their therapy was put to the test was too large. Introducing VR would supposedly reduce these problems with generalization but still does not yield a significant effect. This may be due to the quality of the virtual environments, as they are far from photorealistic. Addiction remains a multidimensional problem, and CET (in VR or a more traditional setting) may target a component that is too small to be effective as a standalone intervention. Alternatively, the lack of results may be a consequence of the quality of the virtual worlds. Since its introduction, VR has improved greatly, but even as graphics and technology have improved the environments (or at least the ones usually developed for academic studies, ie, on a budget) are far from photorealistic and rarely closely mimic the actual world of the users. Future studies with current VRT may be able to create more realistic and personalized virtual scenarios and thus potentially obtain more reliable results.

Three other types of intervention with a behavioral component emerged: AAT-based [79,80], gamified behavioral training [56,83-85], and skill training [86,87]. The proof of concept has been provided for VR-AAT [81,82], and the 2 immersive games based on AAT demonstrated some success [56,83,84]. Similarly, tentative initial findings suggest a potential for gamified interventions that combine behavioral therapy with psychological coaching [85]. Overall, the potential of gamified interventions as an (effective) treatment program is further enhanced through the addition of a gaming element. Well-designed games lead the user to play out of intrinsic rather than extrinsic motivation [103]; thus, designing an intervention in this format could elicit higher participant engagement, which would, in turn, lead to better therapy compliance and lower dropout rates than conventional interventions. Given the high dropout rates of smoking cessation therapies [104,105] and the importance of compliance in the effectiveness of an intervention [105], these would be major advantages over conventional (nongamified) therapies and interventions.

Only 1 paper was found that combined VR and ENDS in a smoking intervention, possibly because of the ongoing controversy surrounding ENDS [36,37]. Although no consensus has been reached yet on whether the downsides of ENDS outweigh the improvements over conventional cigarettes [38,39], implementing ENDS in future interventions as an NRT component may be explored more as a viable method for smoking cessation.

A final factor that needs consideration in the discussion of VR-based interventions is cost. With the gaming industry's growing interest in the technology [51], VR has become increasingly accessible over the last few years, with a variety of affordable systems and VR template environments readily available. Nonetheless, developing VRT materials is far from cheap, and the technology thus ought to have a considerable benefit over conventional methods if it is to replace them. Considering this drawback, AR may prove to be an easier and more economical alternative, as it does not require an entire world to be created from scratch but rather uses the existing one as a template to be enhanced or altered.

Limitations

Overview

A few limitations must be noted when interpreting the outcomes of this literature review. Some of these are the consequence of the methodology and design of the papers and interventions reviewed, and some are the result of the methodology of this review itself. All of these are presented in the following section.

Limitations of the Reviewed Literature

The first limitation is related to the sample size and sample population reported in the studies. Sample sizes varied widely (between 8 and 102 or 541 for the meta-analysis). Large samples reduce the chance of false negatives and increase the chance of finding small effects [106]. In addition to the size of the sample, not all studies differentiated between treatment-seeking and nontreatment-seeking smokers, which may have been a relevant background variable. Bordnick et al [62] indicated that this may not be a problem for CE–induced craving, but there may still be an effect of motivation to quit on other outcome measures.

The second limitation concerns the experimental design of the studies. The lack of control for potentially confounding variables such as motivation to quit or the number of years of smoking addiction has been mentioned before. Many studies used a 1D measure of effectiveness; for example, some of the VR-CET studies only measured self-reported cravings. Assuming that a therapy or an intervention is considered effective when it has led to a reduction or even cessation of smoking behavior, extrapolating a single-dimension measure to indicate the overall effectiveness of a treatment introduces method bias [107].

Ideally, studies should target more than one type of measure. For example, adding physiological measures for stress (heart rate and skin conductance), smoking behavior (CO exhaled), behavioral measures such as the amount of time the participant is willing to wait until the next cigarette, how much they would be willing to spend on a cigarette, and actual smoking behavior could be included.

Moreover, few studies included control groups, a blinded design, or a follow-up measure. Together with multidimensional measurements that target psychological as well as behavioral responses, these are all methodological aspects fundamental to fully assessing the potential of an intervention. However, these were not implemented in the majority of the studies reviewed in this paper. The lack of assessment of these aspects means that the results obtained in these studies should be interpreted and generalized with care.

Limitations of This Review

With regard to this literature review, publication bias [108,109] most likely prevented some studies on VRT smoking interventions from being included in this review. By searching the Cochrane database, at least the study protocols could be found as an indication of the experiments that never made it to publication. It is encouraging that of the 3 protocols that were found with no publication of the results, only one [89] had set out to finish their study before 2020 and could thus conceivably have been published in time for this literature review. Although there is no perfect measure, this can be taken as an indication that publication bias may have been low.

In addition, the literature selection and data extraction were performed by a single researcher, without a second independent researcher confirming the results. Having these decisions depending on 1 person may introduce bias.

Conclusions

The studies presented in this review suggest that VRT can be considered a promising addition to smoking cessation therapies. Although VR-CET by itself has not yielded consistent results, tentative initial findings on behavioral interventions as well as the combination of VR-CET with these interventions are promising. Moreover, the potential of ENDS in combination with VRT may offer an alternative for future research. More rigorous testing, especially in terms of larger sample sizes, the inclusion of control groups or placebo interventions, and follow-up measures, is still needed.

Acknowledgments

The project was funded by the Health Research Council of New Zealand (Explorer grant 18/738).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview table with details of all publications selected for this review, including details about the population, intervention method, design, and outcome of the study.
[DOCX File , 24 KB - jmir_v23i9e24307_app1.docx]

References

1. Degenhardt L, Glantz M, Evans-Lacko S, Sadikova E, Sampson N, Thornicroft G, World Health Organization's World Mental Health Surveys collaborators. Estimating treatment coverage for people with substance use disorders: an analysis of data from the World Mental Health Surveys. World Psychiatry 2017 Oct 21;16(3):299-307 [FREE Full text] [doi: 10.1002/wps.20457] [Medline: 28941090]
2. Xu X, Bishop EE, Kennedy SM, Simpson SA, Pechacek TF. Annual healthcare spending attributable to cigarette smoking: an update. Am J Prev Med 2015 Mar;48(3):326-333 [FREE Full text] [doi: 10.1016/j.amepre.2014.10.012] [Medline: 25498551]
3. Leone A. Biochemical markers of cardiovascular damage from tobacco smoke. Curr Pharm Des 2005 Jul 01;11(17):2199-2208. [doi: 10.2174/1381612054367391] [Medline: 16026289]
4. Hecht SS. Tobacco smoke carcinogens and lung cancer. J Natl Cancer Inst 1999 Jul 21;91(14):1194-1210. [doi: 10.1093/jnci/91.14.1194] [Medline: 10413421]
5. Samanic C, Kogevinas M, Dosemeci M, Malats N, Real FX, Garcia-Closas M, et al. Smoking and bladder cancer in Spain: effects of tobacco type, timing, environmental tobacco smoke, and gender. Cancer Epidemiol Biomark Prev 2006 Jul 01;15(7):1348-1354. [doi: 10.1158/1055-9965.epi-06-0021]
6. Peto R, Boreham J, Lopez A, Thun M, Heath C. Mortality from tobacco in developed countries: indirect estimation from national vital statistics. Lancet 1992 May;339(8804):1268-1278. [doi: 10.1016/0140-6736(92)91600-d]
7. Willemse B, Postma D, Timens W, ten Hacken N. The impact of smoking cessation on respiratory symptoms, lung function, airway hyperresponsiveness and inflammation. Eur Respir J 2004 Mar 01;23(3):464-476 [FREE Full text] [doi: 10.1183/09031936.04.00012704] [Medline: 15065840]

8. Nusselder W, Looman C, Marang-van de Mheen PJ, van de Mheen H, Mackenbach J. Smoking and the compression of morbidity. J Epidemiol Community Health 2000 Aug;54(8):566-574 [FREE Full text] [doi: 10.1136/jech.54.8.566] [Medline: 10890867]
9. Ferrucci L, Izmirlian G, Leveille S, Phillips CL, Corti M, Brock DB, et al. Smoking, physical activity, and active life expectancy. Am J Epidemiol 1999 Apr 01;149(7):645-653. [doi: 10.1093/oxfordjournals.aje.a009865] [Medline: 10192312]
10. World Health Organization. Tobacco or health: a global status report. Geneva: World Health Organization; 1997:1-495.
11. American Psychiatric Association. Tobacco-related disorders. In: Diagnostic and Statistical Manual of Mental Disorders, 5th Ed. Washington, DC: American Psychiatric Publishing; 2013:571-577.
12. Creamer MR, Wang TW, Babb S, Cullen KA, Day H, Willis G, et al. Tobacco product use and cessation indicators among adults - United States, 2018. MMWR Morb Mortal Wkly Rep 2019 Nov 15;68(45):1013-1019 [FREE Full text] [doi: 10.15585/mmwr.mm6845a2] [Medline: 31725711]
13. Garvey AJ, Bliss RE, Hitchcock JL, Heinold JW, Rosner B. Predictors of smoking relapse among self-quitters: a report from the normative aging study. Addict Behav 1992 Jul;17(4):367-377. [doi: 10.1016/0306-4603(92)90042-t]
14. Mothersill KJ, McDowell I, Rosser W. Subject characteristics and long term post-program smoking cessation. Addict Behav 1988 Jan;13(1):29-36. [doi: 10.1016/0306-4603(88)90022-6]
15. Carmody TP. Affect regulation, nicotine addiction, and smoking cessation. J Psychoactive Drugs 1989 Apr;21(3):331-342. [doi: 10.1080/02791072.1989.10472175] [Medline: 2681631]
16. Jimenez-Ruiz CA. Psychological and behavioural interventions for smoking cessation. In: Nardini S, editor. Smoking Cessation. Lausanne, Switzerland: European Respiratory Society Monograph; 2008:61-73.
17. Hughes JR, Higgins ST, Bickel WK. Nicotine withdrawal versus other drug withdrawal syndromes: similarities and dissimilarities. Addiction 1994 Nov;89(11):1461-1470. [doi: 10.1111/j.1360-0443.1994.tb03744.x] [Medline: 7841857]
18. Center for Substance Abuse Treatment. Chapter 1—Introduction to brief interventions and therapies. In: Brief Interventions and Brief Therapies for Substance Abuse. Rockville (MD): Substance Abuse and Mental Health Services Administration (US). (Treatment Improvement Protocol (TIP) Series, No. 34.); 1999:1-12.
19. Unrod M, Drobes DJ, Stasiewicz PR, Ditre JW, Heckman B, Miller RR, et al. Decline in cue-provoked craving during cue exposure therapy for smoking cessation. Nicotine Tob Res 2013 Sep 28;16(3):306-315. [doi: 10.1093/ntr/ntt145]
20. Kotlyar M, Vogel RI, Dufresne SR, Mills AM, Vuchetich JP. Effect of nicotine lozenge use prior to smoking cue presentation on craving and withdrawal symptom severity. Drug Alcohol Depend 2020 Jan;206:107706. [doi: 10.1016/j.drugalcdep.2019.107706]
21. De La Garza LG, Yoon J, Yammine L, Holst M, Salas R. Preliminary evaluation of the effects of electronic cigarettes versus own cigarette on withdrawal, craving, and smoking severity in tobacco-dependent volunteers. Neuropsychopharmacology 2016;41:S380 [FREE Full text] [doi: 10.1038/npp.2016.241]
22. Conklin CA, Tiffany ST. Applying extinction research and theory to cue-exposure addiction treatments. Addiction 2002 Feb;97(2):155-167. [doi: 10.1046/j.1360-0443.2002.00014.x] [Medline: 11860387]
23. Phaf RH, Mohr SE, Rotteveel M, Wicherts JM. Approach, avoidance, and affect: a meta-analysis of approach-avoidance tendencies in manual reaction time tasks. Front Psychol 2014 May 08;5:378 [FREE Full text] [doi: 10.3389/fpsyg.2014.00378] [Medline: 24847292]
24. Wittekind CE, Feist A, Schneider BC, Moritz S, Fritzsche A. The approach-avoidance task as an online intervention in cigarette smoking: a pilot study. J Behav Ther Exp Psychiatry 2015 Mar;46:115-120. [doi: 10.1016/j.jbtep.2014.08.006] [Medline: 25306247]
25. Heuer K, Rinck M, Becker ES. Avoidance of emotional facial expressions in social anxiety: The approach–avoidance task. Behav Res Ther 2007 Dec;45(12):2990-3001. [doi: 10.1016/j.brat.2007.08.010]
26. Becoña E, Vázquez FL. Does using relapse prevention increase the efficacy of a program for smoking cessation?: an empirical study. Psychol Rep 1997 Aug;81(1):291-296. [doi: 10.2466/pr0.1997.81.1.291] [Medline: 9293218]
27. Stead L, Perera R, Bullen C, Mant D, Hartmann-Boyce J, Cahill K, et al. Nicotine replacement therapy for smoking cessation. Cochrane Database Syst Rev 2012 Nov 14;11:CD000146. [doi: 10.1002/14651858.CD000146.pub4] [Medline: 23152200]
28. Lancaster T, Stead LF. Individual behavioural counselling for smoking cessation. Cochrane Database Syst Rev 2017 Mar 31;3(3):CD001292 [FREE Full text] [doi: 10.1002/14651858.CD001292.pub3] [Medline: 28361496]
29. Stead L, Koilpillai P, Fanshawe T, Lancaster T. Combined pharmacotherapy and behavioural interventions for smoking cessation. Cochrane Database Syst Rev 2016 Mar 24;3(3):CD008286 [FREE Full text] [doi: 10.1002/14651858.CD008286.pub3] [Medline: 27009521]
30. Grana R, Benowitz N, Glantz SA. E-cigarettes: a scientific review. Circulation 2014 May 12;129(19):1972-1986. [doi: 10.1161/CIRCULATIONAHA.114.007667]
31. Bennani I, Chentoufi MA, El Karbane M, Cheikh A, Bouatia M. E-cigarette quality control: impurity and nicotine level analysis in electronic cigarette refill liquids. Sci World J 2020 Apr 13;2020:3050189 [FREE Full text] [doi: 10.1155/2020/3050189] [Medline: 32327940]
32. Weaver SR, Kim H, Glasser AM, Sutfin EL, Barrington-Trimis J, Payne TJ, et al. Establishing consensus on survey measures for electronic nicotine and non-nicotine delivery system use: current challenges and considerations for researchers. Addict Behav 2018 Apr;79:203-212. [doi: 10.1016/j.addbeh.2017.11.016]

33. Grana RA, Ling PM. "Smoking Revolution": a content analysis of electronic cigarette retail websites. Am J Prev Med 2014 Apr;46(4):395-403. [doi: 10.1016/j.amepre.2013.12.010]
34. Yong H, Borland R, Cummings KM, Gravely S, Thrasher JF, McNeill A, et al. Reasons for regular vaping and for its discontinuation among smokers and recent ex-smokers: findings from the 2016 ITC Four Country Smoking and Vaping Survey. Addiction 2019 Oct 23;114 Suppl 1(S1):35-48 [FREE Full text] [doi: 10.1111/add.14593] [Medline: 30821861]
35. Sood AK, Kesic MJ, Hernandez ML. Electronic cigarettes: one size does not fit all. J Allergy Clin Immunol 2018 Jun;141(6):1973-1982. [doi: 10.1016/j.jaci.2018.02.029]
36. Cao DJ, Aldy K, Hsu S, McGetrick M, Verbeck G, De Silva I, et al. Review of health consequences of electronic cigarettes and the outbreak of electronic cigarette, or vaping, product use-associated lung injury. J Med Toxicol 2020 Jul 16;16(3):295-310 [FREE Full text] [doi: 10.1007/s13181-020-00772-w] [Medline: 32301069]
37. Friedman AS, Xu S. Associations of flavored e-cigarette uptake with subsequent smoking initiation and cessation. JAMA Netw Open 2020 Jun 01;3(6):e203826 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3826] [Medline: 32501490]
38. Hering T. [E-cigarettes: toxicological fiasco or better than not giving up smoking?]. Internist (Berl) 2020 Jun 15;61(6):634-643. [doi: 10.1007/s00108-020-00794-0] [Medline: 32415499]
39. Hartmann-Boyce J, McRobbie H, Lindson N, Bullen C, Begh R, Theodoulou A, et al. Electronic cigarettes for smoking cessation. Cochrane Database Syst Rev 2020 Oct 14;10:CD010216. [doi: 10.1002/14651858.CD010216.pub4] [Medline: 33052602]
40. Bell K, Keane H. Nicotine control: E-cigarettes, smoking and addiction. Int J Drug Policy 2012 May;23(3):242-247. [doi: 10.1016/j.drugpo.2012.01.006] [Medline: 22365155]
41. Cook R, Davidson P, Martin R, NIHR Dissemination Centre. E-cigarettes helped more smokers quit than nicotine replacement therapy. Br Med J 2019 May 10;365:l2036. [doi: 10.1136/bmj.l2036] [Medline: 31076461]
42. Pericot-Valverde I, Secades-Villa R, Gutiérrez-Maldonado J. A randomized clinical trial of cue exposure treatment through virtual reality for smoking cessation. J Subst Abuse Treat 2019 Jan;96:26-32. [doi: 10.1016/j.jsat.2018.10.003] [Medline: 30466545]
43. Machulska A, Eiler TJ, Grünewald A, Brück R, Jahn K, Niehaves B, et al. Promoting smoking abstinence in smokers willing to quit smoking through virtual reality-approach bias retraining: a study protocol for a randomized controlled trial. Trials 2020 Feb 26;21(1):227 [FREE Full text] [doi: 10.1186/s13063-020-4098-5] [Medline: 32102685]
44. Lee J, Jung S, Kim JW, Biocca F. Applying spatial augmented reality to anti-smoking message: focusing on spatial presence, negative emotions, and threat appraisal. Int J Hum Comput Stud 2018 Jul 09;35(9):751-760. [doi: 10.1080/10447318.2018.1489581]
45. Clifford R, Khan H, Hoermann S, Billinghurst M, Lindeman R. The effect of immersive displays on situation awareness in virtual environments for aerial firefighting air attack supervisor training. In: Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR). 2018 Presented at: IEEE Conference on Virtual Reality and 3D User Interfaces (VR); March 18-22, 2018; Tuebingen/Reutlingen, Germany. [doi: 10.1109/VR.2018.8446139]
46. García-Rodríguez O, Pericot-Valverde I, Gutiérrez-Maldonado J, Ferrer-García M, Secades-Villa R. Validation of smoking-related virtual environments for cue exposure therapy. Addict Behav 2012 Jun;37(6):703-708. [doi: 10.1016/j.addbeh.2012.02.013] [Medline: 22385732]
47. Litt M, Kadden R, Kabela-Cormier E. Individualized assessment and treatment program for alcohol dependence: results of an initial study to train coping skills. Addiction 2009;104(11):1837-1848. [doi: 10.1111/j.1360-0443.2009.02693.x]
48. Trahan MH, Maynard BR, Smith KS, Farina ASJ, Khoo YM. Virtual reality exposure therapy on alcohol and nicotine: a systematic review. Res Soc Work Pract 2019 Feb 17;29(8):876-891. [doi: 10.1177/1049731518823073]
49. Pericot-Valverde I, Germeroth LJ, Tiffany ST. The use of virtual reality in the production of cue-specific craving for cigarettes: a meta-analysis. Nicotine Tob Res 2016 May 08;18(5):538-546. [doi: 10.1093/ntr/ntv216] [Medline: 26453669]
50. Hone-Blanchet A, Wensing T, Fecteau S. The use of virtual reality in craving assessment and cue-exposure therapy in substance use disorders. Front Hum Neurosci 2014 Oct 17;8:844 [FREE Full text] [doi: 10.3389/fnhum.2014.00844] [Medline: 25368571]
51. Segawa T, Baudry T, Bourla A, Blanc J, Peretti C, Mouchabac S, et al. Virtual Reality (VR) in assessment and treatment of addictive disorders: a systematic review. Front Neurosci 2019 Jan 10;13:1409 [FREE Full text] [doi: 10.3389/fnins.2019.01409] [Medline: 31998066]
52. Karelitz JL. Differences in magnitude of cue reactivity across durations of smoking history: a meta-analysis. Nicotine Tob Res 2020 Jul 16;22(8):1267-1276 [FREE Full text] [doi: 10.1093/ntr/ntz071] [Medline: 31050735]
53. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
54. Keijsers M. Systematic literature review plan: VR smoking craving. Centre for Open Science Storage. 2020 Jul 17. URL: https://osf.io/fx34y/ [accessed 2020-11-06]
55. Pericot-Valverde I, Garcia-Rodriguez O, López-Núñez C, Weidberg S, Secades-Villa R. Does cue exposure treatment improve outcomes in smoking cessation treatment? A controlled trial. Drug Alcohol Depend 2015 Nov;156:e176. [doi: 10.1016/j.drugalcdep.2015.07.477]

56. Metcalf M, Rossie K, Stokes K, Tallman C, Tanner B. Virtual reality cue refusal video game for alcohol and cigarette recovery support: summative study. JMIR Serious Games 2018 Apr 16;6(2):e7 [FREE Full text] [doi: 10.2196/games.9231] [Medline: 29661748]
57. Muñoz M, Idrissi S, Sánchez-Barrera MB, Fernández-Santaella MC, Vila J. Tobacco craving and eyeblink startle modulation using 3D immersive environments: a pilot study. Psychol Addict Behav 2013 Mar;27(1):243-248. [doi: 10.1037/a0028745] [Medline: 22686962]
58. Culbertson CS, Shulenberger S, De La Garza R, Newton TF, Brody AL. Virtual reality cue exposure therapy for the treatment of tobacco dependence. J Cyber Ther Rehabil 2012;5(1):57-64 [FREE Full text] [Medline: 25342999]
59. Thompson-Lake D, Cooper K, Mahoney IJ, Bordnick P, Salas R, Kosten T, et al. Withdrawal symptoms and nicotine dependence severity predict virtual reality craving in cigarette-deprived smokers. Nicotine Tob Res 2015 Jul;17(7):796-802 [FREE Full text] [doi: 10.1093/ntr/ntu245] [Medline: 25475087]
60. Traylor AC, Bordnick PS, Carter BL. Assessing craving in young adult smokers using virtual reality. Am J Addict 2008 Jan;17(5):436-440. [doi: 10.1080/10550490802268876] [Medline: 18770087]
61. Malbos E, Borwell B, Cantalupi R, Lancon C. Virtual reality cue exposure for smoking relapse prevention: a comparative trial. In: Annual Review of Cybertherapy And Telemedicine. San Diego, CA: Interactive Media Institute; 2018:124-130.
62. Bordnick PS, Yoon JH, Kaganoff E, Carter B. Virtual reality cue reactivity assessment. Res Soc Work Pract 2013 Mar 28;23(4):419-425. [doi: 10.1177/1049731513482377]
63. Acker J, MacKillop J. Behavioral economic analysis of cue-elicited craving for tobacco: a virtual reality study. Nicotine Tob Res 2013 Aug;15(8):1409-1416 [FREE Full text] [doi: 10.1093/ntr/nts341] [Medline: 23322768]
64. Kaganoff E, Bordnick PS, Carter BL. Feasibility of using virtual reality to assess nicotine cue reactivity during treatment. Res Soc Work Pract 2012 Mar 01;22(2):159-165 [FREE Full text] [doi: 10.1177/1049731511428617] [Medline: 25110453]
65. Pericot-Valverde I, García-Rodríguez O, Gutierrez-Maldonado J, Ferrer-García M, Secades-Villa R. Evolution of smoking urge during exposure through virtual reality. In: Annual Review of Cybertherapy and Telemedicine. Amsterdam, Netherlands: IOS Press; 2011:74-79.
66. Paris MM, Carter BL, Traylor AC, Bordnick PS, Day SX, Armsworth MW, et al. Cue reactivity in virtual reality: the role of context. Addict Behav 2011 Jul;36(7):696-699 [FREE Full text] [doi: 10.1016/j.addbeh.2011.01.029] [Medline: 21349649]
67. Choi J, Park S, Lee J, Jung H, Lee H, Jin C, et al. The effect of repeated virtual nicotine cue exposure therapy on the psychophysiological responses: a preliminary study. Psychiatry Investig 2011 Jun;8(2):155-160 [FREE Full text] [doi: 10.4306/pi.2011.8.2.155] [Medline: 21852993]
68. Ferrer-García M, García-Rodríguez O, Gutiérrez-Maldonado J, Pericot-Valverde I, Secades-Villa R. Efficacy of virtual reality in triggering the craving to smoke: its relation to level of presence and nicotine dependence. Stud Health Technol Inform 2010;154:123-127. [Medline: 20543283]
69. Bordnick PS, Graap KM, Copp HL, Brooks J, Ferrer M. Virtual reality cue reactivity assessment in cigarette smokers. Cyberpsychol Behav 2005 Oct;8(5):487-492. [doi: 10.1089/cpb.2005.8.487] [Medline: 16232041]
70. Bordnick PS, Graap KM, Copp H, Brooks J, Ferrer M, Logue B. Utilizing virtual reality to standardize nicotine craving research: a pilot study. Addict Behav 2004 Dec;29(9):1889-1894. [doi: 10.1016/j.addbeh.2004.06.008] [Medline: 15530734]
71. Lee JH, Ku J, Kim K, Kim B, Kim IY, Yang B, et al. Experimental application of virtual reality for nicotine craving through cue exposure. Cyberpsychol Behav 2003 Jun;6(3):275-280. [doi: 10.1089/109493103322011560] [Medline: 12855083]
72. García-Rodríguez O, Weidberg S, Gutiérrez-Maldonado J, Secades-Villa R. Smoking a virtual cigarette increases craving among smokers. Addict Behav 2013 Oct;38(10):2551-2554. [doi: 10.1016/j.addbeh.2013.05.007] [Medline: 23793042]
73. Traylor AC, Bordnick PS, Carter BL. Using virtual reality to assess young adult smokers' attention to cues. Cyberpsychol Behav 2009 Aug;12(4):373-378 [FREE Full text] [doi: 10.1089/cpb.2009.0070] [Medline: 19630582]
74. Ferrer-García M, García-Rodríguez O, Pericot-Valverde I, Yoon JH, Secades-Villa R, Gutiérrez-Maldonado J. Predictors of smoking craving during virtual reality exposure. Presence Teleop Virt Environ 2012 Nov;21(4):423-434. [doi: 10.1162/pres_a_00125]
75. Pericot-Valverde I, Secades-Villa R, Gutiérrez-Maldonado J, García-Rodríguez O. Effects of systematic cue exposure through virtual reality on cigarette craving. Nicotine Tob Res 2014 Nov 23;16(11):1470-1477 [FREE Full text] [doi: 10.1093/ntr/ntu104] [Medline: 24962558]
76. Pericot-Valverde I, Ferrer-García M, Pla-Sanjuanelo J, Secades-Villa R, Gutiérrez J. Cue exposure treatment through virtual reality reduce cigarette craving in real life environments. In: Annual Review of Cybertherapy and Telemedicine. San Diego, CA: Interactive Media Institute; 2016:137-142.
77. Moon J, Lee J. Cue exposure treatment in a virtual environment to reduce nicotine craving: a functional MRI study. Cyberpsychol Behav 2009 Feb;12(1):43-45. [doi: 10.1089/cpb.2008.0032] [Medline: 19025463]
78. Park C, Choi J, Park SM, Lee J, Jung HY, Seol J, et al. Comparison of the effectiveness of virtual cue exposure therapy and cognitive behavioral therapy for nicotine dependence. Cyberpsychol Behav Soc Netw 2014 Apr;17(4):262-267 [FREE Full text] [doi: 10.1089/cyber.2013.0253] [Medline: 24555521]
79. Papini S, Young CC, Gebhardt CS, Perrone A, Morikawa H, Otto MW, et al. Isradipine enhancement of virtual reality cue exposure for smoking cessation: rationale and study protocol for a double-blind randomized controlled trial. Contemp Clin Trials 2020 Jul;94:106013 [FREE Full text] [doi: 10.1016/j.cct.2020.106013] [Medline: 32335287]

80. Goldenhersch E, Thrul J, Ungaretti J, Rosencovich N, Waitman C, Ceberio MR. Virtual reality smartphone-based intervention for smoking cessation: pilot randomized controlled trial on initial clinical efficacy and adherence. J Med Internet Res 2020 Jul 29;22(7):e17571 [FREE Full text] [doi: 10.2196/17571] [Medline: 32723722]
81. Eiler T, Grünewald A, Brück R. Fighting substance dependency combining AAT therapy and virtual reality with game design elements. In: Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - HUCAPP. 2019 Presented at: 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - HUCAPP; 2019; Prague, Czech Republic p. 28-37 URL: https://www.scitepress.org/Link.aspx?doi=10.5220/0007362100280037 [doi: 10.5220/0007362100280037]
82. Eiler T, Grünewald A, Wahl M, Brück R. AAT meets virtual reality. In: Computer Vision, Imaging and Computer Graphics Theory and Applications. VISIGRAPP 2019. Communications in Computer and Information Science, vol 1182. Cham: Springer; 2020:153-176.
83. Girard B, Turcotte V, Bouchard S, Girard B. Crushing virtual cigarettes reduces tobacco addiction and treatment discontinuation. Cyberpsychol Behav 2009 Oct;12(5):477-483. [doi: 10.1089/cpb.2009.0118] [Medline: 19817561]
84. Girard B, Turcotte V. A virtual arm to stop smoking, a pilot perceptual learning experiment. Annual Review of Cybertherapy and Telemedicine. 2007. URL: https://www.arctt.info/volume-5-summer-2007 [accessed 2021-08-02]
85. Blevins D, Khanna S, Ait-Daoud N. A pilot study for video game therapy for substance cessation and recovery for nicotine dependence. Am J Addict 2017:278-279. [doi: 10.1111/ajad.12545] [Medline: 28370765]
86. Bordnick PS, Traylor AC, Carter BL, Graap KM. A feasibility study of virtual reality-based coping skills training for nicotine dependence. Res Soc Work Pract 2012 May 08;22(3):293-300 [FREE Full text] [doi: 10.1177/1049731511426880] [Medline: 25484549]
87. Pericot-Valverde I, Rodríguez OG, Ferrer-García M, Villa RS, Gutiérrez-Maldonado J. Virtual reality for smoking cessation: a case report. Annu Rev Cyber Therapy Telemed 2012 Jun;10:292-296 [FREE Full text] [doi: 10.1037/t04044-000]
88. Caponnetto P, Maglia M, Lombardo D, Demma S, Polosa R. The role of virtual reality intervention on young adult smokers' motivation to quit smoking: a feasibility and pilot study. J Addict Dis 2018 Oct 15;37(3-4):217-226. [doi: 10.1080/10550887.2019.1664364] [Medline: 31613186]
89. Batra A. Benefits of a cognitive behavioral therapy intervention for smoking cessation supported by Virtual Reality Smoking Cue Exposure (ViReTa). Protocol - NCT03707106. 2018. URL: https://clinicaltrials.gov/ct2/show/NCT03707106 [accessed 2021-08-02]
90. Pericot-Valverde I, García-Rodríguez O, Gutiérrez-Maldonado J, Secades-Villa R. Individual variables related to craving reduction in cue exposure treatment. Addict Behav 2015 Oct;49:59-63. [doi: 10.1016/j.addbeh.2015.05.013] [Medline: 26047836]
91. Gamito P, Oliveira J, Baptista A, Morais D, Lopes P, Rosa P, et al. Eliciting nicotine craving with virtual smoking cues. Cyberpsychol Behav Soc Netw 2014 Aug;17(8):556-561. [doi: 10.1089/cyber.2013.0329]
92. Carter BL, Bordnick P, Traylor A, Day SX, Paris M. Location and longing: the nicotine craving experience in virtual reality. Drug Alcohol Depend 2008 May 01;95(1-2):73-80 [FREE Full text] [doi: 10.1016/j.drugalcdep.2007.12.010] [Medline: 18243586]
93. Baumann SB, Sayette MA. Smoking cues in a virtual world provoke craving in cigarette smokers. Psychol Addict Behav 2006;20(4):484-489. [doi: 10.1037/0893-164x.20.4.484]
94. Gamito P, Oliveira J, Baptista A, Pereira E, Morais D, Saraiva T, et al. Virtual reality exposure on nicotine craving. Stud Health Technol Inform 2011;167:63-68. [doi: 10.3233/978-1-60750-766-6-63] [Medline: 21685643]
95. Gamito P, Oliveira J, Baptista A, Morais D, Lopes P, Saraiva T, et al. Nicotine craving: ERPs correlates after VR exposure to smoking cues. Annu Rev CyberTherapy Telemed 2012;10:78-82 [FREE Full text] [doi: 10.3233/978-1-61499-121-2-78]
96. Keijsers M. Systematic literature review searches. Centre for Open Science Storage. 2020 Jul 17. URL: https://osf.io/f5qvy/ [accessed 2020-11-06]
97. Raw M, Russell M. Rapid smoking, cue exposure and support in the modification of smoking. Behav Res Ther 1980;18(5):363-372. [doi: 10.1016/0005-7967(80)90001-7]
98. Niaura R, Abrams D, Shadel W, Rohsenow D, Monti P, Sirota A. Cue exposure treatment for smoking relapse prevention: a controlled clinical trial. Addiction 1999 May;94(5):685-695. [doi: 10.1046/j.1360-0443.1999.9456856.x] [Medline: 10563033]
99. Belita E, Sidani S. Attrition in Smoking Cessation Intervention Studies: A Systematic Review. Can J Nurs Res 2015 Dec 11;47(4):21-40 [FREE Full text] [Medline: 29509476]
100. Curtin L, Brown RA, Sales SD. Determinants of attrition from cessation treatment in smokers with a history of major depressive disorder. Psychol Addict Behav 2000;14(2):134-142. [doi: 10.1037/0893-164x.14.2.134]
101. Mellentin AI, Skøt L, Nielsen B, Schippers GM, Nielsen AS, Stenager E, et al. Cue exposure therapy for the treatment of alcohol use disorders: a meta-analytic review. Clin Psychol Rev 2017 Nov;57:195-207 [FREE Full text] [doi: 10.1016/j.cpr.2017.07.006] [Medline: 28781153]
102. Thewissen R, Snijders SJ, Havermans RC, van den Hout M, Jansen A. Renewal of cue-elicited urge to smoke: implications for cue exposure treatment. Behav Res Ther 2006 Oct;44(10):1441-1449. [doi: 10.1016/j.brat.2005.10.010] [Medline: 16375853]

103. Ryan RM, Rigby CS, Przybylski A. The motivational pull of video games: a self-determination theory approach. Motiv Emot 2006 Nov 29;30(4):344-360. [doi: 10.1007/s11031-006-9051-8]
104. Bahadir A, Iliaz S, Yurt S, Ortakoylu MG, Bakan ND, Yazar E. Factors affecting dropout in the smoking cessation outpatient clinic. Chron Respir Dis 2016 Feb 04;13(2):155-161. [doi: 10.1177/1479972316629953] [Medline: 26846679]
105. Raupach T, Brown J, Herbec A, Brose L, West R. A systematic review of studies assessing the association between adherence to smoking cessation medication and treatment success. Addiction 2014 Jan;109(1):35-43. [doi: 10.1111/add.12319] [Medline: 23919621]
106. Cohen J. A power primer. Psychol Bull 1992;112(1):155-159. [doi: 10.1037/0033-2909.112.1.155]
107. Podsakoff PM, MacKenzie SB, Podsakoff NP. Sources of method bias in social science research and recommendations on how to control it. Annu Rev Psychol 2012 Jan;63:539-569. [doi: 10.1146/annurev-psych-120710-100452] [Medline: 21838546]
108. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. Lancet 1991 Apr 13;337(8746):867-872. [doi: 10.1016/0140-6736(91)90201-y] [Medline: 1672966]
109. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS One 2008 Aug 28;3(8):e3081 [FREE Full text] [doi: 10.1371/journal.pone.0003081] [Medline: 18769481]

Abbreviations

AAT: approach-avoidance task
AR: augmented reality
CBT: cognitive behavioral therapy
CE: cue exposure
CET: cue exposure therapy
CO: carbon monoxide
ENDS: Electronic Nicotine Delivery Systems
NRT: nicotine replacement therapy
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
VR: virtual reality
VR-CET: cue exposure therapy in virtual reality
VRT: virtual reality technology

Edited by R Kukafka; submitted 14.09.20; peer-reviewed by S Bhattacharjya, K Brandon; comments to author 31.10.20; revised version received 22.01.21; accepted 21.06.21; published 17.09.21.

Please cite as:
Keijsers M, Vega-Corredor MC, Tomintz M, Hoermann S
Virtual Reality Technology Use in Cigarette Craving and Smoking Interventions (I "Virtually" Quit): Systematic Review
J Med Internet Res 2021;23(9):e24307
URL: https://www.jmir.org/2021/9/e24307
doi:10.2196/24307
PMID:34533471

Original Paper

The Effect of Reminiscence Therapy Using Virtual Reality on Apathy in Residential Aged Care: Multisite Nonrandomized Controlled Trial

Dimitrios Saredakis[1], BPsych (Hons); Hannah AD Keage[1], PhD; Megan Corlis[2], RN, BAppSc; Erica S Ghezzi[1], BPsych (Hons); Helen Loffler[3], RN; Tobias Loetscher[1], PhD

[1]UniSA Justice & Society, University of South Australia, Adelaide, South Australia, Australia

[2]UniSA Clinical & Health Sciences, University of South Australia, Adelaide, South Australia, Australia

[3]Helping Hand Aged Care, Adelaide, South Australia, Australia

Corresponding Author:
Dimitrios Saredakis, BPsych (Hons)
UniSA Justice & Society
University of South Australia
St Bernards Road, Magill
Adelaide, South Australia, 5072
Australia
Phone: 61 8 8302 4083
Email: dimitrios.saredakis@mymail.unisa.edu.au

Abstract

Background: Apathy is a frequent and underrecognized neurological disorder symptom. Reduced goal-directed behavior caused by apathy is associated with poor outcomes for older adults in residential aged care. Recommended nonpharmacological treatments include person-centered therapy using information and communication technology. Virtual reality (VR) in the form of head-mounted displays (HMDs) is a fully immersive technology that provides access to a wide range of freely available content. The use of VR as a therapy tool has demonstrated promise in the treatment of posttraumatic stress disorder and anxiety. In addition, VR has been used to improve conditions including depression, anxiety, cognitive function, and balance in older adults with memory deficits, Alzheimer disease, and Parkinson disease. Research using VR for the symptoms of apathy in older adults living in residential aged care facilities is limited.

Objective: This study aims to examine whether using HMDs as a tool for reminiscence therapy improves the symptoms of apathy compared with using a laptop computer and physical items with older adults living in residential aged care.

Methods: In this multisite trial, 43 participants were allocated to one of three groups: reminiscence therapy intervention using VR in the form of HMDs, reminiscence therapy using a laptop computer supplemented by physical items if required (active control), and a usual care (passive control) group. The primary outcome was apathy, and the secondary outcomes included cognition and depression. The side effects of using HMDs were also measured in the VR group.

Results: Mixed model analyses revealed no significant group interaction over time in outcomes between the VR and laptop groups (estimate=−2.24, SE 1.89; t_{40}=−1.18; P=.24). Pooled apathy scores in the two intervention groups compared with the passive control group also revealed no significant group interaction over time (estimate=−0.26, SE 1.66; t_{40}=−0.16; P=.88). There were no significant secondary outcomes. Most participants in the VR group stated that they would prefer to watch content in VR than on a flat screen (X^2_2=11.2; P=.004), side effects from HMD use were negligible to minimal according to the Simulator Sickness Questionnaire cutoff scores.

Conclusions: Although there were no significant results in outcome measures, this study found that participants engaged in the research and enjoyed the process of reminiscing using both forms of technology. It was found that VR can be implemented in an aged care setting with correct protocols in place. Providing residents in aged care with a choice of technology may assist in increasing participation in activities. We cannot dismiss the importance of immediate effects while the therapy was in progress, and this is an avenue for future research.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12619001510134; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=378564.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2020-046030

(J Med Internet Res 2021;23(9):e29210) doi:10.2196/29210

KEYWORDS

reminiscence; head-mounted display; apathy; cognitive aging; dementia; residential facilities; virtual reality

Introduction

Background

Apathy

Apathy is a neurological disorder with a high prevalence found in older adults living in residential aged care facilities [1,2]. Characterized as a lack of interest and diminished motivation [3], the presence of apathy results in reduced goal-directed behavior [4,5], which can result in withdrawal from activities. In long-term residential aged care facilities, lower engagement in activities and social interaction is associated with poorer quality of life [6].

Prognostic features of apathy include accelerated cognitive decline [7], and in aged care, apathy has been associated with an increased risk of mortality [8]. Although apathy is not as disruptive as other neuropsychiatric symptoms, including aggression and agitation, it can be a predictor of caregiver burden [9]. There is a lack of convincing evidence for pharmacological interventions for apathy [10-12]. Therefore, nonpharmacological individualized interventions are recommended [13,14].

Reminiscence Therapy

One such approach that can be individualized and tailored to a person is reminiscence therapy [15]. Commonly used in residential aged care facilities, reminiscence is a strength-based, person-centered approach that is recommended for consideration in clinical practice when caring for people with dementia [16]. The process of reminiscence involves recalling memories through the experience and discussion of past events and is commonly done with the assistance of familiar items [17]. The focus is on early memories. Reminiscence approaches can be classified into three types [18]. Simple reminiscence involves recalling positive memories to increase positive feelings [19]. Life review aims to consolidate both positive and negative previous life events into a meaningful life story, and finally, life review therapy involves redefining the negative interpretations of a person's past [19]. It is possible that the recall of both positive and negative memories may cause distress [20,21]; however, no reported harm has been found in a review examining reminiscence therapy [17]. In this study, a semistructured simple reminiscence approach was used, focusing on positive memories.

Positive outcomes of reminiscence found in people with Alzheimer disease include improvements in cognition, depression, and quality of life [22]. Results from a meta-analysis have reported medium effects for depression (g=0.57, 95% CI 0.44 to 0.70) and small effects for overall mental health symptoms (g=0.33, 95% CI 0.16 to 0.51) [18]. The retrieval of autobiographical memories through reminiscence also provides the social function of sharing memories with other people [23].

Autobiographical memories are an important aspect of the sense of self [24], which can be impaired in those with apathy [25,26]. In aged care, the sharing of memories or life stories can positively influence the sense of self for people with dementia [27]. Therefore, reliving autobiographical memories may translate into improvements in apathy. Although reminiscence can be performed in a group or individually, it has been recommended that individual approaches may provide increased benefits [28]. An individual approach allows sessions to be tailored specifically to a person by focusing on appropriate and relevant memories depending on the person's background. Apathy can cause a lack of interest in participation in traditional forms of activities or therapy and places the person at risk of loneliness, a prevalent problem in aged care [29]. Therefore, alternative solutions for treating apathy are required. Technology provides easy access to content and assists in maintaining engagement in therapy [15,30]. The use of immersive technology may increase therapeutic outcomes.

Virtual Reality

Technology can provide different levels of immersion, for example, computers or tablets are described as nonimmersive, and large screen televisions or projectors are semi-immersive [31]. An example of highly immersive virtual reality (VR) technology is the cave automatic virtual environment, where the image is displayed on the walls, and in some cases, on the floor and ceiling; however, this system is expensive, takes time to set up, and requires a dedicated room [32]. Another example of highly immersive VR technology is the head-mounted display (HMD); modern HMDs consist of a monitor for each eye to provide a stereoscopic image via a device worn on the head [33]. Available applications for HMDs can be passive, for example, providing a 360° view of a scenic virtual environment that changes as the viewer turns their head [34]. Applications can also allow users to pick up objects or physically move in a virtual environment, providing an interactive experience [34].

The successful use of HMDs as a tool for interventions includes exposure and distraction-based therapies [35,36]. Being a fully immersive technology, VR using HMDs eliminates any distractions and can increase the sense of presence, the psychological feeling of being in the virtual environment [37]. A high sense of presence may increase the response to the virtual environment [38]. For example, a study consisting of a sample of young adults found that autobiographical memory improved when content was viewed in VR using HMDs compared with a flat screen monitor [39]. Similar results of improved memory performance were found in participants who used HMDs compared with a desktop display in a sample of participants from a university campus and community [40]. Using HMDs

may provide increased stimulation in the process of reliving autobiographical memories in older clinical populations because of the use of a fully immersive and realistic environment. Whether this translates to improved outcomes for older adults living in residential aged care is uncertain.

Two recent reviews have reported that VR HMD applications demonstrated potential for rehabilitation in neurological diseases, including dementia, Parkinson disease, multiple sclerosis, and stroke, or of those with pain and memory deficits [41,42]. Positive outcomes, ranging from small to large effects, found in the review by Dermody et al [42], included improved pain scores in community-dwelling older adults using VR games (d=−1.54) [43], and improvements in long-term recall (d=0.70) using VR memory training in aged care residents [44]. However, no significant outcomes were reported in psychological health, activities of daily living, depression, and social life in the studies included in this review [43,45]. A recent study of 236 older adults from community centers used familiar VR content to stimulate memory. This study found significant increases in positive affect scores (mean difference 2.09) and decreases in negative affect scores (mean difference −1.99) [46]. There is evidence that using VR with HMDs with older adults can be successful [42]; however, there is a lack of comparison between VR and flat screen technology to establish the differences and advantages of using VR.

There is a paucity of evidence comparing nonimmersive or flat screen technology with fully immersive technology (HMDs) in a therapeutic context in clinical older populations. A single-session study in aged care residents compared viewing Google Maps Street View using a tablet and HMD [47]. This study found that both forms of technology provided emotional experiences for older adults. In addition, none of the 7 participants preferred using HMD over the tablet. However, an exercise study in aged care residents reported an increase in interest and enjoyment when using HMDs compared with when watching a flat screen [48]. In another study, greater improvements in physical well-being were seen in older adults in an assisting living community when viewing content related to their past, travel, and relaxation using VR as compared with television [49].

Exposure to virtual environments can cause side effects or after effects, including eyestrain, nausea, and headaches [50,51]. This can have health and safety implications when using HMDs. Therefore, there is a need to increase our understanding of the side effects of HMD use in older adults and clinical populations who are taking medication and have symptoms related to their condition that may exacerbate any VR side effects.

Objectives

The primary aim of this study is to examine whether using HMDs as a tool for reminiscence therapy improved outcomes compared with using a laptop computer supplemented by physical items if required (active control group). The primary outcome was apathy. Secondary outcomes included cognition and depression. Exploratory outcomes included loneliness and quality of life. The side effects of using HMDs were also examined in the VR group. We hypothesize the following: (1) the VR group would have lower apathy scores than the active control group using a laptop computer and physical items after the intervention and (2) both the VR group and active control combined would have lower apathy scores than the passive control (usual care) group after the intervention.

Methods

Study Design

This was a multisite nonrandomized controlled trial. Data were collected across 3 residential aged care facilities in Adelaide, South Australia. Each site was allocated a group determined by the aged care provider. Assessors conducting baseline and follow-up measures were blinded to group allocation. Participants in the intervention group were blinded to the presence of other conditions. The study was performed between December 2019 and February 2021. The CONSORT (Consolidated Standards of Reporting Trials) 2010 statement guidelines were followed [52]. A summary of the study methods is provided below, and further details of the study methods are available in the published study protocol [53].

Participants

Eligible participants were aged 65 years or older and included only those with up to moderate impairment according to the Psychogeriatric Assessment Scale [54] as assessed by the aged care facility. Participants had to be English speaking and willing to undertake follow-up assessments. Participants in the VR group had to be able to tolerate wearing an HMD and have vision that can be corrected using their eyeglasses. The eyeglass frame also needed to fit the HMD. Participants were excluded if their score on the Psychogeriatric Assessment Scale was higher than 15 or if they had significant disorders, conditions, or behaviors that would make assessment difficult. Those with confusion or disorientation issues and who might become distressed because of confusion regarding time and place were also excluded. Participants who were interested and met the eligibility criteria were provided written and verbal information about the study. A dedicated research nurse employed by the residential aged care facility or primary researcher obtained informed consent. All participants were given the opportunity to discuss their participation with family members or other responsible persons close to the participant. Consent was continually monitored during the research by asking participants at the start and end of each session if they wanted to continue. Ethics approval was obtained from the University of South Australia Human Research Ethics Committee. We refer to the published protocol for sample size calculation [53].

Materials

The terminology VR is an umbrella term that can refer to both nonimmersive and immersive technologies [55]; for the purposes of this study, VR will refer to immersive HMDs.

VR Software

YouTube VR (developed by Google LLC), a VR version of YouTube, was used to view personalized videos in VR. Wander (developed by Parkline Interactive), which uses data from Google Street View, was used to view personalized places of interest. The active control group used a laptop for viewing

content; therefore, Google Street View, YouTube, and the internet in general were used to view content on the laptop.

VR Hardware

The Oculus Quest [56] HMD, a commercially available stand-alone headset, was used for the VR group. This HMD provides stereoscopic vision at a resolution of 1440 × 1600 per eye with a 72 Hz refresh rate and provides access to the two VR software applications used.

Primary Outcome: Apathy Evaluation Scale Clinician Version

The Apathy Evaluation Scale (AES) clinician version, an 18-item scale, was used to measure apathy [3]. This scale also consists of an interview with the participant before completion to build rapport and gain an understanding of the participant. Each item is recorded with responses ranging from *not at all characteristic*, *slightly characteristic*, *somewhat characteristic*, and *a lot*. Four items were self-evaluated, and the remaining 14 items were assessed by the researcher. Five items assessed by the researcher need to be quantified, for example, a participant is required to give three or more examples to meet the criteria for *a lot*. Scores range from 18 to 72, with higher scores indicating increased apathy. The clinician version of the AES has a test-retest reliability of .88 and good internal consistency (α=.90) [3].

Secondary Outcomes

Addenbrooke Cognitive Examination III

Addenbrooke Cognitive Examination III (ACE-III) was used to assess cognitive ability [57]. ACE-III comprises subtests of attention, memory, fluency, language, and visuospatial functioning. Its maximum score is 100, with higher scores indicating higher cognitive functioning. ACE-III reported good internal consistency (α=.88) [58].

Geriatric Depression Scale Short Form

The Geriatric Depression Scale (GDS) was used to measure depression [59]. This scale was developed for older populations and is suitable for those with cognitive impairment and living in long-term institutional environments [60]. The GDS has 15 items that a participant needs to respond to with a *yes* or *no* answer. Scores range from 0 to 15. A score of >5 suggests depression, whereas a score of ≥ 10 indicates depression. The GDS has 92% sensitivity and 89% specificity when compared with diagnostic criteria [60], and good internal consistency has been demonstrated (α=.80) [61].

Exploratory Outcomes

The Quality of Life in Alzheimer Disease

Quality of life was measured using the Quality of Life in Alzheimer Disease (QOL-AD) 13-item scale [62]. Responses to items range from *poor*, *fair*, and *good* to *excellent*, with a score ranging from 13 to 52. Higher scores indicate a higher quality of life. Good internal consistency for the QOL-AD scale has been reported (α=.82) [63].

Three-Item Loneliness Scale

The Three-Item Loneliness Scale, a shortened version of the Revised University of California, Los Angeles Loneliness Scale [64], was used to assess the participants' level of loneliness. Three questions were rated on a three-point scale, with scores ranging from 3 to 9. Responses range from *hardly ever* to *some of the time* to *often*. Higher scores indicated higher levels of loneliness. Acceptable internal consistency has been reported for the Three-Item Loneliness Scale (α=.72) [64].

Simulator Sickness Questionnaire

The Simulator Sickness Questionnaire (SSQ) was used to measure the side effects of HMD for the VR group [50]. The SSQ comprises 16 items on a four-point scale, ranging from *none*, *slight*, and *moderate*, to *severe*. The three subscales measure symptoms related to nausea, oculomotor, and disorientation. Higher scores indicate higher symptoms of sickness. The SSQ is the most commonly used questionnaire in VR research [65].

Staff Questionnaire

The Staff Questionnaire was developed by the primary researcher to measure improvement or deterioration in participants from the staff members' point of view. Responses were reported on a five-point scale ranging from *not at all* to *very much so*. Domains measured included social involvement, cognitive awareness, pain, activities of daily living, behavior, and communication. The same domains were assessed for improvement and deterioration. Higher scores on the improvement questions meant more improvement, whereas higher scores on the deterioration questions meant more deterioration.

Session Record

To measure participant attendance and responses to the reminiscence sessions, a Bender session record [66] was completed for both the VR and active control groups after each reminiscence session by the researchers delivering the reminiscence content. This measure examined five subscales including *Attendance of session*, *Memory*, *Interaction*, *Responsiveness*, and *Enjoyment*. Each subscale was measured on a four-point scale ranging from 0 to 3. In addition, for both groups, participants were asked, "Would you like to do it again?" In the VR group, participants were additionally asked, "If given a choice, would you prefer to view content in VR or on a flat screen?" This was compared with the participants' previous experience of viewing television or large screen displays.

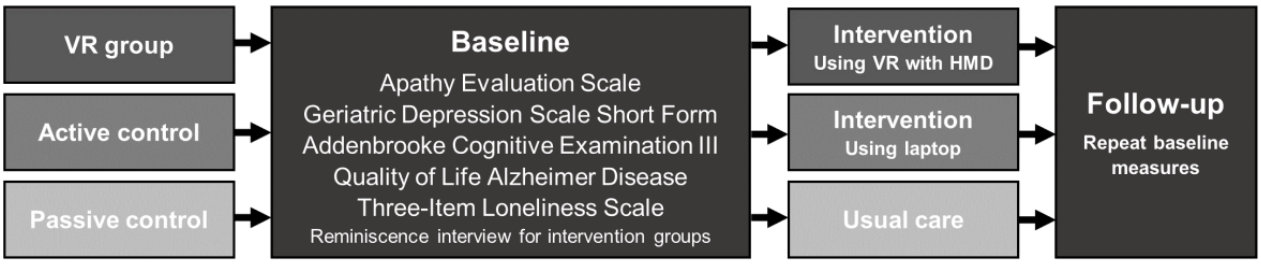
Procedure

Recruitment

Participants were selected for recruitment by senior staff at the residential aged care facility according to the inclusion and exclusion criteria. Recruitment was undertaken by a dedicated research nurse employed by the aged care facility or the lead researcher. Participants who were interested in participating in the VR group were given a demonstration of the HMD to ensure that they could tolerate wearing it and could view images to a

satisfactory standard before consent. The study flowchart is shown in Figure 1.

Figure 1. Study flow chart of procedure. HMD: head-mounted display; VR: virtual reality.



Baseline

During the baseline session, the AES, ACE-III, GDS, QOL-AD scale, and Three-Item Loneliness Scale were completed for all three groups by trained research assistants (blinded to group allocation). This process took approximately 60 minutes. For both intervention groups, a qualitative interview was conducted by a different trained researcher (unblinded to group allocation) to establish topics for the reminiscence sessions after the baseline measures were completed. This process took approximately 60 minutes. Therefore, the baseline sessions for the two intervention groups took approximately two hours. The qualitative interview was semistructured and covered themes throughout the lifespan of each participant. This included early childhood memories, adolescence, young adulthood, older adulthood, most recent memories, and musical memories. The focus was on positive experiences using reminiscence therapy guidelines [15,66].

Intervention

The VR and active control groups undertook three individual reminiscence sessions that were completed within a period of approximately 2 weeks at least 1 day apart. The reminiscence component was timed for 20 minutes. Additional time was taken to introduce and summarize the sessions. SSQ was completed before and after reminiscence sessions 1 and 3 for the VR group, adding approximately 10 minutes. Sessions were run by the lead researcher or a trained research assistant who was not involved in taking outcome measures. During the reminiscence sessions, there was continual conversation between the researcher and the participant regarding the content being viewed. Participants in the active control group viewed the reminiscence content on a laptop computer. The Oculus Quest HMD was used in the VR group. During the VR reminiscence experience, participants were seated in a swivel chair with arm rests when possible. If they were unable to sit in a chair, they remained in their bed; therefore, this was a passive VR experience. Content relating to the participant's background gained from the reminiscence interview was viewed using 360° videos from YouTube VR and street view content from the Wander app. All interactions were performed by the researcher, and sound was provided by the integrated speakers of the Oculus Quest HMD. Participants were closely monitored and asked about any symptoms or side effects of viewing content in the HMD during the sessions. Content from the HMD was mirrored onto a laptop computer to enable the researcher to select content for the participant and view the content simultaneously. The passive control group did not receive the intervention and continued with their usual care during the 2 weeks.

Objective Measures

This study collected data on objective measures that will not be reported here. Heart rate variability, galvanic skin response, and speech measures were collected for both intervention groups during the intervention sessions, and approximately 30 minutes were added to the sessions. Activity was measured 48 hours before baseline and 48 hours after follow-up for all three groups [53].

Follow-up

For both intervention groups, follow-up measures were completed the day after the final reminiscence session. The average time between baseline and follow-up for participants in all three groups was approximately 2 weeks plus 4 days for the measuring of activity, as described above. For 2 participants in the passive control group, restricted site access because of COVID-19 increased the time between baseline and follow-up sessions to almost 4 weeks. None of the participants were restricted from participating in lifestyle activities offered by the aged care facility during the research period. Lifestyle activities were reduced because of COVID-19 restrictions.

Statistical Methods

Data were entered using Research Electronic Data Capture [67] hosted by the University of South Australia. The analysis was performed using Jamovi (version 1.6.15) [68]. The primary outcome figure was created in R Studio (version 3.6.3) [69] using open visualizations [70]. The primary outcomes at baseline and follow-up were analyzed using linear mixed modeling with Helmert contrasts. This consisted of fixed factors of group (VR, active control, and passive control) and time (baseline and follow-up), with the intercept of the participant as a random factor. For comparisons of the two intervention groups with the usual care group, Helmert contrasts were used where the contrasts compared (1) combined intervention groups with the usual care group and (2) the two interventions. Additional analyses were performed by including only participants who met the criteria for a diagnosis of apathy at baseline using a cutoff score of 37.5 [3,71]. The influence of potential covariates, including depression and cognition at baseline, was examined. ACE-III, GDS, QOL-AD scale, and the Three-Item Loneliness Scale were analyzed as per the primary outcome. Assumptions of normality using the Shapiro-Wilk test found SSQ total and subscale scores to significantly deviate from a normal

distribution; therefore, Wilcoxon signed-rank tests were performed for all SSQ comparisons. Assumptions of normality using the Shapiro-Wilk test found session record total and subscale scores, except for memory, to significantly deviate from a normal distribution; therefore, Mann-Whitney U tests were performed for all session record comparisons. No adjustments for multiple comparisons were made for all planned analyses; therefore, significance was set at <.05. Effect sizes for linear mixed modeling were calculated using partial η^2 calculated from the *t* statistic and sample size with the following cutoffs: small=0.01, medium=0.06, and large=0.14.

Results

Demographic and Baseline Scores

A total of 74 participants were identified and approached to participate in the study, of whom 46 consented (Figure 2). In total, 28 participants declined because they were not interested in participating or for personal reasons. A total of 15 participants were initially allocated to each of the three conditions; 3 participants withdrew from the study. One participant in the VR group was admitted to the hospital after completion of the baseline measures and was unable to continue upon their return. A second participant in the active control withdrew after the baseline measures, as they did not wish to continue. A final participant in the passive control withdrew when follow-up measures began, as they did not wish to continue. As this was a per-protocol analysis, the participant in the VR group was replaced with another participant; however, COVID-19 restrictions prevented the replacement of participants in the other two conditions because of site access restrictions during the research period.

The background characteristics and baseline results of the outcome measures of participants who completed the study are reported in Table 1. Using a cutoff score of 37.5 for the AES [3,71], 65% (28/43) of participants met the criteria for a diagnosis of apathy at baseline. Of the remaining participants, 7% (3/43) scored 37, 12% (5/43) scored between 33 and 36, and 16% (7/43) scored between 22 and 28.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of study enrollment and analyses (modified for nonrandomized study).

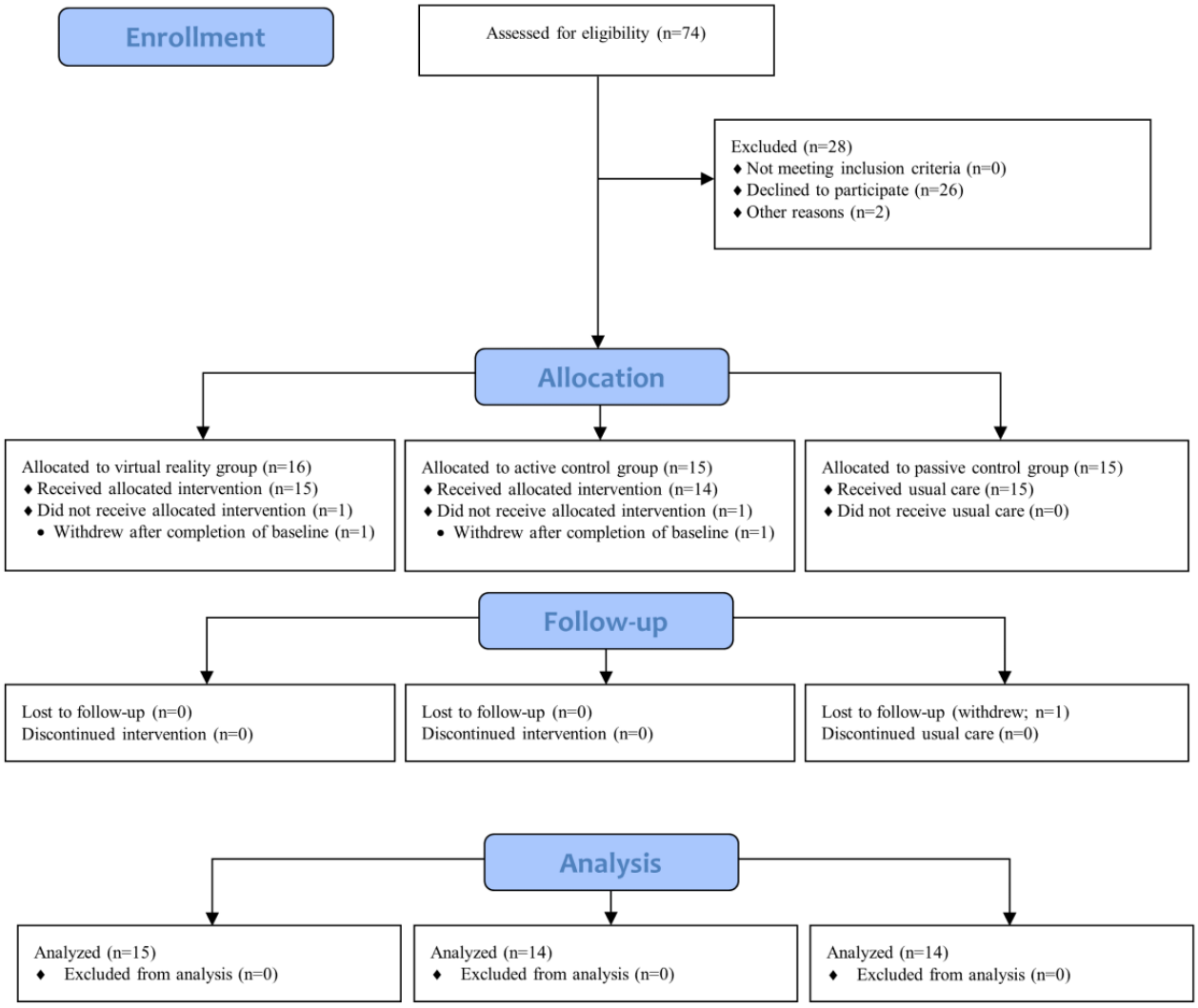


Table 1. Participant demographic and clinical characteristics by study group at baseline (N=43).

Participant characteristics	Virtual reality (n=15)	Active control (n=14)	Passive control (n=14)	Total (n=43)
Age (years), mean (SD; range)	81.7 (6.6; 72-93)	85.9 (8.1; 71-95)	87 (8.7; 73-103)	84.8 (8.0; 71-103)
Sex (female), n (%)	10 (67)	9 (64)	9 (64)	28 (65)
Years of education, mean (SD)	11.7 (3.1)	10.7 (2.6)	11.0 (4.6)	11.2 (3.4)
Marital status, n (%)				
Married	3 (20)	2 (14)	5 (36)	10 (23)
Divorced or widowed	11 (73)	11 (79)	8 (57)	30 (70)
Single	1 (7)	1 (7)	1 (7)	3 (7)
Depression and anxiety medication, n (%)	5 (33)	3 (21)	8 (57)	16 (37)
Primary diagnosis, n (%)				
Memory-related, dementia, or Parkinson disease	5 (33)	3 (21)	3 (21)	11 (26)
Heart disease	8 (53)	4 (29)	8 (57)	20 (47)
Stroke	2 (13)	2 (14)	2 (14)	6 (14)
Other	0 (0)	5 (36)	1 (7)	6 (14)
Outcome measures at baseline, mean (SD)				
Apathy Evaluation Scale	35.3 (8.7)	41.8 (7.1)	44.3 (9.5)	40.3 (9.1)
Geriatric Depression Scale	4.1 (3.6)	3.5 (2.5)	4.4 (3.0)	4.0 (3.0)
ACE-III[a]	72.9 (17.7)	71.2 (13.7)	72.2 (13.6)	72.1 (14.9)
QOL-AD[b]	35.4 (5.8)	36.1 (5.2)	32.9 (6.0)	34.8 (5.7)
Three-Item Loneliness Scale	4.5 (1.5)	4.2 (1.4)	5.0 (2.5)	4.6 (1.9)

[a]ACE-III: Addenbrooke Cognitive Examination III.

[b]QOL-AD: Quality of Life in Alzheimer Disease.

Primary Outcome

For contrast 1 (pooled VR and active control groups compared with passive control), there was a significant main effect of group, with higher apathy scores in the passive control group. For contrast 2 (VR group compared with active control), the main effect of group was not significant (Table 2).

Table 2. Fixed effects parameter estimates for the Apathy Evaluation Scale.

Parameter	Estimate (SE; 95% CI)	*t* statistic (*df*)	*P* value	Partial η^2
Main effect				
Time	−0.54 (0.78; −2.06 to 0.98)	−0.69 (40)	.49	0.012
Contrast 1[a]: passive versus (active and VR[b])	5.60 (2.55; 0.61 to 10.59)	2.20 (40)	*.03*[c]	0.108
Contrast 2[d]: active versus VR	5.33 (2.91; −0.36 to 11.03)	1.84 (40)	.07	0.078
Interaction				
Contrast 1: time×passive versus (active and VR)	−0.26 (1.66; −3.51 to 2.99)	−0.16 (40)	.88	0.001
Contrast 2: time×active versus VR	−2.24 (1.89; −5.95 to 1.47)	−1.18 (40)	.24	0.034

[a]Contrast 1 compares the pooled interventions (virtual reality and active control) with the passive control group.

[b]VR: virtual reality.

[c]Italics indicates significant values.

[d]Contrast 2 compares both intervention groups (virtual reality and active control).

Contrast 1 (pooled VR and active control groups compared with passive control) revealed that an intervention did not significantly change AES scores from baseline to follow-up (ie, the interaction between group and time was not significant). Contrast 2 (VR group compared with active control) was also not significant (Table 2). The addition of covariates (cognition and depression) did not change the statistical significance.

When restricting the sample to the subgroup that met the AES cutoff of 37.5, the same pattern of results was found (Table 3). The sample size was reduced to 28 (7 in the VR group, 10 in the active control group, and 11 in the passive control group).

Figure 3 presents the baseline and follow-up AES scores for the participants in each group.

Table 3. Fixed effects parameter estimates for the Apathy Evaluation Scale for subgroup meeting the Apathy Evaluation Scale cutoff of 37.5 at baseline.

Parameter	Estimate (SE; 95% CI)	*t* statistic (*df*)	*P* value	Partial η^2
Main effect				
Time	−1.38 (1.01; −3.36 to 0.60)	−1.36 (25)	.19	0.069
Contrast 1[a]: passive versus (active and VR[b])	5.12 (2.10; 1.01 to 9.23)	2.44 (25)	*.02*[c]	0.193
Contrast 2[d]: active versus VR	2.30 (2.65; −2.90 to 7.50)	0.87 (25)	.39	0.029
Interaction				
Contrast 1: time×passive versus (active and VR)	1.66 (2.04; −2.35 to 5.66)	0.81 (25)	.43	0.026
Contrast 2: time×active versus VR	−0.14 (2.58; −5.21 to 4.92)	−0.06 (25)	.96	<0.001

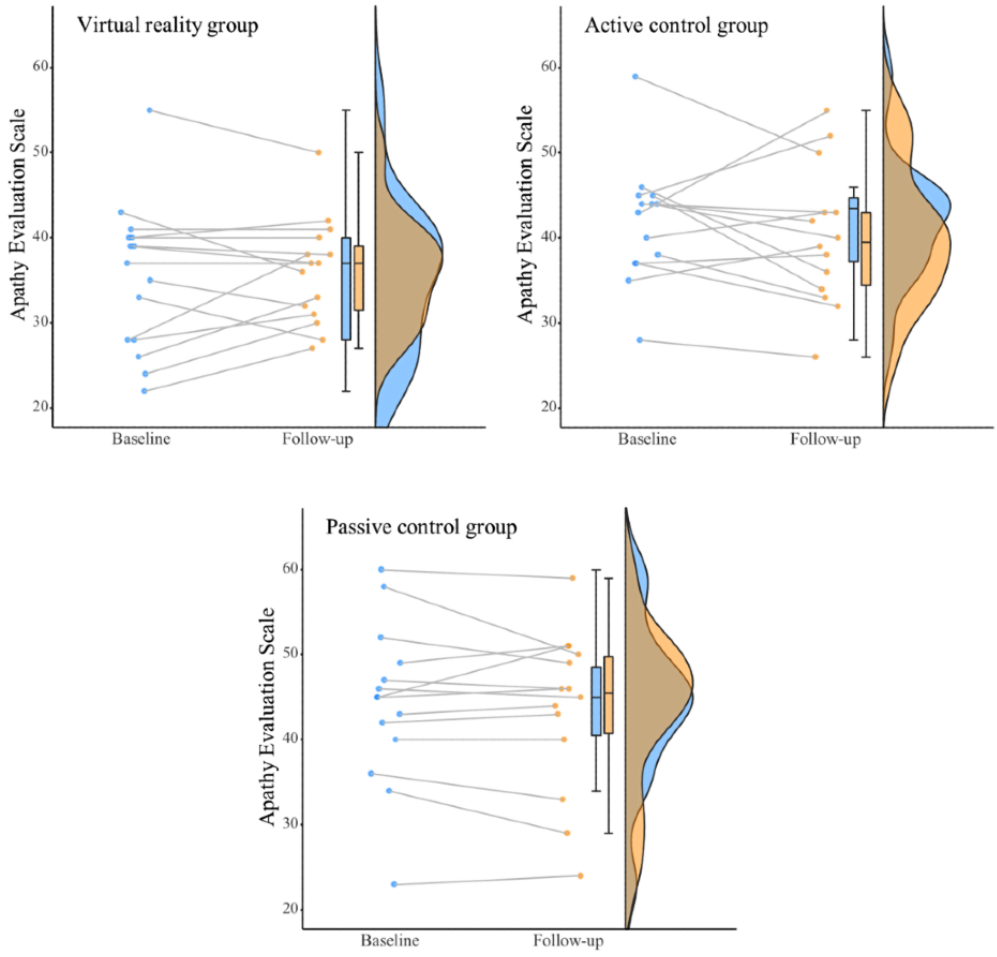
[a]Contrast 1 compares the pooled interventions (virtual reality and active control) with the passive control group.

[b]VR: virtual reality.

[c]Italics indicates significant values.

[d]Contrast 2 compares both intervention groups (virtual reality and active control).

Figure 3. Baseline and follow-up apathy scores for each group with each line representing one participant. Boxplots and raincloud indicate distribution in baseline and follow-up scores.



Secondary Outcomes

No significant results were observed for the ACE-III and GDS (Table S1 in Multimedia Appendix 1). Table S2 in Multimedia Appendix 1 presents results restricting the sample to the subgroup that met the AES cutoff of 37.5 at baseline.

Exploratory Outcomes

No significant results were observed for the QOL-AD scale and the Three-Item Loneliness Scale (Table S1 in Multimedia Appendix 1). Table S2 in Multimedia Appendix 1 presents results restricting the sample to the subgroup that met the AES cutoff of 37.5 at baseline.

SSQ Results

The average pre-VR SSQ scores and post-VR SSQ scores for sessions 1 and 3 were compared. The results were not statistically significant (Table 4). There were no dropouts because of side effects from the HMD use. Two participants reported back the after effects that occurred in the evening after their first morning VR session. This included a headache in one participant and a heavy-head feeling in another participant. The after effects were short-lived and did not cause significant discomfort.

Table 4. SSQ[a] means and SDs with statistics and effect size.

SSQ	Pre-SSQ, mean (SD)	Post-SSQ, mean (SD)	Statistic	*P* value	*d* value
Total	4.86 (6.11)	5.49 (5.98)	6.50	.46	−0.38
Nausea	5.72 (7.90)	6.36 (10.30)	5.50	.68	−0.27
Oculomotor	5.56 (7.14)	7.83 (9.11)	4.50	.25	−0.57
Disorientation	5.10 (10.01)	5.57 (8.40)	3.50	.71	−0.30

[a]SSQ: Simulator Sickness Questionnaire.

Staff Questionnaire

The Staff Questionnaire data were not included in the analysis. This was because of missing data for 12 participants caused by COVID-19 restrictions that did not allow follow-up with staff. Therefore, this would have made the results unreliable.

Session Record

The sum of the three sessions was used; therefore, the scores ranged from 0 to 9 for each subscale. The total score of the combined subscales ranged from 0 to 45 for all three sessions. Results comparing the VR group with active control for total and subscale scores were not statistically significant (Table 5). Results from the questions asked in the session record found that 73% (11/15) of participants in the VR group preferred viewing content in VR to a flat screen if given a choice, 7% (1/15) stated that they preferred a flat screen to VR, and 20% (3/15) were undecided. The proportion of participants who responded a preference for VR was significant (X^2_2=11.2; *P*=.004). Participants in both the VR and active control groups stated that they would like to do reminiscence again.

Table 5. Session record means and SDs for each group with statistics and effect size.

Session record	Virtual reality group, mean (SD)	Active control, mean (SD)	Statistic	*P* value	*d* value
Total	36.40 (7.21)	38.86 (5.02)	84.0	.37	0.20
Attendance	7.93 (1.44)	8.14 (1.61)	94.5	.60	0.10
Memory	5.60 (2.90)	7.36 (2.06)	66.0	.09	0.37
Interaction	8.20 (1.26)	8.50 (0.76)	99.5	.80	0.05
Responsiveness	7.53 (1.85)	8.57 (1.09)	72.0	.08	0.31
Enjoyment	7.20 (1.42)	6.29 (1.33)	68.5	.10	0.35

Discussion

Summary

To the best of our knowledge, this is the first clinical trial to compare the use of VR with flat screen technology for reminiscence in an aged care setting with a usual care control group. Well-established and validated measures were used, and there was a 100% adherence rate once the intervention commenced. It was also found that with correct procedures, side effects of using HMDs can be minimized and that VR technology can be implemented in an aged care setting. Participants also enjoyed the process of reminiscing indicated by the session record with all participants stating that they would like to do reminiscence again.

Content

It was found that the content for both intervention groups for all three sessions could be sourced using readily available apps, as per the feasibility study [72]. Common types of content the participants viewed included their original family home, the school they attended, places of employment, and travel destinations. In many instances, it was possible to tour inside buildings or places of interest and remote tourist destinations because of the increased access available in both the Wander app for the VR group and Google Street View for the active control group. Both YouTube and YouTube VR also provided access to travel destinations in addition to the music that the participant had memories of.

VR Group and Active Control Comparison

The results of this study are inconclusive, with no significant differences found between the intervention groups. Median scores for apathy decreased in the active control group and remained the same in the VR group from baseline to follow-up (Figure 3). Results from the session record for the VR group indicated that there was a preference for viewing content in VR than on a flat screen. In addition, a nonsignificant small effect size was found for enjoyment with higher scores in the VR group; however, similar small effect sizes with higher scores in the active control group were reported for responsiveness and memory. Anecdotally, there were more positive emotions observed in participants in the VR group when participants viewed familiar content compared with the active control group. The enjoyment or preference for VR did not translate to improved outcomes over flat screen technology. The findings from our study are consistent with those of a previous study [73] examining the use of VR by older adults with dementia, where no significant changes were found in outcome measures; however, VR did provide an engaging experience from the perspective of both participants and caregivers.

The availability of content when using the apps in the VR group (Wander and YouTube VR) was equivalent to that of the apps used for the active control group (Google Street View and YouTube); however, there were instances where the internet in general was used for the active control group, increasing the availability of content. In addition, in the active control, physical items were used in some instances (photo albums) if the participant requested this; however, the laptop was still always used, and the physical items only supplemented the reminiscence experience. This means that there were differences other than the display technology between the two intervention groups.

In both YouTube VR and Wander, participants had 360° vision by turning in the swivel chair and if the researcher used the controller to turn in the virtual environment. The Wander app provided increased movement backward and forward through the environment navigated by the researcher while participants remained stationary. The Oculus Quest HMD used in this study is capable of six degrees of freedom (head and body movement). As the participants were stationary and the apps used did not use the full capabilities of movement in the virtual environment, the participants were not provided with the optimum VR experience. Interactive VR with software using six degrees of freedom where the participant moves in the virtual environment has been found to increase the sense of presence and positive affect [34], which may not be an option for older adults in residential aged care because of safety aspects, including increased risk of falling. However, research using HMDs limited to three degrees of freedom (head movement only) for pain and posttraumatic stress disorder has reported positive outcomes [36,74].

Although VR provides increased immersion and realism, the interaction between the participant and researcher or therapist can be compromised by the HMD. Improving outcomes related to neuropsychiatric symptoms or quality of life may be influenced by interactions with the researcher or therapist. For example, reminiscence therapy traditionally involves attentive behaviors, including eye contact and body language [75]. Wearing the HMD during the interventions did not provide the same level of personal interaction as in the active control group. This may explain why VR in this study did not demonstrate the same positive outcomes seen in VR research using exposure and distraction-based therapies [35,36], where there is increased focus on the content being viewed rather than interaction with the therapist.

It is important not to discount the possibility that immediate effects provide enjoyment and stimulation during an intervention for participants that may not translate to longer-term measures [76]. Positive results related to apathy and emotion reported in previous studies using HMDs have been measured directly after a VR experience [46,72,77]. What has not been reported in this study is any immediate physiological effect when taking measures directly after the VR experience and how this compares with flat screen technology.

Effect of Reminiscence Therapy

There were no significant differences in outcomes from baseline to follow-up between the pooled intervention and passive control groups. These results contrast with the participant enthusiasm and enjoyment during the intervention in both groups evidenced by 100% participation once commencing, measured enjoyment levels from the session record, wanting to participate in reminiscence again, and anecdotal feedback from staff. Consistent with the meta-analysis by Pinquart and Forstmeier [18], a medium effect size was observed for depression, with reduced depression in the intervention groups over time, but this effect was not statistically significant (P=.09; partial η^2=0.069; Table S1 in Multimedia Appendix 1).

Compared with previous traditional reminiscence studies including apathy as an outcome, our study did have a relatively small dosage and was conducted over a short period [78,79], which may have contributed to the lack of statistically significant results. However, in an aged care setting, it may not be necessary to see improvement but a reduction or stabilization in the rate of decline of conditions, including apathy for the intervention to benefit participants [80]. Entry into aged care facilities results in accelerated cognitive decline compared with people living in the community [81]. It is unknown whether the intervention in this study was successful in altering the trajectory of decline from apathy to more severe symptoms that would require a longer intervention period. The number of sessions was selected to increase the feasibility of the trial, reduce attrition bias in an aged care context, and determine from immediate effects reported from our feasibility study [72]. Dose-effect models do have a nonlinear relationship [82], and previous studies have seen improvements in three sessions using reminiscence to improve agitation [83], verbal fluency and communication [84], increased engagement [85], self-esteem and depression [86,87], and anxiety [88]. As VR can provide increased stimulation compared with traditional therapies, the dosage required is yet to be established. There was also a significant difference between the groups at baseline, which may have also contributed to the lack of statistical differences.

The pooled scores of the intervention groups on apathy were significantly lower at baseline compared with those of the passive control group and may have accounted for a lack of treatment effect. Equality between groups is more likely to be difficult to achieve in smaller samples [89], particularly if the sample includes older adults [90]. In an aged care setting, this is further complicated because of the focus on providing support for people to age in their homes. This means that residents entering aged care are now older, have complex needs, and are taking multiple medications [91]. Although participants with severe cognitive decline were excluded and covariates controlled for baseline differences, there were day-to-day differences in a participant's condition commonly observed throughout the research period. Despite the use of reliable and valid measures, the selection of outcome measures is a challenging issue in residential aged care [92], and they may not be sensitive to changes in all residents at all times.

VR Side Effects

Results from SSQ revealed negligible to minimal side effects according to the cutoff scores [93]. There were 40% (6/15) of participants who reported symptoms before VR because of existing conditions. Therefore, it is important to take measures before exposure of this population to VR to accurately understand and differentiate symptoms occurring during HMD use. Two participants reported possible after effects in the VR group. This occurred in the evening after a morning VR session in 2 participants, one participant reported a headache and the second had a heavy-head feeling. The symptoms were not long-lasting and did not cause any significant discomfort. A consideration for future research is to advise the participants and staff that these symptoms may occur.

Limitations

Participants in this study were not classified as having apathy at baseline according to the AES; therefore, there were participants in the study who had relatively low scores at baseline, leaving little room for improvement. However, this was not supported by subgroup analysis. The approach of not using a cutoff score to classify participants as having apathy was taken because of positive changes seen in our feasibility study [72] in participants with low apathy scores below recommended cutoffs, and the high prevalence of apathy in aged care.

The study included participants with different conditions who were taking a range of medications, and the sample size was relatively small; these factors affected generalizability and was also a possible reason for the lack of significant differences in outcome measures over time. In addition, only those interested in reminiscing may have participated in the research in the intervention groups. Participant selection for recruitment was performed by staff at the aged care facility. This may have resulted in selection bias within the sample. Although 3 separate sites were used in this study, they were run by the same aged care facility, therefore providing a similar quality of care and within a 20-km radius of each other. To maximize comparability between participants, a cutoff score on the Psychogeriatric Assessment Scale was used; therefore, participants with severe cognitive impairment were not included in the trial. Maximizing comparability further would have severely restricted recruitment numbers and the feasibility of the trial. This was overcome by including covariates at baseline in the final analyses, which did not significantly affect the results. As participants with severe cognitive impairment were excluded from this study, we cannot make any assumptions as to whether using HMD is suitable for this stage of cognitive impairment. A sensory approach to reminiscence using physical objects or items, including pictures, may be the best approach for those with severe cognitive impairment [94].

The results of this study may have been affected by the global pandemic. Additional hygiene measures introduced during the research period included wearing masks. Many participants in this age group and setting can have hearing problems, and wearing a mask reduces the communication of facial expressions and reduces the projection of a person's voice and clarity. How this influenced both outcome measures and intervention sessions was not known. Apathy has been reported as one of the most affected symptoms in patients with Alzheimer disease [95] and increased caregiver burden in patients with neuropsychiatric symptoms [96] during the pandemic. Participants also had severely restricted visitations from family and friends at times during the research period, which may have exacerbated the symptoms and reduced the likelihood of statistically significant outcomes of the intervention.

Conclusions

This study examined changes in apathy using VR. Although no statistically significant difference was found between the VR group and the active control group, we have demonstrated that VR can be implemented in an aged care setting with correct protocols in place and that residents in aged care enjoy the process of reminiscing. The use of VR provides access to a wide range of content that is always increasing, and aged care facilities may be able to use VR in other contexts, for example, in lifestyle activities for music or travel.

To facilitate implementation in residential aged care, potential co-designers of VR activities could consist of key stakeholders, including management of aged care facilities, staff who organize and deliver activities, residents, and family or friends of residents. Staff members who deliver activities in residential aged care facilities would normally already have the necessary skills for dealing with and engaging with people in residential aged care. Clinician involvement in training in the use of HMDs would assist in managing health and safety risks and providing the best experience for both implementing it as an activity in residential aged care and for optimum data collection in research. This would include recommendations of HMD type, monitoring of side effects, correct fitment and setup of HMDs, and sourcing of content.

The success of using VR in exposure- and distraction-based therapies may not transfer to therapeutic uses in an aged care setting. However, the results from this trial are not definitive, and longer-term research including more sessions is required. Giving participants a choice between immersive and flat screen technology can assist in increasing enjoyment and participation rates in lifestyle activities, ultimately providing more engaged residents for aged care facilities, particularly for those with

apathy. Physiological measures for assessing immediate effects and the use of actigraphy as an objective measure of apathy are an avenue for further research that will be reported from this study.

Acknowledgments

DS was supported by the Australian Government Research Training Program Scholarship.

TL and HADK were funded by National Health and Medical Research Council Dementia Research Leadership Fellowships (GNT1136269 and GNT1135676, respectively).

The funding bodies listed above did not contribute to the design of the study, data collection, interpretation of data, or writing of the manuscript.

The authors would like to thank Veronica Ogilvy for recruiting the participants.

In addition, the authors would like to thank the following research assistants: Charlotte Sloan, Tyler Ross, and Sara Knayfati.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Secondary and exploratory results.
[DOCX File , 27 KB - jmir_v23i9e29210_app1.docx]

References

1. Selbæk G, Engedal K, Bergh S. The prevalence and course of neuropsychiatric symptoms in nursing home patients with dementia: a systematic review. J Am Med Dir Assoc 2013 Mar;14(3):161-169. [doi: 10.1016/j.jamda.2012.09.027] [Medline: 23168112]
2. Zuidema S, Koopmans R, Verhey F. Prevalence and predictors of neuropsychiatric symptoms in cognitively impaired nursing home patients. J Geriatr Psychiatry Neurol 2007 Mar;20(1):41-49. [doi: 10.1177/0891988706292762] [Medline: 17341770]
3. Marin R, Biedrzycki R, Firinciogullari S. Reliability and validity of the Apathy Evaluation Scale. Psychiatry Res 1991 Aug;38(2):143-162. [doi: 10.1016/0165-1781(91)90040-v] [Medline: 1754629]
4. Brown R, Pluck G. Negative symptoms: the 'pathology' of motivation and goal-directed behaviour. Trends Neurosci 2000 Sep;23(9):412-417. [doi: 10.1016/s0166-2236(00)01626-x] [Medline: 10941190]
5. Robert P, Lanctôt KL, Agüera-Ortiz L, Aalten P, Bremond F, Defrancesco M, et al. Is it time to revise the diagnostic criteria for apathy in brain disorders? The 2018 international consensus group. Eur Psychiatry 2018 Oct;54:71-76 [FREE Full text] [doi: 10.1016/j.eurpsy.2018.07.008] [Medline: 30125783]
6. Ballard C, O'Brien J, James I, Mynt P, Lana M, Potkins D, et al. Quality of life for people with dementia living in residential and nursing home care: the impact of performance on activities of daily living, behavioral and psychological symptoms, language skills, and psychotropic drugs. Int Psychogeriatr 2001 Mar;13(1):93-106. [doi: 10.1017/s1041610201007499] [Medline: 11352339]
7. van Dalen JW, van Wanrooij LL, van Charante EP, Brayne C, van Gool WA, Richard E. Association of apathy with risk of incident dementia: a systematic review and meta-analysis. JAMA Psychiatry 2018 Oct 01;75(10):1012-1021 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.1877] [Medline: 30027214]
8. Nijsten JM, Leontjevas R, Pat-El R, Smalbrugge M, Koopmans RT, Gerritsen DL. Apathy: risk factor for mortality in nursing home patients. J Am Geriatr Soc 2017 Oct;65(10):2182-2189. [doi: 10.1111/jgs.15007] [Medline: 28791690]
9. Kumfor F, Zhen A, Hodges J, Piguet O, Irish M. Apathy in Alzheimer's disease and frontotemporal dementia: distinct clinical profiles and neural correlates. Cortex 2018 Jun;103:350-359. [doi: 10.1016/j.cortex.2018.03.019] [Medline: 29704671]
10. Theleritis CG, Siarkos KT, Politis AM. Unmet needs in pharmacological treatment of apathy in Alzheimer's disease: a systematic review. Front Pharmacol 2019 Oct 4;10:1108 [FREE Full text] [doi: 10.3389/fphar.2019.01108] [Medline: 31680942]
11. Ruthirakuhan M, Herrmann N, Abraham E, Chan S, Lanctôt KL. Pharmacological interventions for apathy in Alzheimer's disease. Cochrane Database Syst Rev 2018 May 04;5(5):CD012197 [FREE Full text] [doi: 10.1002/14651858.CD012197.pub2] [Medline: 29727467]
12. Bogdan A, Manera V, Koenig A, David R. Pharmacologic approaches for the management of apathy in neurodegenerative disorders. Front Pharmacol 2020 Jan 23;10:1581 [FREE Full text] [doi: 10.3389/fphar.2019.01581] [Medline: 32038253]

13. Theleritis C, Siarkos K, Politis AA, Katirtzoglou E, Politis A. A systematic review of non-pharmacological treatments for apathy in dementia. Int J Geriatr Psychiatry 2018 Feb;33(2):177-192. [doi: 10.1002/gps.4783] [Medline: 28960446]
14. Lanctôt KL, Agüera-Ortiz L, Brodaty H, Francis P, Geda Y, Ismail Z, et al. Apathy associated with neurocognitive disorders: recent progress and future directions. Alzheimers Dement 2017 Jan;13(1):84-100. [doi: 10.1016/j.jalz.2016.05.008] [Medline: 27362291]
15. Westphal A, Callega D, LoGiudice D, Lautenschlager N. The Academic Unit for Psychiatry of Old Age, Department of Psychiatry. 2017 Jul. URL: http://medicine.unimelb.edu.au/__data/assets/pdf_file/0020/2471312/Using-reminiscence-with-people-with-dementia-in-acute-and-subacute-care-manual.pdf [accessed 2021-08-13]
16. Dyer S, Laver K, Pond C, Cumming R, Whitehead C, Crotty M. Clinical practice guidelines and principles of care for people with dementia in Australia. Aust Fam Physician 2016 Dec;45(12):884-889 [FREE Full text] [Medline: 27903038]
17. Woods B, O'Philbin L, Farrell EM, Spector AE, Orrell M. Reminiscence therapy for dementia. Cochrane Database Syst Rev 2018 Mar 01;3(3):CD001120 [FREE Full text] [doi: 10.1002/14651858.CD001120.pub3] [Medline: 29493789]
18. Pinquart M, Forstmeier S. Effects of reminiscence interventions on psychosocial outcomes: a meta-analysis. Aging Ment Health 2012;16(5):541-558. [doi: 10.1080/13607863.2011.651434] [Medline: 22304736]
19. Webster JD, Bohlmeijer ET, Westerhof GJ. Mapping the future of reminiscence: a conceptual guide for research and practice. Res Aging 2010 May 24;32(4):527-564. [doi: 10.1177/0164027510364122]
20. Wong PT, Watt LM. What types of reminiscence are associated with successful aging? Psychol Aging 1991 Jun;6(2):272-279. [doi: 10.1037//0882-7974.6.2.272] [Medline: 1863396]
21. Cappeliez P, Guindon M, Robitaille A. Functions of reminiscence and emotional regulation among older adults. J Aging Stud 2008 Aug;22(3):266-272 [FREE Full text] [doi: 10.1016/j.jaging.2007.06.003]
22. Cuevas P, Davidson P, Mejilla J, Rodney T. Reminiscence therapy for older adults with Alzheimer's disease: a literature review. Int J Ment Health Nurs 2020 Jun;29(3):364-371. [doi: 10.1111/inm.12692] [Medline: 31984570]
23. Bluck S. Autobiographical memory: exploring its functions in everyday life. Memory 2003 Mar;11(2):113-123. [doi: 10.1080/741938206] [Medline: 12820825]
24. Strohminger N, Nichols S. The essential moral self. Cognition 2014 Apr;131(1):159-171 [FREE Full text] [doi: 10.1016/j.cognition.2013.12.005] [Medline: 24503450]
25. Stuss D, Van RR, Murphy K. Differentiation of states and causes of apathy. In: Borod JC, editor. The Neuropsychology of Emotion. New York: Oxford University Press; 2000:340-363.
26. Sockeel P, Dujardin K, Devos D, Denève C, Destée A, Defebvre L. The Lille apathy rating scale (LARS), a new instrument for detecting and quantifying apathy: validation in Parkinson's disease. J Neurol Neurosurg Psychiatry 2006 May;77(5):579-584 [FREE Full text] [doi: 10.1136/jnnp.2005.075929] [Medline: 16614016]
27. Surr C. Preservation of self in people with dementia living in residential care: a socio-biographical approach. Soc Sci Med 2006 Apr;62(7):1720-1730. [doi: 10.1016/j.socscimed.2005.08.025] [Medline: 16168543]
28. Park K, Lee S, Yang J, Song T, Hong GS. A systematic review and meta-analysis on the effect of reminiscence therapy for people with dementia. Int Psychogeriatr 2019 Nov;31(11):1581-1597. [doi: 10.1017/S1041610218002168] [Medline: 30712519]
29. Gardiner C, Laud P, Heaton T, Gott M. What is the prevalence of loneliness amongst older people living in residential and nursing care homes? A systematic review and meta-analysis. Age Ageing 2020 Aug 24;49(5):748-757. [doi: 10.1093/ageing/afaa049] [Medline: 32396600]
30. Lazar A, Thompson H, Demiris G. A systematic review of the use of technology for reminiscence therapy. Health Educ Behav 2014 Oct;41(1 Suppl):51-61 [FREE Full text] [doi: 10.1177/1090198114537067] [Medline: 25274711]
31. Bharathi A, Tucker C. Investigating the impact of interactive immersive virtual reality environments in enhancing task performance in online engineering design activities. In: Proceedings of the ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. 2015 Presented at: ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference; August 2-5, 2015; Boston, Massachusetts, USA. [doi: 10.1115/detc2015-47388]
32. Manjrekar S, Sandilya S, Bhosale D, Kanchi S, Pitkar A, Gondhalekar M. CAVE: an emerging immersive technology -- a review. In: Proceedings of the 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation. 2014 Presented at: 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation; Mar 26-28, 2014; Cambridge, UK. [doi: 10.1109/uksim.2014.20]
33. Hale K, Stanney K. Handbook of Virtual Environments: Design, Implementation, and Applications, Second Edition. Boca Raton, Florida: CRC Press; Sep 10, 2014.
34. Yeo N, White M, Alcock I, Garside R, Dean S, Smalley A, et al. What is the best way of delivering virtual nature for improving mood? An experimental comparison of high definition TV, 360° video, and computer generated virtual reality. J Environ Psychol 2020 Dec;72:101500 [FREE Full text] [doi: 10.1016/j.jenvp.2020.101500] [Medline: 33390641]
35. Rizzo A, Shilling R. Clinical virtual reality tools to advance the prevention, assessment, and treatment of PTSD. Eur J Psychotraumatol 2017 Jan 16;8(sup5):1414560 [FREE Full text] [doi: 10.1080/20008198.2017.1414560] [Medline: 29372007]

36. Spiegel B, Fuller G, Lopez M, Dupuy T, Noah B, Howard A, et al. Virtual reality for management of pain in hospitalized patients: a randomized comparative effectiveness trial. PLoS One 2019 Aug 14;14(8):e0219115 [FREE Full text] [doi: 10.1371/journal.pone.0219115] [Medline: 31412029]
37. Witmer BG, Jerome CJ, Singer MJ. The factor structure of the presence questionnaire. Presence 2005 Jun;14(3):298-312. [doi: 10.1162/105474605323384654]
38. Cummings JJ, Bailenson JN. How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. Media Psychol 2015 May 15;19(2):272-309. [doi: 10.1080/15213269.2015.1015740]
39. Schöne B, Wessels M, Gruber T. Experiences in virtual reality: a window to autobiographical memory. Curr Psychol 2017 Jul 20;38(3):715-719. [doi: 10.1007/s12144-017-9648-y]
40. Krokos E, Plaisant C, Varshney A. Virtual memory palaces: immersion aids recall. Virtual Real 2018 May 16;23(1):1-15. [doi: 10.1007/s10055-018-0346-3]
41. Schiza E, Matsangidou M, Neokleous K, Pattichis CS. Virtual reality applications for neurological disease: a review. Front Robot AI 2019 Oct 16;6:100 [FREE Full text] [doi: 10.3389/frobt.2019.00100] [Medline: 33501115]
42. Dermody G, Whitehead L, Wilson G, Glass C. The role of virtual reality in improving health outcomes for community-dwelling older adults: systematic review. J Med Internet Res 2020 Jun 01;22(6):e17331 [FREE Full text] [doi: 10.2196/17331] [Medline: 32478662]
43. Benham S, Kang M, Grampurohit N. Immersive virtual reality for the management of pain in community-dwelling older adults. OTJR (Thorofare N J) 2019 Apr;39(2):90-96. [doi: 10.1177/1539449218817291] [Medline: 30595096]
44. Optale G, Urgesi C, Busato V, Marin S, Piron L, Priftis K, et al. Controlling memory impairment in elderly adults using virtual reality memory training: a randomized controlled pilot study. Neurorehabil Neural Repair 2010 May;24(4):348-357. [doi: 10.1177/1545968309353328] [Medline: 19934445]
45. Levy F, Leboucher P, Rautureau G, Komano O, Millet B, Jouvent R. Fear of falling: efficacy of virtual reality associated with serious games in elderly people. Neuropsychiatr Dis Treat 2016 Apr 15;12:877-881 [FREE Full text] [doi: 10.2147/NDT.S97809] [Medline: 27143889]
46. Chan J, Chan T, Wong M, Cheung R, Yiu K, Tsoi K. Effects of virtual reality on moods in community older adults. A multicenter randomized controlled trial. Int J Geriatr Psychiatry 2020 Aug;35(8):926-933. [doi: 10.1002/gps.5314] [Medline: 32346896]
47. Webber S, Baker S, Waycott J. Virtual visits: Reminiscence in residential aged care with digital mapping technologies. Australas J Ageing 2021 Feb 04:12902 (forthcoming). [doi: 10.1111/ajag.12902] [Medline: 33538405]
48. Bruun-Pedersen J, Serafin S, Kofoed L. Going outside while staying inside — exercise motivation with immersive vs. non–immersive recreational virtual environment augmentation for older adult nursing home residents. In: Proceedings of the 2016 IEEE International Conference on Healthcare Informatics (ICHI). 2016 Presented at: 2016 IEEE International Conference on Healthcare Informatics (ICHI); Oct 4-7 Oct, 2016; Chicago, IL, USA. [doi: 10.1109/ichi.2016.31]
49. Lin C, Lee C, Lally D, Coughlin J. Impact of Virtual Reality (VR) experience on older adults' well-being. In: Human Aspects of IT for the Aged Population. Applications in Health, Assistance, and Entertainment. Basel, Switzerland: Springer; 2018.
50. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. Int J Aviat Psychol 1993 Jul;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
51. Szpak A, Michalski SC, Saredakis D, Chen CS, Loetscher T. Beyond feeling sick: the visual and cognitive aftereffects of virtual reality. IEEE Access 2019 Sep 9;7:130883-130892. [doi: 10.1109/access.2019.2940073]
52. Moher D, Hopewell S, Schulz K, Montori V, Gøtzsche PC, Devereaux P, CONSORT. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surg 2012;10(1):28-55 [FREE Full text] [doi: 10.1016/j.ijsu.2011.10.001] [Medline: 22036893]
53. Saredakis D, Keage HA, Corlis M, Loetscher T. Virtual reality intervention to improve apathy in residential aged care: protocol for a multisite non-randomised controlled trial. BMJ Open 2021 Feb 01;11(2):e046030 [FREE Full text] [doi: 10.1136/bmjopen-2020-046030] [Medline: 33526504]
54. Jorm AF, Mackinnon AJ, Henderson AS, Scott R, Christensen H, Korten AE, et al. The Psychogeriatric Assessment Scales: a multi-dimensional alternative to categorical diagnoses of dementia and depression in the elderly. Psychol Med 1995 May;25(3):447-460. [doi: 10.1017/s0033291700033377] [Medline: 7480426]
55. Høeg ER, Povlsen TM, Bruun-Pedersen JR, Lange B, Nilsson NC, Haugaard KB, et al. System immersion in virtual reality-based rehabilitation of motor function in older adults: a systematic review and meta-analysis. Front Virtual Real 2021 Apr 12;2:647993. [doi: 10.3389/frvir.2021.647993]
56. Oculus. URL: https://www.oculus.com/ [accessed 2021-08-13]
57. Noone P. Addenbrooke's cognitive examination-III. Occup Med (Lond) 2015 Jul 17;65(5):418-420. [doi: 10.1093/occmed/kqv041] [Medline: 26187808]
58. Hsieh S, Schubert S, Hoon C, Mioshi E, Hodges JR. Validation of the Addenbrooke's Cognitive Examination III in frontotemporal dementia and Alzheimer's disease. Dement Geriatr Cogn Disord 2013;36(3-4):242-250. [doi: 10.1159/000351671] [Medline: 23949210]

59. Yesavage JA, Sheikh JI. 9/Geriatric Depression Scale (GDS). Clin Gerontol 1986;5(1-2):165-173. [doi: 10.1300/J018v05n01_09]
60. Kurlowicz L, Greenberg S. The Geriatric Depression Scale (GDS). Am J Nurse 2007 Oct;107(10):67-68. [doi: 10.1097/01.naj.0000292207.37066.2f]
61. D'Ath P, Katona P, Mullan E, Evans S, Katona C. Screening, detection and management of depression in elderly primary care attenders. I: the acceptability and performance of the 15 item Geriatric Depression Scale (GDS15) and the development of short versions. Fam Pract 1994 Sep;11(3):260-266. [doi: 10.1093/fampra/11.3.260] [Medline: 7843514]
62. Logsdon RG, Gibbons LE, McCurry SM, Teri L. Assessing quality of life in older adults with cognitive impairment. Psychosom Med 2002;64(3):510-519. [doi: 10.1097/00006842-200205000-00016] [Medline: 12021425]
63. Thorgrimsen L, Selwood A, Spector A, Royan L, de Madariaga Lopez M, Woods RT, et al. Whose quality of life is it anyway? The validity and reliability of the Quality of Life-Alzheimer's Disease (QoL-AD) scale. Alzheimer Dis Assoc Disord 2003;17(4):201-208. [doi: 10.1097/00002093-200310000-00002] [Medline: 14657783]
64. Hughes ME, Waite LJ, Hawkley LC, Cacioppo JT. A short scale for measuring loneliness in large surveys: results from two population-based studies. Res Aging 2004;26(6):655-672 [FREE Full text] [doi: 10.1177/0164027504268574] [Medline: 18504506]
65. Rebenitsch L, Owen C. Review on cybersickness in applications and visual displays. Virtual Real 2016 Apr 26;20(2):101-125. [doi: 10.1007/s10055-016-0285-9]
66. Gibson F. Reminiscence and Life Story Work: A Practice Guide. USA: Jessica Kingsley Publishers; Apr 1, 2011.
67. Harris P, Taylor R, Thielke R, Payne J, Gonzalez N, Conde J. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
68. The jamovi project. Jamovi. URL: https://www.jamovi.org/about.html [accessed 2021-08-13]
69. Integrated development environment. R Studio. URL: https://www.rstudio.com/categories/integrated-development-environment/ [accessed 2021-08-13]
70. jorvlan/open-visualizations: open-visualizations in R and Python. Zenodo. URL: https://zenodo.org/record/3715538#.YRYysogzY2x [accessed 2021-08-13]
71. Clarke DE, Reekum RV, Simard M, Streiner DL, Freedman M, Conn D. Apathy in dementia: an examination of the psychometric properties of the apathy evaluation scale. J Neuropsychiatry Clin Neurosci 2007;19(1):57-64. [doi: 10.1176/jnp.2007.19.1.57] [Medline: 17308228]
72. Saredakis D, Keage HA, Corlis M, Loetscher T. Using virtual reality to improve apathy in residential aged care: mixed methods study. J Med Internet Res 2020 Jun 26;22(6):e17632 [FREE Full text] [doi: 10.2196/17632] [Medline: 32469314]
73. Coelho T, Marques C, Moreira D, Soares M, Portugal P, Marques A, et al. Promoting reminiscences with virtual reality headsets: a pilot study with people with dementia. Int J Environ Res Public Health 2020 Dec 12;17(24):9301 [FREE Full text] [doi: 10.3390/ijerph17249301] [Medline: 33322679]
74. Difede J, Cukor J, Jayasinghe N, Patt I, Jedel S, Spielman L, et al. Virtual reality exposure therapy for the treatment of posttraumatic stress disorder following September 11, 2001. J Clin Psychiatry 2007 Nov;68(11):1639-1647. [Medline: 18052556]
75. Haight BK, Haight BS. The Handbook of Structured Life Review. US: Health Professions Press; 2007.
76. MacPherson S, Bird M, Anderson K, Davis T, Blair A. An art gallery access programme for people with dementia: 'you do it for the moment'. Aging Ment Health 2009 Sep;13(5):744-752. [doi: 10.1080/13607860902918207] [Medline: 19882413]
77. Brimelow RE, Dawe B, Dissanayaka N. Preliminary research: virtual reality in residential aged care to reduce apathy and improve mood. Cyberpsychol Behav Soc Netw 2020 Mar;23(3):165-170. [doi: 10.1089/cyber.2019.0286] [Medline: 31829729]
78. Hsieh C, Chang C, Su S, Hsiao Y, Shih Y, Han W, et al. Reminiscence group therapy on depression and apathy in nursing home residents with mild-to-moderate dementia. J Exp Clin Med 2010 Apr;2(2):72-78 [FREE Full text] [doi: 10.1016/s1878-3317(10)60012-5]
79. İnel Manav A, Simsek N. The effect of reminiscence therapy with internet-based videos on cognitive status and apathy of older people with mild dementia. J Geriatr Psychiatry Neurol 2019 Mar;32(2):104-113. [doi: 10.1177/0891988718819864] [Medline: 30612511]
80. Brodaty H, Burns K. Nonpharmacological management of apathy in dementia: a systematic review. Am J Geriatr Psychiatry 2012 Jul;20(7):549-564. [doi: 10.1097/JGP.0b013e31822be242] [Medline: 21860324]
81. Harmand MG, Meillon C, Rullier L, Avila-Funes J, Bergua V, Dartigues JF, et al. Cognitive decline after entering a nursing home: a 22-year follow-up study of institutionalized and noninstitutionalized elderly people. J Am Med Dir Assoc 2014 Jul;15(7):504-508. [doi: 10.1016/j.jamda.2014.02.006] [Medline: 24679926]
82. Baldwin SA, Berkeljon A, Atkins DC, Olsen JA, Nielsen SL. Rates of change in naturalistic psychotherapy: contrasting dose-effect and good-enough level models of change. J Consult Clin Psychol 2009 Apr;77(2):203-211. [doi: 10.1037/a0015235] [Medline: 19309180]

83. Baillon S, Van Diepen E, Prettyman R, Redman J, Rooke N, Campbell R. A comparison of the effects of Snoezelen and reminiscence therapy on the agitated behaviour of patients with dementia. Int J Geriatr Psychiatry 2004 Nov;19(11):1047-1052. [doi: 10.1002/gps.1208] [Medline: 15481068]
84. Okumura Y, Tanimukai S, Asada T. Effects of short-term reminiscence therapy on elderly with dementia: a comparison with everyday conversation approaches. Psychogeriatrics 2008 Sep;8(3):124-133. [doi: 10.1111/j.1479-8301.2008.00236.x]
85. Morath MA, Okun JG, Müller IB, Sauer SW, Hörster F, Hoffmann GF, et al. Neurodegeneration and chronic renal failure in methylmalonic aciduria--a pathophysiological approach. J Inherit Metab Dis 2008 Feb;31(1):35-43. [doi: 10.1007/s10545-007-0571-5] [Medline: 17846917]
86. Ando M, Morita T, Oshima A. Reminiscence cognitive behavior therapy for spiritual well-being and self-esteem of cancer patients. J Cancer Ther 2011 Jun;02(02):105-109. [doi: 10.4236/jct.2011.22012]
87. Davis MC. Life review therapy as an intervention to manage depression and enhance life satisfaction in individuals with right hemisphere cerebral vascular accidents. Issues Ment Health Nurs 2004;25(5):503-515. [doi: 10.1080/01612840490443455] [Medline: 15204893]
88. Steinhauser KE, Alexander SC, Byock IR, George LK, Olsen MK, Tulsky JA. Do preparation and life completion discussions improve functioning and quality of life in seriously ill patients? Pilot randomized control trial. J Palliat Med 2008 Nov;11(9):1234-1240. [doi: 10.1089/jpm.2008.0078] [Medline: 19021487]
89. Broglio K. Randomization in clinical trials: permuted blocks and stratification. J Am Med Assoc 2018 Jun 05;319(21):2223-2224. [doi: 10.1001/jama.2018.6360] [Medline: 29872845]
90. Netz Y, Lidor R, Ziv G. Small samples and increased variability - discussing the need for restricted types of randomization in exercise interventions in old age. Eur Rev Aging Phys Act 2019 Oct 27;16:17 [FREE Full text] [doi: 10.1186/s11556-019-0224-3] [Medline: 31673298]
91. Sluggett J, Ilomäki J, Seaman K, Corlis M, Bell J. Medication management policy, practice and research in Australian residential aged care: current and future directions. Pharmacol Res 2017 Feb;116:20-28. [doi: 10.1016/j.phrs.2016.12.011] [Medline: 27965033]
92. Lam HR, Chow S, Taylor K, Chow R, Lam H, Bonin K, et al. Challenges of conducting research in long-term care facilities: a systematic review. BMC Geriatr 2018 Oct 12;18(1):242 [FREE Full text] [doi: 10.1186/s12877-018-0934-9] [Medline: 30314472]
93. Kennedy R, Drexler J, Compton D, Stanney K, Lanham D, Harm D. Configural scoring of simulator sickness, cybersickness and space adaptation syndromeimilarities and differences. In: Hettinger LJ, Haas MW, editors. Virtual and Adaptive Environments: Applications, Implications, and Human Performance Issues. Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers; 2003:247.
94. Kverno KS, Black BS, Nolan MT, Rabins PV. Research on treating neuropsychiatric symptoms of advanced dementia with non-pharmacological strategies, 1998-2008: a systematic literature review. Int Psychogeriatr 2009 Oct;21(5):825-843 [FREE Full text] [doi: 10.1017/S1041610209990196] [Medline: 19586562]
95. Lara B, Carnes A, Dakterzada F, Benitez I, Piñol-Ripoll G. Neuropsychiatric symptoms and quality of life in Spanish patients with Alzheimer's disease during the COVID-19 lockdown. Eur J Neurol 2020 Sep;27(9):1744-1747 [FREE Full text] [doi: 10.1111/ene.14339] [Medline: 32449791]
96. Penteado CT, Loureiro JC, Pais MV, Carvalho CL, Sant'Ana LF, Valiengo LC, et al. Mental health status of psychogeriatric patients during the 2019 new coronavirus disease (COVID-19) pandemic and effects on caregiver burden. Front Psychiatry 2020 Nov 17;11:578672 [FREE Full text] [doi: 10.3389/fpsyt.2020.578672] [Medline: 33312138]

Abbreviations

ACE-III: Addenbrooke Cognitive Examination III
AES: Apathy Evaluation Scale
CONSORT: Consolidated Standards of Reporting Trials
GDS: Geriatric Depression Scale
HMD: head-mounted displays
QOL-AD: Quality of Life in Alzheimer Disease
SSQ: Simulator Sickness Questionnaire
VR: virtual reality

Edited by R Kukafka; submitted 12.04.21; peer-reviewed by G Dermody, G Wilson, M Orrell, J Persson; comments to author 09.06.21; revised version received 23.06.21; accepted 27.07.21; published 20.09.21.

Please cite as:
Saredakis D, Keage HAD, Corlis M, Ghezzi ES, Loffler H, Loetscher T
The Effect of Reminiscence Therapy Using Virtual Reality on Apathy in Residential Aged Care: Multisite Nonrandomized Controlled Trial
J Med Internet Res 2021;23(9):e29210
URL: https://www.jmir.org/2021/9/e29210
doi:10.2196/29210
PMID:34542418

Original Paper

Comparing a Virtual Reality–Based Simulation App (VR-MRI) With a Standard Preparatory Manual and Child Life Program for Improving Success and Reducing Anxiety During Pediatric Medical Imaging: Randomized Clinical Trial

Chelsea Stunden[1,2*], MPH; Kirsten Stratton[3], CCLS; Sima Zakani[1,2], PhD; John Jacob[1,2,4*], MSc, MBA

[1]Department of Pediatrics, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

[2]Digital Lab at BC Children's Hospital, Vancouver, BC, Canada

[3]Department of Child Life, BC Children's Hospital, Vancouver, BC, Canada

[4]City University of London, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
John Jacob, MSc, MBA
Department of Pediatrics
Faculty of Medicine
University of British Columbia
Rm 2D19
4480 Oak Street
Vancouver, BC, V6H 3V4
Canada
Phone: 1 604 875 2345 ext 2393
Email: john.jacob@cw.bc.ca

Abstract

Background: The experience of undergoing magnetic resonance imaging (MRI) can be anxiety provoking, particularly for pediatric patients and their families. Alternative methods to improve success and experiences without the use of sedation are needed.

Objective: This study aims to compare the effectiveness of a virtual reality (VR)—based simulation app (VR-MRI) with a standard preparatory manual (SPM) and a hospital-based Child Life Program (CLP) on success and anxiety during a simulated pediatric MRI scan. Our secondary aim is to compare caregivers' reported anxiety, procedural data, caregiver usability, child satisfaction, and fun.

Methods: This unblinded, randomized, triple-arm clinical trial involved 92 children aged 4-13 years and their caregivers. Recruitment was conducted through posters, public libraries, community centers, and social media. At a 2-hour session, participants were instructed to prepare for a simulated MRI head scan using one of three randomly assigned preparation materials: the VR-MRI app, SPM, or the CLP. Data were collected before preparation, during a simulated MRI head scan, and after the simulated scan. The primary outcomes were the success of the simulated MRI scan (MoTrak head motion tracking system), and child-reported anxiety (Venham picture test). We secondarily measured caregivers' reported anxiety (short State-Trait Anxiety Inventory), procedural data (minutes), usability (Usefulness, Satisfaction, and Ease of Use Questionnaire), and child-reported satisfaction and fun (visual analog scales).

Results: A total of 84 participants were included in the final analysis (VR-MRI: 30/84, 36%; SPM: 24/84, 29%; and CLP: 30/84, 36%). There were no *clinically* significant differences between the groups in terms of success during the MRI simulation (P=.27) or the children's reported anxiety at any timepoint (timepoint 1, P=.99; timepoint 2, P=.008; timepoint 3, P=.10). Caregivers reported being significantly more anxious after preparing with the manual than caregivers in the other 2 groups (P<.001). Child and caregiver anxiety had a significant relationship, increasing together with moderate effect (r_{84}=0.421; P<.001). Participants using VR-MRI took the most time to prepare (P<.001) and participants using the manual took the least time (P<.001). No statistically significant relationships were found between time preparing and time completing the simulated assessment (P=.13). There were no differences found in ease of use (P=.99), ease of learning (P=.48), and usefulness (P=.11) between the

groups; however, caregivers reported being significantly more satisfied with the VR-MRI app and CLP than SPM (P<.001). Children reported the most satisfaction with the CLP (P<.001). There were no differences in how much fun the preparation materials were perceived to be (P=.37).

Conclusions: Digital preparation experiences using VR-based media could be a viable solution to improve the success of nonsedated MRI scans, with outcomes comparable with hospital-based in-person preparatory programs. Future research should focus on validating the results in a real MRI setting.

Trial Registration: Clinicaltrials.gov NCT03931382; https://clinicaltrials.gov/ct2/show/NCT03931382

(J Med Internet Res 2021;23(9):e22942) doi:10.2196/22942

KEYWORDS

virtual reality; magnetic resonance imaging simulation; certified child life specialists; pediatrics; magnetic resonance imaging procedures; alternatives to sedation; preparing children for MRI

Introduction

Background

Previous reports estimate that between 50% and 75% of pediatric patients experience elevated anxiety and distress before a new medical procedure, such as medical imaging [1,2]. Poor management of this anxiety has an impact on patient experience by causing undue psychological and physiological distress [1-4] and movement during the procedure, which can significantly reduce the diagnostic quality of medical imaging [5]. It also has a ripple effect by causing additional logistical, operational, and economic burdens on the health care system and patients [6,7]. Short-term consequences include pain or discomfort, anxiety, crying, and poor cooperation during the procedure [1,8,9]. In the long term, negative experiences can evolve into posttraumatic stress syndrome, fear, changes in pain perception and coping effectiveness, avoidance of medical care, and phobias, causing difficulties during future medical experiences [9,10]. These consequences can lead to additional resource requirements, system impacts, and costs of care that are reported to be 3.24-9.56 times higher for sedated patients than for those who can complete the procedure unsedated [6]. Thus, it is important to properly and effectively assess pediatric patients before medical procedures and improve compliance with appropriate preparation techniques whenever feasible.

At our institution, >4500 pediatric patients undergo magnetic resonance imaging (MRI) each year, with approximately 45% of patients requiring sedation. In the 2017 to 2018 fiscal year, the waitlist for sedated MRI reached a significant point where some patients were forecasted to endure a wait of >1 year for their scheduled appointments. To address this, the hospital launched a coordinated strategy that included increased operational capacity for medical imaging and funding for clinical and support services. Certified child life specialists (CCLSs), trained to prepare patients and families for procedures, were among these resources. CCLSs make targeted efforts to reduce the need for procedural sedation through training and exposure therapy techniques involving simulated procedures and therapeutic play [11-14]. Workshops or programs delivered by CCLSs have been implemented in >400 North American health care settings [13] and are considered a key factor in enabling some patients to undergo imaging without the use of sedation, where it may have otherwise been indicated to address high preprocedural anxiety [11,15].

The Child Life Program (CLP) at our hospital uses a replica of an MRI unit to orientate and practice the process with patients before the true imaging procedure. This method has been documented in the literature by other investigators [15-17]. However, capacity limitations still exist, and there are socioeconomic costs and logistical considerations of only having these units available on-site at tertiary care facilities. To mitigate this, our hospital offers physical materials such as preparatory manuals and telephone and email consultations to orient patients before the procedure. These methods have also been discussed in the literature [12,14,18,19]; however, there is some conflicting evidence regarding their efficacy [20].

Recently, virtual reality (VR), a computer-generated simulation of a 3D environment that can be explored and interacted with by the use of a head-mounted display (HMD), has emerged as an effective solution for reducing anxiety in a variety of pediatric psychological applications [21,22], including preprocedural anxiety [23-32]. Although the use of VR in hospital settings is promising, to our knowledge, only one study has compared it with the CLP regarding anxiety [30]. We also found only one small meta-analysis comparing different types of preparation programs regarding objective image quality [33]. Of the studies that have been conducted to evaluate VR in procedural preparation thus far, most have focused primarily on self-reporting and survey metrics [27,28,30]. There is a need for rigorous randomized clinical trials using innovative methods to establish the safety, feasibility, and efficacy of alternatives to sedation in preparing pediatric patients for medical imaging procedures before applying them in real pediatric clinical settings [21,34-36].

Objectives

In this context, this study aimed to compare the effectiveness of a custom-developed VR-based intervention with established hospital alternatives in preparing children aged 4-13 years for a simulated medical imaging procedure. We hypothesized that a VR app (VR-MRI) based on experiential learning [37] and social cognitive theory [38] would be effective in preparing pediatric patients for a successful MRI experience, reducing periprocedural anxieties and, thus, noncompliant behaviors contributing to poor image quality or acquisition. We secondarily hypothesize that:

- VR-MRI would reduce caregiver anxiety;

- Children's anxiety would be related to their caregiver's anxiety;
- More time practicing would result in periprocedural efficiency;
- Caregivers would be satisfied with VR-MRI, and it would be perceived as useful and easy to learn;
- Children would be satisfied with VR-MRI, and it would be perceived as fun to use.

Ultimately, we suggest that digital preparation apps using immersive media, such as our VR-MRI solution, could improve outcomes and patient experience by introducing standardized, accessible, and independently repeatable opportunities for experiential learning and preprocedural simulated practice. Furthermore, we suggest that interventions such as this could reduce costs and burden of health systems [7] through increased efficiency and reduced need for pharmaceutical intervention or anesthesia services to improve compliance during medical imaging.

Methods

VR-MRI Design

The VR-MRI media was custom-designed by the British Columbia Children's Hospital Digital Lab. The development process was informed by the literature [39], encompassing recent design strategy recommendations for VR [34]. The development of VR-MRI included iterative consultation and testing with a multidisciplinary team of approximately 8 stakeholders, including CCLSs, radiology technicians, child psychologists, health system administrators, and the research team. The curriculum focused on the same material included in the standard preparatory manual (SPM) and CLP: developing rapport with medical professionals, getting comfortable with the hospital setting and medical equipment, assessing reactions to pictures of a real MRI, discussing the upcoming medical procedure, and getting comfortable with earplugs, headphones, loud noises, restraints, the head coil, going into and remaining inside the bore, and holding still.

We used the agile development methodology to cycle through user experience design, development, alpha testing, and beta testing [40]. The content in our study was custom developed in Unity (Unity Technologies version 2018.4.9f1) and displayed on a Samsung S9 mobile phone that was used with a MERGE VR headset (Merge Labs Inc). The headset was selected for its balance of quality (eg, repeat use, compatibility with hygiene solutions, and interpupillary adjustments designed for children) and affordability, priced at Can $40 (US $32). The headset requires a mobile phone to be inserted into an embedded front panel that is viewable to the users. We used AirServer Connect (App Dynamic ehf) to mirror the VR-MRI sequence in real time on a tablet device for caregivers to watch in parallel.

VR-MRI App

Tutorial

A tutorial was designed to help the user learn how to interact with the different elements presented in the virtual environment. The tutorial included a dinosaur in outer space that taught the user how to interact with the elements, see in 360°, and interact with hotspots (referred to as teleportation devices). An image of the tutorial is shown in Figure 1.

Figure 1. A screenshot of the virtual reality magnetic resonance imaging tutorial.

Tour Sequence

The tour sequence aimed to get the user comfortable at the hospital, with the medical staff, and with the medical equipment. For creating a sense of safety in the hospital, users were first introduced to a radiologist and a peer in the reception area (Figure 2). They led the user through an interactive guided tour of the hospital reception areas, the imaging room, and the steps of a head scan. Each room can be passively explored by rotating the head in 360°. To introduce some autonomy, we also implemented interactive hotspots that needed to be activated by the user to transition between rooms or sequences (Figure 3).

Figure 2. A screenshot of the introduction to hospital staff and a peer in the reception area.

Figure 3. A screenshot of the interactive hotspots used to transition through rooms and sequences.

In the MRI room, the procedural steps of the head scan were introduced. The MRI procedure involved wearing earplugs, headphones, putting on the restraints (referred to as seatbelts), putting on the head coil (referred to as a helmet), listening to sounds (referred to as familiar sounds), and going into the bore (referred to as the tunnel). If the user was distracted during the introduction of important information, cued by not looking in the appropriate direction, the sequence would stall until the user's attention was refocused appropriately (Figure 4). To improve the child's ability to cope with the MRI sounds, which are very loud and noisy, we introduced them as familiar sounds using a narrative element while the sounds played. To further build comfort with the loud noises, users were given the opportunity to play with the sounds via buttons (Figure 5).

Figure 4. A screenshot of the sequence focusing the user's attention on the head coil.

Figure 5. A screenshot of narrative elements used to describe the magnetic resonance imaging sounds.

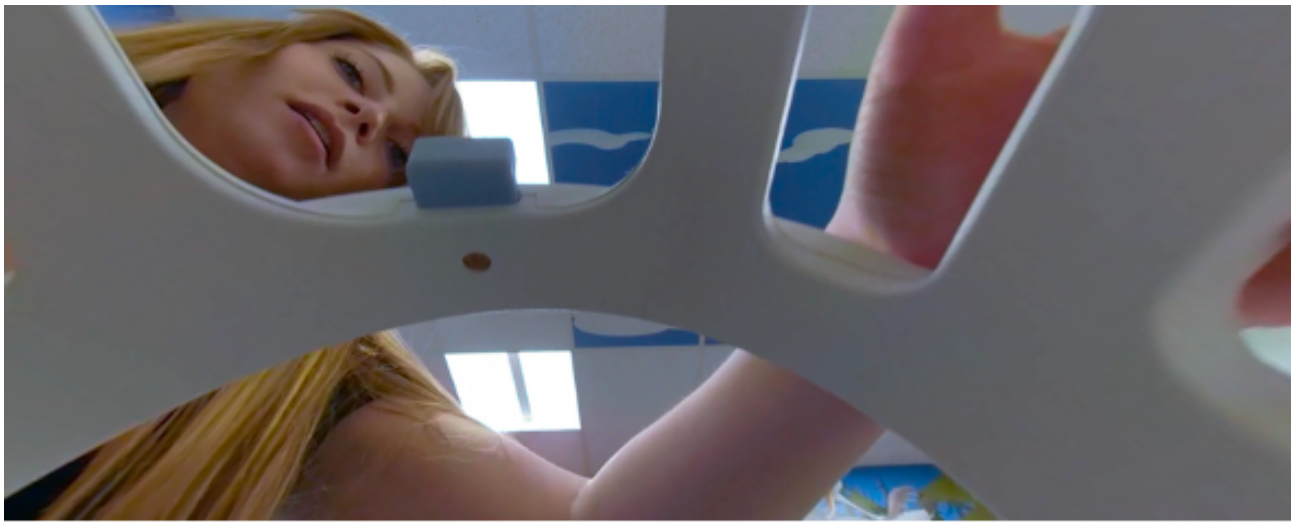
After the introduction to the procedure, the user was invited to watch a peer successfully complete an MRI scan and then invited to complete one themselves. At this point, the user could choose to either watch the peer again or attempt it themselves by selecting an interactive hotspot.

The Virtual MRI Experience and Real-time Feedback Game

During the user's virtual imaging experience (first-person view), the radiologist prepared them for the scan by inserting the earplugs and putting on the headphones (Figure 6). Once the safety equipment was put on, the user was instructed to lie down. For safety purposes, the research assistant helped facilitate this movement on a yoga mat. After the participant laid down, the scene progressed to attaching the head coil (Figure 7). The radiologist then placed them in the tunnel.

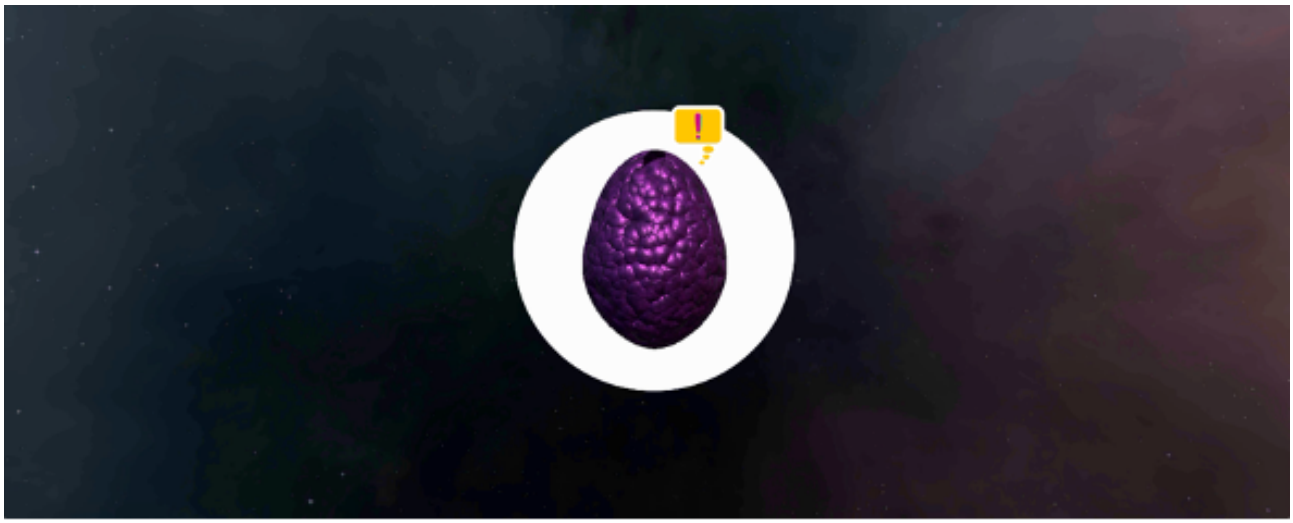
Figure 6. A screenshot of putting in earplugs during the user's virtual imaging experience.

Figure 7. A screenshot of putting on the head coil during the user's virtual imaging experience.



Once in the tunnel, a sequence began to engage the patient in game-based activities to build capacity for self-regulation of their movements during the scan. The user was invited to interact with a dinosaur egg that offered real-time feedback on indicators of movement (such as tilt, shake, rotation, or swing) measured by the mobile phone's gyroscope (Figure 8).

Figure 8. A screenshot of the real-time feedback during movement, measured using the mobile phone gyroscope.



As a means of progression, the game has three levels that become more challenging as feedback mechanisms are eliminated, and presence in the bore is reintroduced through sounds and visual cues (Figure 9). The time to complete each game differs between participants, as success can only be achieved by staying still. After the third level was achieved, users were invited to try again by staying still or exit the bore. Once the system was prompted to exit, the user was removed from the tunnel, and the experience was then noted as complete.

Figure 9. Screenshots of the levels where feedback mechanisms are eliminated and presence in the bore is reintroduced through sounds and visual cues.

Standard Preparatory Manual

The SPM group received the hospital's SPM for nonsedated MRI. The manual contains a series of photos showing the MRI experience step-by-step and is intended to help children and their families prepare for medical imaging (Multimedia Appendix 1). Caregivers were instructed to use the manual to prepare for an MRI as they would at home; however, standardized preparation was not enforced as this would not have been reflective of the environment. Materials (eg, chair and MRI sounds) referenced in the manual for practicing were provided.

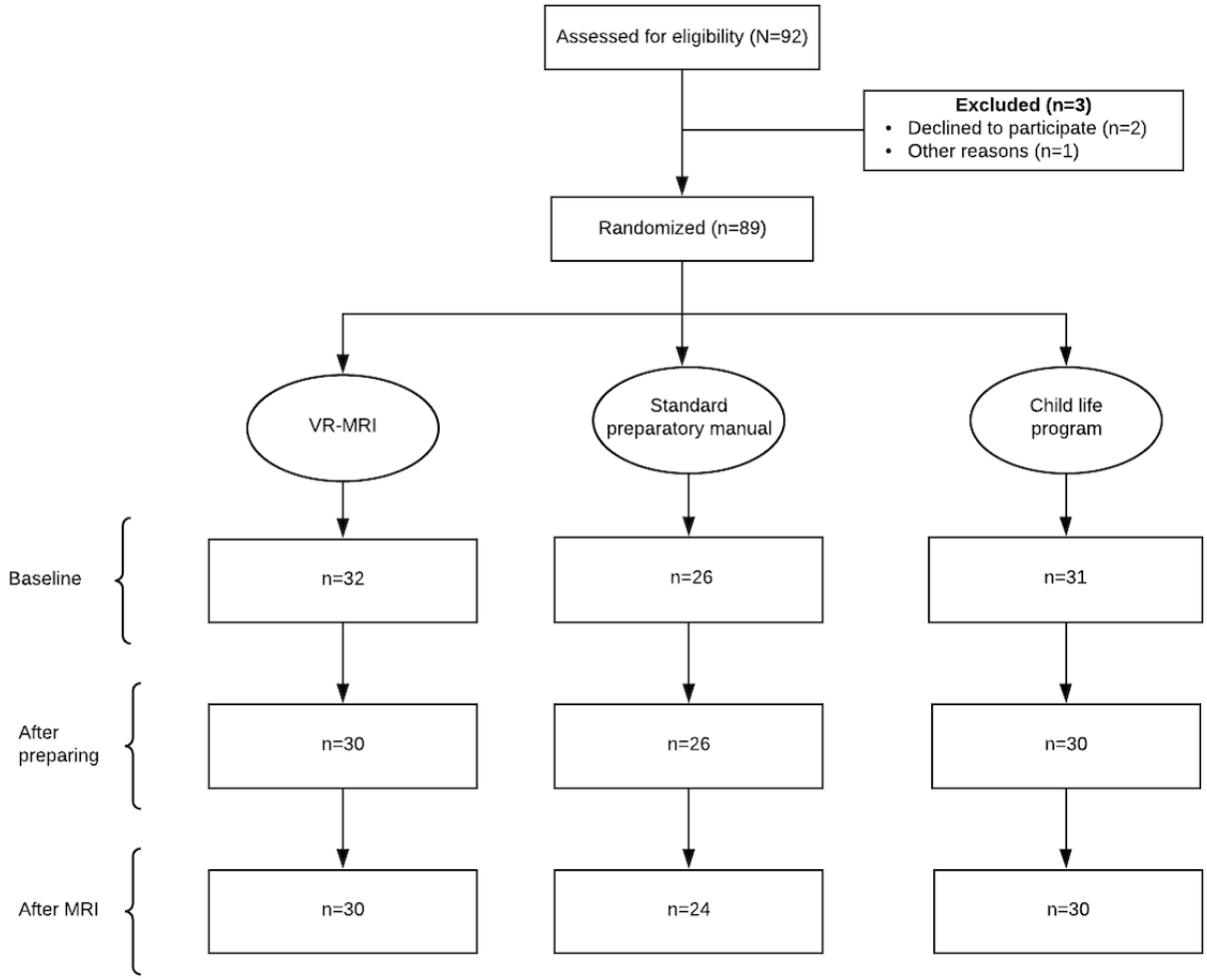
Child Life Program

The CLP group received conventional care, where a CCLS was introduced and prepared the participant with the hospital MRI simulator. The curriculum focused on developing rapport with medical professionals, getting comfortable with the hospital setting and medical equipment, assessing reactions to pictures of a real MRI, discussing the upcoming medical procedure, and getting comfortable with earplugs, headphones, loud noises, restraints, the head coil, going into and remaining inside the bore, and holding still. Preparation was not standardized as it would not have been reflective of the current environment. Individualization and adaptation are the tenets of CLP processes.

Principal Objectives

The primary objectives were to evaluate the effectiveness of VR-MRI in preparing children for the MRI simulator experience and reduce child-reported procedural anxiety compared with conventional methods. As a secondary aim, we evaluated caregivers' reported anxiety, procedural data, parental usability, and child-reported satisfaction with all the preparation materials. Data were collected at 3 distinct timepoints (Figure 10).

Figure 10. Flow diagram for virtual reality magnetic resonance imaging study. MRI: magnetic resonance imaging; VR-MRI: virtual reality magnetic resonance imaging.



Participants, Randomization, and Procedures

A nonblinded, triple-arm randomized clinical trial was performed at a large provincial hospital in Vancouver (British Columbia), between July 2019 and February 2020. Ethics approval was granted by the University of British Columbia Children's & Women's Research Ethics Board (#H19-00371), and the study was prospectively registered at the US National Library of Medicine (#NCT03931382).

The participants were aged 4-13 years. Participants were excluded from the study if they had mental disability, current concussion, significant visual or auditory impairment, inability

to speak and understand English, history of seizures or epilepsy, facial or head wounds, or inability to move their head in all directions. All children provided assent, and caregivers or legal guardians provided written consent. Participants received Can $20 (US $16) and parking remuneration.

Participants were recruited through posters at the hospital, public libraries, community centers, and through social media. We assigned participants in the ratio 1:1:1 to VR-MRI, CLP and SPM and then tested for compliance during a simulated 6-minute MRI scan of the head, designed to replicate an authentic scanning environment (Figure 11). Blinding was not feasible.

Figure 11. The magnetic resonance imaging simulation room.

Outcome Measurements

Primary Outcomes

The primary outcomes were success in the MRI simulation experience and child anxiety.

Success in the MRI Simulation Experience

Movement in the MRI simulation was captured by fitting participants with a motion sensor headband (MoTrak System 1.0, Psychology Software Tools Inc). The sensor system is susceptible to movements of the muscles of facial expression, and has been proposed as one of the most accurate ways to measure the movement of pediatric patients in the absence of estimates from actual MRI [41,42]. Head movement was collected at 8 samples per second, yielding approximately 3300 data points per axis per participant. The threshold for a *successful* MRI, as defined by the department of radiology, is approximately 3-4 mm. Consequently, if at any point during the scan the participant moved >4 mm of cumulative displacement, it was noted as a *fail*.

Child Anxiety

Child anxiety was measured with the Venham picture test (VPT; score 0-8) [43]. The VPT has been validated with children for assessing situational anxiety during medical procedures [44] and has a moderately high retest reliability of 0.70 and a high degree of internal consistency with a coefficient α of .838 [43]. Children completed the assessment at three timepoints (before preparing [T1], after preparing and upon entering the simulation room for their 6-minute *scan* [T2], and after the assessment [T3]). The level of the patient's anxiety was classified as *anxiety-free* (score 0), *low anxiety* (scores 1-3), *middle anxiety* (scores 4-6), and *high anxiety* (scores 7-8).

Secondary Outcomes

The secondary outcomes included caregiver anxiety, procedural data, parental usability, child satisfaction, and fun. We also administered a baseline survey to collect information about demographics and clinical characteristics predicted to influence or confound outcomes, such as previous experience with medical imaging.

Caregiver Anxiety

Caregiver anxiety was measured with the short State-Trait Anxiety Inventory (STAI; score 6-24), a 6-item, adapted version of the well-validated Spielberger STAI Scale [45]. This measure has a moderately high retest reliability ranging from 0.65-0.75 and a high degree of internal consistency with a coefficient α of .86-.95 [45]. Parents or caregivers were asked to complete the assessment on a tablet at the same time as children (before preparing [T1], after preparing and upon entering the simulation room for their child's 6-minute *scan* [T2], and after the assessment [T3]).

Procedural Data

Procedural data included preparation and assessment times. In accordance with the experiential learning theory [37], we postulated that children who prepared for longer would have a more efficient assessment. Side effects during preparation (eg, nausea and headaches), dropouts, and noncompliance were also recorded.

The *preparation time* started when the study staff finished describing the preparation program, and the researcher indicated that it was time to begin the active preparation. The preparation time was stopped upon indication that the participant felt ready to take the assessment. A maximum preparation time of 45 minutes was allowed. This time frame was selected as it was

the allotted appointment time provided by our hospital CLP to prepare patients and their families for medical imaging procedures.

The *assessment time* was defined as the time spent in the simulated MRI room until the participant was *discharged* from the simulation experience, either as successful or noncompliant. A maximum assessment time of 20 minutes was allowed. The time spent transitioning between activities or breaks required for reasons unrelated to the study was not accounted as time.

Caregiver Usability

The caregivers of our participants were prompted to provide usability feedback on the preparation materials at the end of the study. Caregivers were asked to complete the USE (Usefulness, Satisfaction, and Ease of Use) Questionnaire administered at the end of the study activities on a tablet. The USE Questionnaire is a 7-point Likert rating scale and is a validated and reliable measure for assessing the subjective usability of a product or service [46]. We limited our testing to the strongest factors of the survey [47].

In addition to this, we asked caregivers if they would be comfortable with their child using the preparation programs and if they had any recommendations for improvement.

Child Satisfaction and Fun

For measuring child satisfaction, participants were asked to indicate how satisfied they were with the preparation program by pointing to a visual analog scale ranging from 0 (terrible) to 100 (fantastic). After this, we asked the children if they would recommend the preparation to a friend who needed an MRI and if there was anything that would make their experience better.

Fun was measured using the Smilyometer Likert Scale, a part of the Fun Toolkit [48]. We selected this as a surrogate measure to inform potential adherence and uptake in the real world, assuming that fun would influence use. The Smileyometer was used before and after the children interacted with a preparation program. The rationale for using it before is that it can measure their expectations and for using it afterward is that it is assumed that the child is reporting experienced fun. It has been widely adopted in testing technologies with children to measure satisfaction and fun as it requires no writing [48-52]. After allocation to a preparatory program, children were asked how good they thought the preparation would be by pointing to the Smilyometer Likert Scale to indicate their expectation of using the intervention. After using the preparation and completing the assessment (timepoint 3 [T3]), the children were again asked how good they thought the preparation actually was using the same Smileyometer Likert scale.

Data Analysis

Power Calculation

A priori power analysis was performed using G*Power 3 [53]. Assuming a small-to-moderate effect size (Cohen f=0.20) with 90% power and the probability of a type 1 error of 0.1, a total sample size of 69 was needed (23 in each group).

Statistical Analysis

Statistical analysis was performed using SPSS (version 22, IBM Corp). Continuous variables were expressed as mean (95% CI) and ordinal variables as median (IQR). Categorical variables were expressed as percentages. Normality conditions were checked for all variables to apply a proper test of significance. Many of the outcome variables were ordinal in nature and were measured in scores. A chi-square test was used to test the independence of association between categorical variables. Analysis of variance (ANOVA; for normal distribution) or the Kruskal–Wallis test (for nonparametric distribution) was applied for 1-way analysis to compare the average scores of the three interventions among the three timepoints. Post hoc Bonferroni analysis was applied to statistically significant findings to confirm the differences between the groups. If the equal variance assumption was not met during the ANOVA process, pairwise comparisons were based on the statistics of Dunnett T3 [54]. To test for relationships between 2 continuous variables, we used the bivariate Pearson correlation. In the case of missing values, a single value was filled for each missing value by averaging the collected scores for each participant.

Results

Demographic and Clinical Characteristics

A total of 92 participants were recruited during the study period; 1 did not consent, and 1 participant did not show up for the appointment. One participant who provided consent initially later withdrew. A total of 89 participants were enrolled. Of the consenting participants, 5% (5/92) were excluded because of equipment malfunction. The remaining 84 participants were included in the analysis (VR-MRI: 30/84, 36%; SPM: 24/84, 29%; CLP: 30/84, 36%).

The demographic and clinical characteristics of the patients are shown in Table 1. Most participants were male (51/84 61%) and had no history of MRI (77/84, 92%) or simulator experience (81/84, 96%). Approximately half the participants (43/84, 51%) had experience with other medical imaging procedures, and many had used VR before participating in the study (70/84, 83%). Chi-square tests were conducted for demographic variables and ANOVA for continuous variables. No significant differences in demographic variables were found among the groups (Table 1).

Table 1. Demographics and clinical characteristics of participants (N=84).

Characteristics	Total population (n=84)	VR-MRI[a] (n=30)	Standard preparatory manual (n=24)	Child Life Program (n=30)	*P* value
Age (years), mean (SD)	9.1 (2.7)	9.3 (2.6)	8.9 (2.8)	9.2 (2.7)	.51
Males, n (%)	51 (61)	18 (60)	13 (54)	20 (67)	.64
History with, n (%)					
Magnetic resonance imaging	7 (8)	3 (10)	3 (12)	1 (3)	.44
Any other medical imaging	43 (51)	13 (43)	12 (50)	15 (50)	.84
Magnetic resonance imaging simulator	3 (4)	2 (7)	—[b]	1 (3)	.42
Virtual reality	70 (83)	24 (80)	18 (75)	25 (83)	.75

[a]VR-MRI: virtual reality magnetic resonance imaging.

[b]Not available.

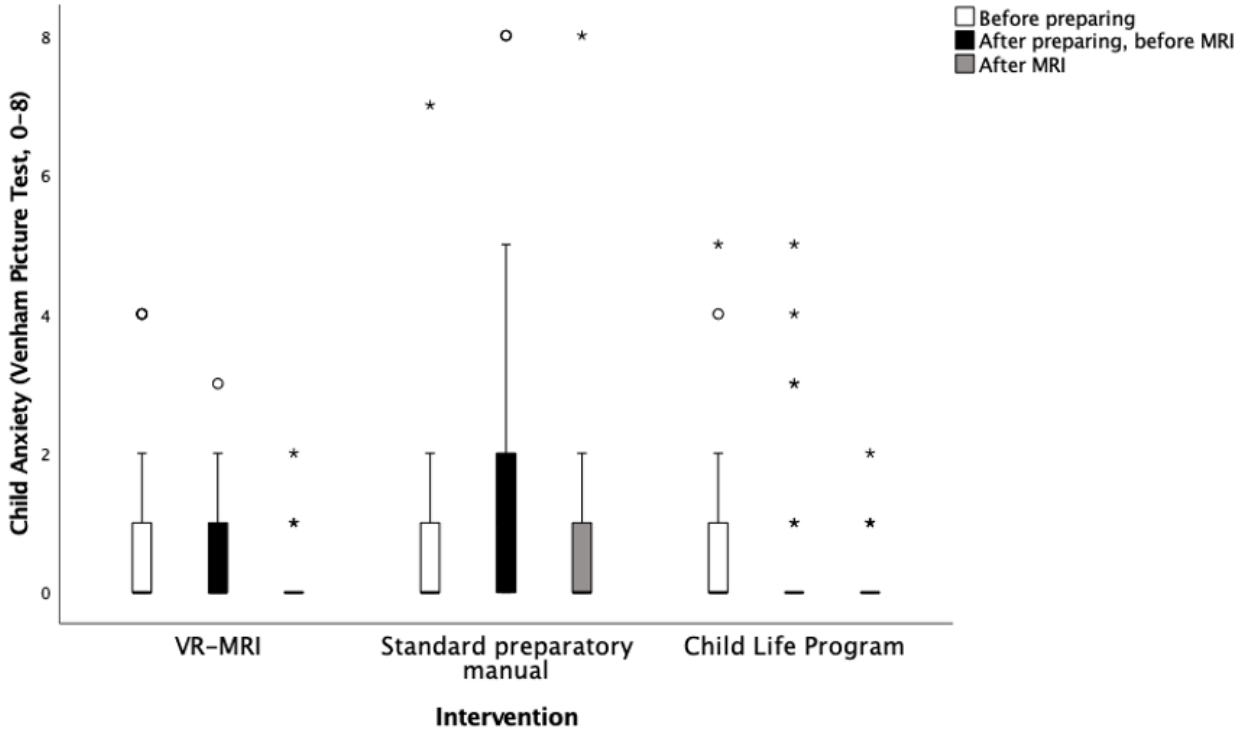
Success in the MRI Simulation Experience

Success was indicated if participants were able to complete a 6-minute head scan without surpassing 4 mm of movement at any of the 3300 data points collected. The average number of times participants scored above the threshold for the 3300 data points was not statistically significant between the groups (χ^2_2=2.7; *P*=.07). Similarly, no statistically significant differences were found when calculating success in the simulated MRI experience among VR-MRI, SPM, and CLP (χ^2_2=2.6; *P*=.27). On average, 30% (95% CI 13%-47%) participants in the VR-MRI group were successful, compared to 50% (95% CI 28%-72%) in the SPM group and 47% (95% CI 28%-66%) in the CLP group. Of the participants who failed, 8 (VR-MRI: 2/8, 25%; SPM: 3/8, 37%; CLS: 2/8, 25%) were noncompliant by declining to complete some or all of the requirements of the scan, automatically failing their assessment.

Child Anxiety

VPT was used to determine how anxious children were before preparing (timepoint 1 [T1]), after preparing and upon entering the MRI simulation room (timepoint 2 [T2]), and after completing the assessment [T3]). Participants reported that anxiety remained relatively stable, with no *clinically* significant differences between the groups at any timepoint (Figure 12; T1, *P*=.99; T2, *P*=.008; T3, *P*=.10). On average, children in the VR-MRI group reported being anxiety-free before preparation (median 0, IQR 1; SD 1.311), after preparation (median 0, IQR 1; SD 0.819), and after the assessment (median 0, IQR 1; SD 0.434). Children in the SPM group reported being anxiety-free before preparation (median 0, IQR 1; SD 1.521), having low anxiety after preparation (median 1, IQR 2; SD 2.311), and being anxiety-free again after the assessment (median 0, IQR 1; SD 1.738). Finally, children in the CLP group reported being anxiety-free before preparation (median 0, IQR 0; SD 1.240), after preparation (median 0, IQR 0; SD 1.350), and after the assessment (median 0, IQR 0; SD 0.468).

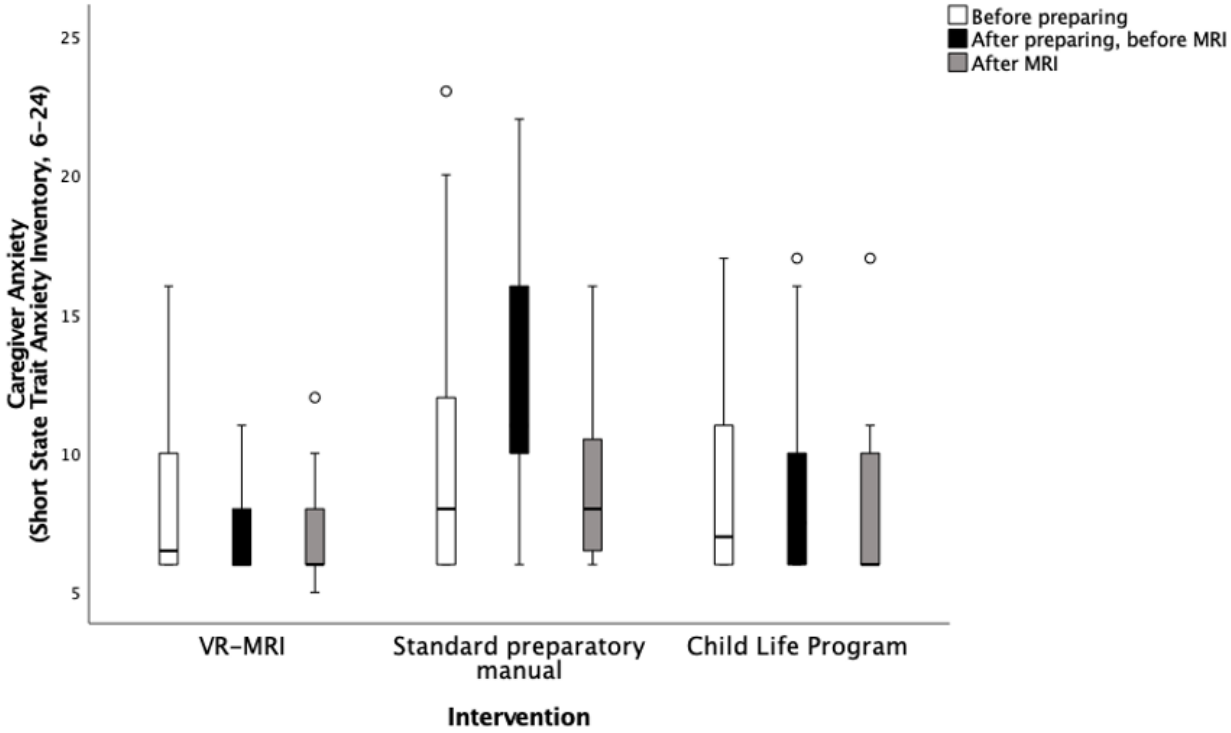
Figure 12. Comparison of self-reported child anxiety across groups and timepoints, measured by the Venham picture test (the circles denote outliers and the asterisks denote extreme outliers). No clinically significant results were indicated (timepoint 1, *P*=.99; timepoint 2, *P*=.008; timepoint 3, *P*=.10). MRI: magnetic resonance imaging; VR-MRI: virtual reality magnetic resonance imaging.



Caregiver Anxiety

The short STAI was used to determine how anxious caregivers were before preparing (T1), after preparing and upon entering the MRI simulation room (T2), and after completing the assessment (T3). As depicted in Figure 13, caregivers using the manual to prepare were significantly more anxious after preparing than caregivers in the VR-MRI (mean difference 5.33, 95% CI 2.93-7.74; Dunnett *P*<.001) and CLP groups (mean difference 3.73, 95% CI 1.07-6.40; Dunnett *P*=.004). The effect size was large (η^2=0.319, 95% CI 0.15-0.448). No clinically significant differences between the groups were found before preparing (T1) or after completing the assessment (T3). Caregivers in the VR-MRI group reported low anxiety before (median 6.5, IQR 4; SD 2.572) and after preparation (median 6, IQR 2; SD 1.744) and after the assessment (median 6, IQR 2; SD 1.810). Caregivers in the manual group reported low anxiety before preparation (median 8, IQR 6; SD 4.945), an increase in anxiety after preparation (median 10, IQR 6; SD 4.394), and again low anxiety after the assessment (median 8, IQR 4; SD 2.924). Finally, caregivers in the CLP group reported low anxiety before (median 7, IQR 6; SD 3.294) and after preparation (median 8, IQR 4; SD 3.224) and after the assessment (median 6, IQR 4; SD 2.545).

Figure 13. Comparison of change in self-reported caregiver anxiety between timepoints and across groups, measured with the short State-Trait Anxiety Inventory (the circles denote outliers and the asterisks denote extreme outliers). Clinically significant differences were found between the manual group in comparison with the virtual reality magnetic resonance imaging (Dunnett *P*<.001) and the Child Life Program (Dunnett *P*=.004). MRI: magnetic resonance imaging; VR-MRI: virtual reality magnetic resonance imaging.



We conducted a bivariate Pearson correlation analysis to determine whether child anxiety was related to caregiver anxiety. After preparation and upon entering the MRI simulator room, child and caregiver anxiety had a statistically significant linear relationship (r_{84}=0.421; *P*<.001), indicating that child anxiety increased with parental anxiety. The strength of this relationship was moderate.

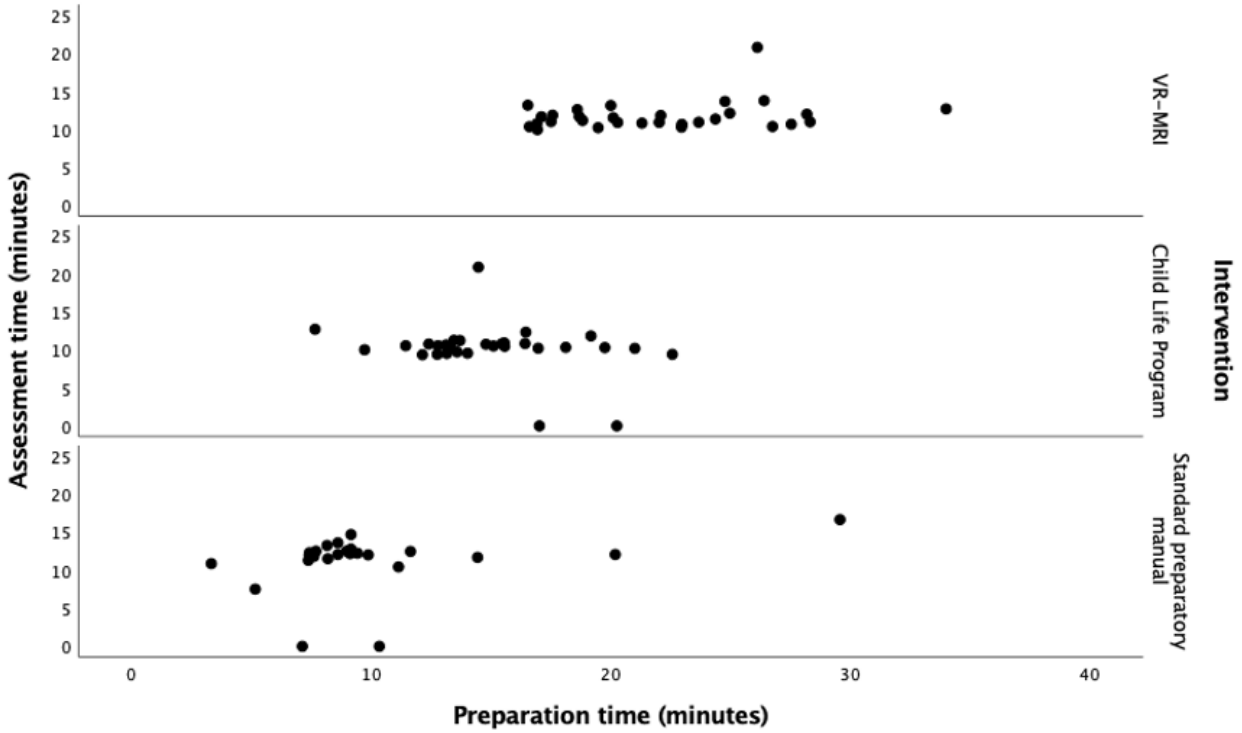
Procedural Data

Time was recorded during preparation and assessment. There was a significant difference in preparation times among all groups ($F_{2.81}$=53.261; Dunnett *P*<.001). The effect size was large (η^2=0.568, 95% CI 0.415-0.660). Participants in the VR-MRI group prepared for the longest time, which was 22.05 minutes (SD 4.41; 95% CI 20.40-23.69), compared with the CLP group, which was in the middle at 15.06 minutes (SD 3.32; 95% CI 13.82-16.30), and the SPM group, which prepared for the shortest time, at 9.98 minutes (SD 5.24; 95% CI 7.76-12.20).

No statistically significant differences were found among the groups in assessment time ($F_{2.81}$=2.063; *P*=.13). Participants in the VR-MRI group, on average, took 11.79 minutes (SD 1.99; 95% CI 11.05-12.55), compared with the SPM group who took 11.19 minutes (SD 3.80; 95% CI 9.58-12.78) and the CLP group who took 10.17 minutes (SD 3.44; 95% 8.89-11.45) to complete their assessment (regardless of success).

We conducted a bivariate Pearson correlation test to determine whether the time spent preparing affected the efficiency of the assessment. Preparation and assessment times did not have a statistically significant linear relationship (r=0.148; *P*=.18), indicating that preparation time did not have a significant effect on assessment time in our study (Figure 14).

Figure 14. Comparison of preparation and assessment times across groups, measured in minutes. No significant relationship was found (r=0.148; *P*=.18). VR-MRI: virtual reality magnetic resonance imaging.



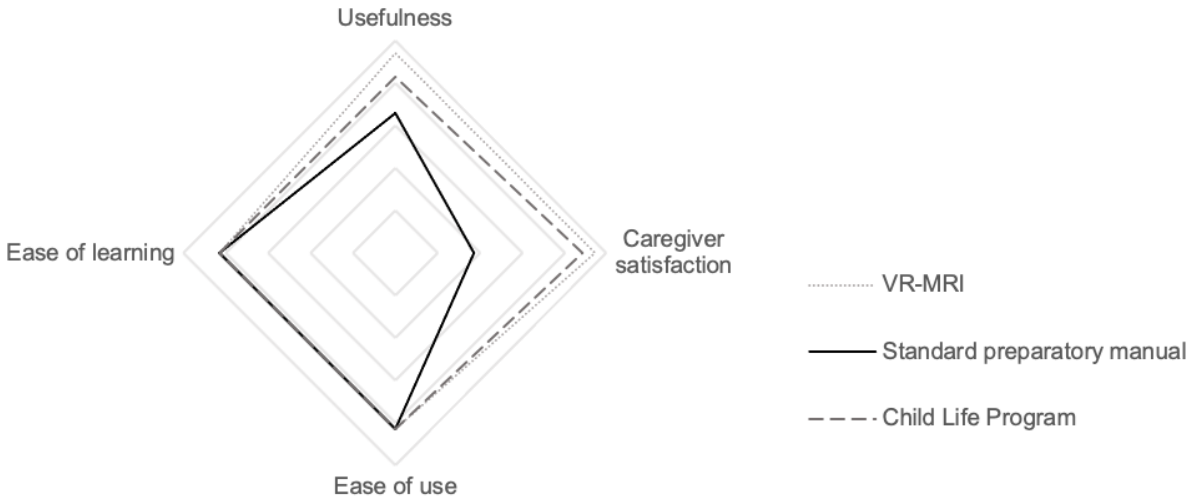
One child (aged 5 years) indicated eye strain and a blurry image when viewing VR-MRI that could not be mitigated by interpupillary adjustments. Two children (aged 6 and 4 years) reported the dinosaur graphic in VR-MRI was *scary*. Six participants (aged 4-12 years) in the manual group expressed being scared of pictures in the manual, particularly the sections of the intravenous or coil pictures. No other side effects were reported.

Caregiver Usability

Caregivers were asked to complete the USE Questionnaire [46] to report how easy the preparation materials were to learn and use, as well as how useful and satisfied they were with using them to prepare their children for the simulated MRI experience.

Caregivers did not report significant differences in ease of use (χ^2_2=0.01; *P*=.99), ease of learning (χ^2_2=1.5; *P*=.48), or usefulness (χ^2_2=4.4; *P*=.11) among the groups. On average, caregivers using VR-MRI to prepare their child agreed that it was useful (median 31, IQR 4; SD 3.562), easy to use (median 24, IQR 3; SD 2.448), and easy to learn (median 18, IQR 2; SD 2.366). As depicted in Figure 15, caregivers using CLP, on average, also *agreed* that it was useful (median 30, IQR 6; SD 4.163), easy to use (median 24, IQR 4; SD 2.937), and easy to learn (median 18, IQR 3; SD 3.059). Caregivers using SPM *somewhat agreed* that it was useful (median 28.5, IQR 6; SD 3.323) and *agreed* that it was easy to use (median 24, IQR 5; SD 2.167) and easy to learn (median 18, IQR 3; SD 2.183).

Figure 15. Comparison of median usability metrics collected across groups, adjusted to total scores of 100. Significant differences were found for caregiver satisfaction when comparing the manual with the virtual reality magnetic resonance imaging and Child Life Program (Bonferroni *P*<.001). VR-MRI: virtual reality magnetic resonance imaging.



However, when reporting satisfaction with the preparation materials, caregivers using VR-MRI and CLP were significantly more satisfied than the caregivers using SPM (Dunnett *P*<.001). The effect size was large (η^2=0.268, 95% CI 0.104-0.402). Caregivers in the VR-MRI (median 31, IQR 4; SD 2.837) and CLP (median 30.5, IQR 6; SD 4.173) groups reported high satisfaction. However, caregivers in the SPM group were only somewhat satisfied (median 26, IQR 6; SD 2.428).

When asked if they would be comfortable with their child using the preparation materials for an actual MRI, 100% of participants in all groups indicated *yes*. In terms of improvements, caregivers in the VR-MRI group indicated they wanted the ability to explore VR-MRI in tandem with their child and wanted more information about when their child would be ready for a successful MRI. Caregivers in the manual group recommended tailoring the manual to the specific procedures being conducted, offering more instruction to parents on how to coach their child through the MRI experience, and offering more interactivity (eg, videos and stickers) to guide children through the steps. Caregivers in the CLP group were pleased and had no recommendations for improvement.

Child Satisfaction and Fun

Child satisfaction is reported in Table 2. Children were asked to point to a visual analog scale, indicating their satisfaction with the preparation materials. Children in the SPM group were significantly less satisfied than those in the CLP group (mean difference −17.509, 95% CI −34.85 to −0.17; Dunnett *P*=.047). On average, children in the manual group were 73.5% (SD 27%; IQR 37%) satisfied compared with children in the CLP group who were 90% (SD 12%; IQR 23%) satisfied. Children in the VR-MRI group were 80% (SD 27%; IQR 22%) satisfied.

Table 2. Comparison of self-reported child satisfaction scores between groups, measured with a visual analog scale.

Characteristics	Participant (n=77), n (%)	Median score (IQR; SD)	Mean rank	Chi-square (*df*)	*P* value
Timepoint 1					
VR-MRI[a]	30 (39)	80 (22; 27)	36.15	7.7 (2)	.02[b]
Standard preparatory manual	20 (26)	73.5 (37; 27)	30.90	—[c]	—
Child Life Program	27 (35)	90 (23;12)	47.17	—	—

[a]VR-MRI: virtual reality magnetic resonance imaging.

[b]Statistically significant difference between study arms (Kruskal–Wallis, *P*<.05).

[c]Not available.

The Smileyometer was used to determine how fun children expected the programs to be and how fun they actually were after completing them. We did not find any statistically significant differences in how much fun children expected the preparation to be ($F_{2.81}$=2.224; *P*=.11). On average, the children in the VR-MRI and CLP groups thought the preparation would be *really good*, whereas children in the manual group thought the preparation would be *okay*.

When asked if they would recommend the preparation materials to a friend, 93% (28/30) of participants in the VR-MRI group said yes compared with 80% (22/27) in the CLP group and only 71% (14/20) in the manual group. Participants in the VR-MRI group thought the game was too difficult and that they spent too much time staying still. The participants reported that they wanted more options in terms of selecting a character in the experience and tailoring the language to their age group. Participants in the manual group recommended only including

the procedures they would encounter and providing more interactivity. The participants in the CLP group also commented on the length of time required to stay still and the importance of listening to the correct volume and repetition of the sounds.

Discussion

Effectiveness of VR-MRI in Preparing Children for the Nonsedated MRI Experience

In this study, we compared a VR preparation program (VR-MRI) with the hospital CLP and SPM for reducing anxiety and noncompliance during a simulated MRI scan of the head. To our knowledge, no other study has compared VR with these alternatives for nonsedated MRI using both anxiety and success as outcomes.

Although VR has been studied in some preprocedural situations, including anesthesiology and medical imaging, its effectiveness in comparison with mock-MRIs [15,16,55-57] and other behavioral techniques [4,12] used by CLPs has not been well examined [12]. We found no significant differences between VR-MRI, our hospital CLP, and our hospital SPM on our measure of success during a simulated MRI experience (P=.27). This finding is comparable with that of another recent study [56] which found cheaper versions of MRI simulations adequate in most cases to provide the desensitization and practice needed for a successful nonsedated pediatric MRI scan. VR-MRI adds to the options for affordable and viable alternatives discussed in recent reviews [12]. It can be used at home or in remote settings for further practice and can easily be stored when not in use. Thus, health care centers could use a digital preparatory sequence leveraging immersive media, such as our VR-MRI app, to prepare pediatric patients aged between 4 and 13 years without the financial, travel, and space requirements of traditional in-person MRI simulations.

VR-MRI and Child Anxiety

In addition to finding no differences in our measure of success, we also found no clinically significant differences among children's reported anxiety when using VR-MRI compared with the other standards of care (T1, P=.99; T2, P=.008; T3, P=.10). Other studies have reported low anxiety during medical imaging simulations [30]. The findings of this study and ours question the relationship suggested by many studies (that anxiety directly influences motion artifacts). Although our study participants reported low anxiety, many were still unsuccessful in the simulated MRI experience. A patient's understanding is another element that can contribute to motion artifacts [58]. However, our findings also question this as we did not find any relationship between success and prior experience with medical imaging. Our sample for this subset was quite small, which could explain why no relationship was found. Future studies could consider measuring heart rate and cortisol as objective measures of stress throughout the training and procedure, as well as a measure of patient understanding to further explore these concepts.

VR-MRI and Caregiver Anxiety

Whereas other research has focused primarily on patient anxiety, our study uniquely evaluated caregivers' anxiety during the process of preparing for and completing a simulated MRI experience. VR-MRI was not statistically different from CLP and was better than SPM in mitigating caregiver-reported anxiety when entering the MRI simulation room (P<.001). We also found that caregiver and child anxiety tended to increase together after preparation (P<.001). Given that it is generally accepted that clinicians provide the opportunity for parents to be present during their child's procedure [59], our study results suggest that preparation is just as beneficial to caregivers as it is to the children undergoing MRI procedures. To assist in caregiver preparation, we mirrored the experience on a tablet for parents to experience it at the same time as the children. To our knowledge, effective ways of engaging caregivers in the preparation process have not yet been studied, particularly in the context of VR.

Time Spent Preparing and Conducting Head Scans

Participants prepared the longest with VR-MRI and the shortest with SPM. This was largely expected given that VR-MRI was a standardized experience compared with the experience of the other two groups that received interventions tailored to the participant by the CCLS (in the CLP group) or the caregiver (in the SPM group). It is interesting to note that participants engaged in programs that provided opportunities for experiential learning (VR-MRI and CLP) for a longer duration than those that offered through didactic learning (SPM). However, these novel elements of preparation, which allowed patients to engage and participate in a practice MRI experience, did not result in differences in success. We did not find any significant differences in the assessment times across the groups (P=.13).

We did not continue to conduct training and assessments with participants until we obtained an appropriate image quality, which would be critical for clinical care scenarios. In fact, many of the children failed the assessments in our study, suggesting that the assessment was actually not efficient, and more training or sedation would be required for these children. An important element of experiential learning is the opportunity to reflect on experiences [37]. As our study comprised a single session on the same day as the assessment, and children aged ≥6 years benefit most if they participate in preparation programs for ≥5 days in advance of the procedure [4], offering VR preparation before a medical imaging procedure and with multiple opportunities to experience, reflect, and practice should be explored by future investigators. In accordance with patient-centered care, children may also benefit from interacting with different practice materials depending on their individual preferences. We echo other investigators in stating that deploying multiple strategies together may provide the best way forward to improve nonsedated medical imaging outcomes and experiences [12].

Experiences With VR-MRI

Caregivers' Experiences

As reported by other investigators [27,29-31], experiences using VR-MRI were positive in our study. Our data indicate that compared with CLP and SPM, the VR-MRI app was just as easy to use (P=.99), easy to learn (P=.48), and useful (P=.11). The technical concerns commonly reported as barriers to perceived ease of use of VR technologies [60] were not

mentioned by the caregivers in our study. In fact, caregivers were significantly more satisfied with VR-MRI than SPM (P<.001). It is important to note that SPM was the only preparation program to require support from the caregiver as the researcher assisted in the setup for the VR-MRI group, and the CCLS assisted in the CLP group. As such, we may not have seen the same result if participants were required to set up the hardware themselves without assistance. Nonetheless, VR-MRI was perceived as a valued and acceptable form of preparation by caregivers.

Children's Experiences

In addition to caregiver experiences, we also measured the children's experiences using VR to prepare for their simulated head scan. We only found significant differences in children's reported satisfaction between SPM and CLP (P<.001) groups. Children using the VR-MRI app generally thought it was *really good* for preparing them for the MRI assessment, meeting the children's expectations for fun. The results of the app meeting expectations are likely influenced by the VR-MRI design, which includes obtaining multiple perspectives for identifying the needs and values of pediatric patients undergoing medical imaging procedures [34,35]. However, it is interesting that there were no significant differences among VR-MRI and the other 2 groups in terms of fun. Other investigators using gamification in VR for preoperative anxiety in pediatric patients undergoing general anesthesia also found that satisfaction was not significantly different between their gamified and control groups [24]. The result suggests that integrating novel technologies does not inherently make a procedural preparation *fun*, and further product and design elements are required to create an optimal experience. On the basis of the feedback from children and their caregivers, the VR-MRI game design may need refinements in terms of tailoring the challenge of staying still to the user's initial skills, as well as more options for customization and immersion to create an optimal experience [21,34,39].

Product and Design

The selection of preprocedural scenarios and hardware that are best for integration with VR is not well understood and has not been the focus of previous pediatric studies [21,39]. In our study, we elected to conduct a head scan as it was one of the most common and anxiety-provoking scans at our hospital. However, pragmatically, the clinical characteristics of patients who require head scans may not be the most appropriate for the use of VR headsets as a significant proportion of patients who require head scans have traumatic brain injuries or are potentially epileptic (which typically precludes them from using VR). In addition, experiences could be affected by a patient's mobility of the head and neck when using HMDs. Further research is needed for safe use with these patients and may include delivering the program through immersive videos rather than with a headset. The program must also be adapted to make it available on other platforms (Android and iOS) and devices so that it can be versatile in deployment [39]. The lack of appropriate patients to recommend VR to had a significant impact on the uptake of VR interventions in other clinical settings [60] and should be an important consideration for development.

Our study responds to the literature calling for studies on products that engage pediatric patients and explore skill-building goals [34]. We introduce intervention and design elements that provide dynamic feedback to the patient and experiential learning to regulate movement in preparation for a nonsedated MRI. We designed and tested a virtual MRI with real-time feedback to explore skill building and introduced some product qualities that enabled the participants to be active in preparing themselves for the MRI experience. Our design considerations included tailoring simplicity and interactivity to improve control, improving a sense of presence, creating a sense of safety through familiar design elements and medical procedures, incorporating narrative elements, and cultivating growth and motivation [34]. The real-time feedback feature, facilitated through a mobile phone gyroscope, is a unique gamification element that aligns with anxiety management strategies [34]; to our knowledge, this has not been reported in the literature to date. Future apps should focus on refining and validating real-time feedback gaming elements in medical imaging preparation so that it could be used as a decision-making tool that informs parents and health care providers about when and if a patient might be ready to attempt a nonsedated MRI, thus reducing the burdens associated with anxiety and noncompliance.

Limitations

There are several potential limitations to this study. Our methodology focused on self-reporting of anxiety of children and caregivers. We used the short STAI for caregivers but VPT for children, which may have introduced confounding factors. The results may also have been affected if users did not fully understand the meaning or how to complete the surveys after the instructions. There may have been a response bias, as children often consciously or subconsciously give responses that they think adults want to hear.

Our study also had several additional biases. The study was subject to information and selection biases, as we recruited participants through posters at the hospital, public libraries, and social media and provided remuneration and parking reimbursements. Motivation and reported outcomes related to using the materials could have been affected by these extrinsic motivations (eg, remuneration). The study was also unblinded to the participants and research staff because of the practicalities of the preparatory processes and logistical limitations.

Our study had a small sample size that met the requirements of our power calculation. The effect sizes were smaller among groups than we anticipated, and, therefore, this study is at risk for type 2 error (accepting a null hypothesis that is actually false). A total of 5 participants did not have movement metrics because of technological malfunction. In our study, we used the MERGE VR headset as no other HMDs have been indicated for use specifically with children. This headset is indicated to match the interpupillary distance of children aged ≥10 years. Younger children may have smaller interpupillary distances than can be adjusted for. Eye strain and a blurry image that could not be mitigated by adjustment were reported by 1 study participant (aged 5 years), which is likely a result of that limitation. Currently available consumer-grade VR hardware has not been designed for use with younger children and, in

some cases, might not be adjustable for the parameters required by them.

Conclusions

The VR preparatory program with a novel real-time feedback feature had comparable findings with our hospital CLP and SPM in successfully preparing children to complete a simulated MRI experience. Furthermore, VR-MRI mitigated situational anxiety in children throughout the process of preparation and completion of a simulated MRI head scan. As such, digital preparatory apps that leverage immersive media, such as our VR-MRI app, may be a viable alternative for preparing children for nonsedated MRI. Further research is required to confirm the findings with actual pediatric patients in a real MRI machine, as our study used a high-fidelity MRI simulator. Nevertheless, the use of VR and the sequence of activities provided through the VR-MRI app show promise for preprocedural anxiety reduction in children and their caregivers.

Acknowledgments

The authors thank all the study participants, radiology and patient experience staff, research team members, and the volunteers who helped support the research activities: Aria Rezapour, Monique Sangha, Amit Sharma, and Jinny Choi.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Standard preparatory manual.
[PDF File (Adobe PDF File), 3096 KB - jmir_v23i9e22942_app1.pdf]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 100 KB - jmir_v23i9e22942_app2.pdf]

References

1. Kain Z, Mayes L, O'Connor TZ, Cicchetti D. Preoperative anxiety in children. Predictors and outcomes. Arch Pediatr Adolesc Med 1996 Dec;150(12):1238-1245. [doi: 10.1001/archpedi.1996.02170370016002] [Medline: 8953995]
2. Ahmed MI, Farrell MA, Parrish K, Karla A. Preoperative anxiety in children risk factors and non-pharmacological management. Middle East J Anaesthesiol 2011 Jun;21(2):153-164. [Medline: 22435267]
3. Alanazi AA. Reducing anxiety in preoperative patients: a systematic review. Br J Nurs 2014 Sep;23(7):387-393 [FREE Full text] [doi: 10.12968/bjon.2014.23.7.387] [Medline: 24732993]
4. McCann ME, Kain ZN. The management of preoperative anxiety in children: an update. Anesth Analg 2001 Jul;93(1):98-105. [doi: 10.1097/00000539-200107000-00022] [Medline: 11429348]
5. Tyc VL, Fairclough D, Fletcher B, Leigh L, Mulhern RK. Children's distress during magnetic resonance imaging procedures. Child Health Care 1995;24(1):5-19. [doi: 10.1207/s15326888chc2401_2] [Medline: 10142086]
6. Vanderby SA, Babyn PS, Carter MW, Jewell SM, McKeever PD. Effect of anesthesia and sedation on pediatric MR imaging patient flow. Radiology 2010 Jul;256(1):229-237. [doi: 10.1148/radiol.10091124] [Medline: 20505061]
7. Saunders R, Davis JA, Kranke P, Weissbrod R, Whitaker DK, Lightdale JR. Clinical and economic burden of procedural sedation-related adverse events and their outcomes: analysis from five countries. Ther Clin Risk Manag 2018 Feb 28;14:393-401 [FREE Full text] [doi: 10.2147/TCRM.S154720] [Medline: 29535525]
8. Marshall SP, Smith MS, Weinberger E. Perceived anxiety of pediatric patients to magnetic resonance. Clin Pediatr (Phila) 1995 Jan;34(1):59-60. [doi: 10.1177/000992289503400114] [Medline: 7720333]
9. Alexander M. Managing patient stress in pediatric radiology. Radiol Technol 2012;83(6):549-560. [Medline: 22763832]
10. Malviya S, Voepel-Lewis T, Prochaska G, Tait AR. Prolonged recovery and delayed side effects of sedation for diagnostic imaging studies in children. Pediatrics 2000 Mar;105(3):E42. [doi: 10.1542/peds.105.3.e42] [Medline: 10699144]
11. Durand DJ, Young M, Nagy P, Tekes A, Huisman TA. Mandatory child life consultation and its impact on pediatric MRI workflow in an academic medical center. J Am Coll Radiol 2015 Jun;12(6):594-598. [doi: 10.1016/j.jacr.2014.12.015] [Medline: 25912290]
12. McGuirt D. Alternatives to sedation and general anesthesia in pediatric magnetic resonance imaging: a literature review. Radiol Technol 2016 Sep;88(1):18-26. [Medline: 27601689]
13. Committee on Hospital CareChild Life Council. Child life services. Pediatrics 2014 May;133(5):e1471-e1478. [doi: 10.1542/peds.2014-0556] [Medline: 24777212]
14. Edwards AD, Arthurs OJ. Paediatric MRI under sedation: is it necessary? What is the evidence for the alternatives? Pediatr Radiol 2011 Nov;41(11):1353-1364. [doi: 10.1007/s00247-011-2147-7] [Medline: 21678113]

15. Mastro KA, Flynn L, Millar TF, DiMartino TM, Ryan SM, Stein MH. Reducing anesthesia use for pediatric magnetic resonance imaging: the effects of a patient- and family-centered intervention on image quality, health-care costs, and operational efficiency. J Radiol Nurs 2019 Mar;38(1):21-27. [doi: 10.1016/j.jradnu.2018.12.003]
16. Carter AJ, Greer MC, Gray SE, Ware RS. Mock MRI: reducing the need for anaesthesia in children. Pediatr Radiol 2010 Aug;40(8):1368-1374. [doi: 10.1007/s00247-010-1554-5] [Medline: 20186541]
17. Cavarocchi E, Pieroni I, Serio A, Velluto L, Guarnieri B, Sorbi S. Kitten Scanner reduces the use of sedation in pediatric MRI. J Child Health Care 2019 Jun;23(2):256-265. [doi: 10.1177/1367493518788476] [Medline: 30049224]
18. Copanitsanou P, Valkeapää K. Effects of education of paediatric patients undergoing elective surgical procedures on their anxiety - a systematic review. J Clin Nurs 2014 Apr;23(7-8):940-954. [doi: 10.1111/jocn.12187] [Medline: 23581540]
19. Wright KD, Stewart SH, Finley GA, Buffett-Jerrott SE. Prevention and intervention strategies to alleviate preoperative anxiety in children: a critical review. Behav Modif 2007 Jan;31(1):52-79. [doi: 10.1177/0145445506295055] [Medline: 17179531]
20. Hartman JH, Bena J, McIntyre S, Albert NM. Does a photo diary decrease stress and anxiety in children undergoing magnetic resonance imaging? A randomized, controlled study. J Radiol Nurs 2009 Dec;28(4):122-128. [doi: 10.1016/j.jradnu.2009.08.002]
21. Eijlers R, Utens E, Staals LM, de Nijs PF, Berghmans JM, Wijnen RM, et al. Systematic review and meta-analysis of virtual reality in pediatrics: effects on pain and anxiety. Anesth Analg 2019 Nov;129(5):1344-1353 [FREE Full text] [doi: 10.1213/ANE.0000000000004165] [Medline: 31136330]
22. Parsons TD, Riva G, Parsons S, Mantovani F, Newbutt N, Lin L, et al. Virtual reality in pediatric psychology. Pediatrics 2017 Nov;140(Suppl 2):86-91 [FREE Full text] [doi: 10.1542/peds.2016-1758I] [Medline: 29093039]
23. Ganry L, Hersant B, Sidahmed-Mezi M, Dhonneur G, Meningaud J. Using virtual reality to control preoperative anxiety in ambulatory surgery patients: a pilot study in maxillofacial and plastic surgery. J Stomatol Oral Maxillofac Surg 2018 Sep;119(4):257-261. [doi: 10.1016/j.jormas.2017.12.010] [Medline: 29317347]
24. Ryu J, Park J, Nahm F, Jeon Y, Oh A, Lee H, et al. The effect of gamification through a virtual reality on preoperative anxiety in pediatric patients undergoing general anesthesia: a prospective, randomized, and controlled trial. J Clin Med 2018 Sep 17;7(9):284 [FREE Full text] [doi: 10.3390/jcm7090284] [Medline: 30227602]
25. Bekelis K, Calnan D, Simmons N, MacKenzie TA, Kakoulides G. Effect of an immersive preoperative virtual reality experience on patient reported outcomes: a randomized controlled trial. Ann Surg 2017 Jun;265(6):1068-1073. [doi: 10.1097/SLA.0000000000002094] [Medline: 27906757]
26. Eijlers R, Legerstee JS, Dierckx B, Staals LM, Berghmans J, van der Schroeff MP, et al. Development of a virtual reality exposure tool as psychological preparation for elective pediatric day care surgery: methodological approach for a randomized controlled trial. JMIR Res Protoc 2017 Sep 11;6(9):e174 [FREE Full text] [doi: 10.2196/resprot.7617] [Medline: 28893727]
27. Liszio S, Masuch M. Virtual reality MRI: playful reduction of children's anxiety in MRI exams. In: Proceedings of the 2017 Conference on Interaction Design and Children. 2017 Jun Presented at: IDC '17: Proceedings of the 2017 Conference on Interaction Design and Children; June 27-30, 2017; Stanford California p. 127-136. [doi: 10.1145/3078072.3079713]
28. Ashmore J, Di Pietro J, Williams K, Stokes E, Symons A, Smith M, et al. A free virtual reality experience to prepare pediatric patients for magnetic resonance imaging: cross-sectional questionnaire study. JMIR Pediatr Parent 2019 Apr 18;2(1):e11684 [FREE Full text] [doi: 10.2196/11684] [Medline: 31518319]
29. Brown RK, Petty S, O'Malley S, Stojanovska J, Davenport MS, Kazerooni EA, et al. Virtual reality tool simulates MRI experience. Tomography 2018 Sep;4(3):95-98 [FREE Full text] [doi: 10.18383/j.tom.2018.00023] [Medline: 30320208]
30. Nakarada-Kordic I, Reay S, Bennett G, Kruse J, Lydon A, Sim J. Can virtual reality simulation prepare patients for an MRI experience? Radiography (Lond) 2020 Aug;26(3):205-213. [doi: 10.1016/j.radi.2019.11.004] [Medline: 32052767]
31. Ryu J, Park S, Park J, Kim J, Yoo H, Kim T, et al. Randomized clinical trial of immersive virtual reality tour of the operating theatre in children before anaesthesia. Br J Surg 2017 Nov;104(12):1628-1633. [doi: 10.1002/bjs.10684] [Medline: 28975600]
32. Hendricks TM, Gutierrez CN, Stulak JM, Dearani JA, Miller JD. The use of virtual reality to reduce preoperative anxiety in first-time sternotomy patients: a randomized controlled pilot trial. Mayo Clin Proc 2020 Jun;95(6):1148-1157. [doi: 10.1016/j.mayocp.2020.02.032] [Medline: 32498771]
33. Li J, Li Q, Dai X, Li J, Zhang X. Does pre-scanning training improve the image quality of children receiving magnetic resonance imaging?: a meta-analysis of current studies. Medicine (Baltimore) 2019 Feb;98(5):e14323 [FREE Full text] [doi: 10.1097/MD.0000000000014323] [Medline: 30702613]
34. Ahmadpour N, Keep M, Janssen A, Rouf AS, Marthick M. Design strategies for virtual reality interventions for managing pain and anxiety in children and adolescents: scoping review. JMIR Serious Games 2020 Jan 31;8(1):e14565 [FREE Full text] [doi: 10.2196/14565] [Medline: 32012042]
35. Ahmadpour N, Weatherall AD, Menezes M, Yoo S, Hong H, Wong G. Synthesizing multiple stakeholder perspectives on using virtual reality to improve the periprocedural experience in children and adolescents: survey study. J Med Internet Res 2020 Jul 17;22(7):e19752 [FREE Full text] [doi: 10.2196/19752] [Medline: 32706671]
36. Birckhead B, Khalil C, Liu X, Conovitz S, Rizzo A, Danovitch I, et al. Recommendations for methodology of virtual reality clinical trials in health care by an international working group: iterative study. JMIR Ment Health 2019 Jan 31;6(1):e11973 [FREE Full text] [doi: 10.2196/11973] [Medline: 30702436]

37. Kolb D, Boyatzis R, Mainemelis C. Experiential learning theory: previous research new directions. In: Perspectives on Thinking, Learning, and Cognitive Styles. Oxfordshire England: Routledge; 2001.
38. Bandura A. Health promotion from the perspective of social cognitive theory. Psychol Health 1998 Jul;13(4):623-649. [doi: 10.1080/08870449808407422]
39. O'Sullivan B, Alam F, Matava C. Creating low-cost 360-degree virtual reality videos for hospitals: a technical paper on the dos and don'ts. J Med Internet Res 2018 Jul 16;20(7):e239 [FREE Full text] [doi: 10.2196/jmir.9596] [Medline: 30012545]
40. Wilson K, Bell C, Wilson L, Witteman H. Agile research to complement agile development: a proposal for an mHealth research lifecycle. NPJ Digit Med 2018 Sep 13;1:46 [FREE Full text] [doi: 10.1038/s41746-018-0053-1] [Medline: 31304326]
41. Greene DJ, Koller JM, Hampton JM, Wesevich V, Van AN, Nguyen AL, et al. Behavioral interventions for reducing head motion during MRI scans in children. Neuroimage 2018 May 01;171:234-245 [FREE Full text] [doi: 10.1016/j.neuroimage.2018.01.023] [Medline: 29337280]
42. Vanderwal T, Kelly C, Eilbott J, Mayes LC, Castellanos FX. Inscapes: a movie paradigm to improve compliance in functional magnetic resonance imaging. Neuroimage 2015 Nov 15;122:222-232 [FREE Full text] [doi: 10.1016/j.neuroimage.2015.07.069] [Medline: 26241683]
43. Venham LL, Gaulin-Kremer E. A self-report measure of situational anxiety for young children. Pediatr Dent 1979 Jun;1(2):91-96. [Medline: 399677]
44. Kim J, Chiesa N, Raazi M, Wright KD. A systematic review of technology-based preoperative preparation interventions for child and parent anxiety. Can J Anaesth 2019 Aug;66(8):966-986. [doi: 10.1007/s12630-019-01387-8] [Medline: 31098960]
45. Marteau TM, Bekker H. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). Br J Clin Psychol 1992 Sep 12;31(3):301-306. [doi: 10.1111/j.2044-8260.1992.tb00997.x] [Medline: 1393159]
46. Gao M, Kortum P, Oswald F. Psychometric evaluation of the USE (Usefulness, Satisfaction, and Ease of use) questionnaire for reliability and validity. Proc Hum Factors Ergon Soc Annu Meet 2018 Sep 27;62(1):1414-1418. [doi: 10.1177/1541931218621322]
47. Lund A. Measuring usability with the USE questionnaire. Usab Interf 2001 Jan;8(2):3-6 [FREE Full text]
48. Read JC. Validating the Fun Toolkit: an instrument for measuring children's opinions of technology. Cogn Tech Work 2007 May 22;10(2):119-128. [doi: 10.1007/s10111-007-0069-9]
49. Barendregt W, Bekker M, Bouwhuis D, Baauw E. Identifying usability and fun problems in a computer game during first use and after some practice. Int J Hum Comput Stud 2006 Sep;64(9):830-846. [doi: 10.1016/j.ijhcs.2006.03.004]
50. Sim G, Horton M, Danino N. Evaluating Game Preference using the Fun Toolkit across Cultures. In: Proceedings of the 26th BCS Conference on Human Computer Interaction (HCI). 2012 Presented at: The 26th BCS Conference on Human Computer Interaction (HCI); Sep 12-14, 2012; Birmingham, UK. [doi: 10.14236/ewic/HCI2012.56]
51. Horton M, Sim G, Zaman B, Slegers K. Evaluating long term user experience with children: comparing the MemoLine with interviews. In: Proceedings of the 18th ACM International Conference on Interaction Design and Children. 2019 Jun Presented at: IDC '19: Proceedings of the 18th ACM International Conference on Interaction Design and Children; June 12-15, 2019; Boise, USA p. 51-57. [doi: 10.1145/3311927.3323128]
52. Cano S, Collazos CA, Aristizábal LF, Gonzalez CS, Moreira F. Towards a methodology for user experience assessment of serious games with children with cochlear implants. Telematics Inform 2018 Jul;35(4):993-1004. [doi: 10.1016/j.tele.2017.09.011]
53. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
54. Lee S, Lee DK. What is the proper way to apply the multiple comparison test? Korean J Anesthesiol 2018 Oct;71(5):353-360 [FREE Full text] [doi: 10.4097/kja.d.18.00242] [Medline: 30157585]
55. de Amorim e Silva CJ, Mackenzie A, Hallowell LM, Stewart SE, Ditchfield MR. Practice MRI: reducing the need for sedation and general anaesthesia in children undergoing MRI. Australas Radiol 2006 Aug;50(4):319-323. [doi: 10.1111/j.1440-1673.2006.01590.x] [Medline: 16884416]
56. Barnea-Goraly N, Weinzimer SA, Ruedy KJ, Mauras N, Beck RW, Marzelli MJ, Diabetes Research in Children Network (DirecNet). High success rates of sedation-free brain MRI scanning in young children using simple subject preparation protocols with and without a commercial mock scanner--the Diabetes Research in Children Network (DirecNet) experience. Pediatr Radiol 2014 Feb;44(2):181-186 [FREE Full text] [doi: 10.1007/s00247-013-2798-7] [Medline: 24096802]
57. de Bie H, Boersma M, Wattjes MP, Adriaanse S, Vermeulen RJ, Oostrom KJ, et al. Preparing children with a mock scanner training protocol results in high quality structural and functional MRI scans. Eur J Pediatr 2010 Sep;169(9):1079-1085 [FREE Full text] [doi: 10.1007/s00431-010-1181-z] [Medline: 20225122]
58. Tugwell J, Goulden N, Mullins P. Alleviating anxiety in patients prior to MRI: a pilot single-centre single-blinded randomised controlled trial to compare video demonstration or telephone conversation with a radiographer versus routine intervention. Radiography (Lond) 2018 May;24(2):122-129. [doi: 10.1016/j.radi.2017.10.001] [Medline: 29605108]

59. Piira T, Sugiura T, Champion G, Donnelly N, Cole A. The role of parental presence in the context of children's medical procedures: a systematic review. Child Care Health Dev 2005 Mar;31(2):233-243 [FREE Full text] [doi: 10.1111/j.1365-2214.2004.00466.x] [Medline: 15715702]
60. Banerjee-Guénette P, Bigford S, Glegg SM. Facilitating the implementation of virtual reality-based therapies in pediatric rehabilitation. Phys Occup Ther Pediatr 2020;40(2):201-216. [doi: 10.1080/01942638.2019.1650867] [Medline: 31416381]

Abbreviations

ANOVA: analysis of variance
CCLS: certified child life specialist
CLP: Child Life Program
HMD: head-mounted display
MRI: magnetic resonance imaging
SPM: standard preparatory manual
STAI: State-Trait Anxiety Inventory
T1: timepoint 1
T2: timepoint 2
T3: timepoint 3
USE: Usefulness, Satisfaction, and Ease of Use
VPT: Venham picture test
VR: virtual reality

Edited by R Kukafka; submitted 20.08.20; peer-reviewed by J Berghmans, K McCarthy; comments to author 30.09.20; revised version received 23.11.20; accepted 19.07.21; published 22.09.21.

Please cite as:
Stunden C, Stratton K, Zakani S, Jacob J
Comparing a Virtual Reality–Based Simulation App (VR-MRI) With a Standard Preparatory Manual and Child Life Program for Improving Success and Reducing Anxiety During Pediatric Medical Imaging: Randomized Clinical Trial
J Med Internet Res 2021;23(9):e22942
URL: https://www.jmir.org/2021/9/e22942
doi:10.2196/22942
PMID:34550072

Original Paper

Barriers and Facilitators to the Implementation of Virtual Reality as a Pain Management Modality in Academic, Community, and Safety-Net Settings: Qualitative Analysis

Urmimala Sarkar[1,2], MD, MPH; Jane E Lee[1,2], MPH; Kim H Nguyen[1,2,3], MPH, ScD; Sarah Lisker[1,2], BA; Courtney R Lyles[1,2,3], PhD

[1]Department of Medicine, University of California San Francisco, San Francisco, CA, United States

[2]Center for Vulnerable Populations, University of California San Francisco, San Francisco, CA, United States

[3]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, United States

Corresponding Author:
Urmimala Sarkar, MD, MPH
Department of Medicine
University of California San Francisco
1001 Potrero Avenue
Box 1364
San Francisco, CA, 94143
United States
Phone: 1 6282064273
Email: urmimala.sarkar@ucsf.edu

Abstract

Background: Prior studies have shown that virtual reality (VR) is an efficacious treatment modality for opioid-sparing pain management. However, the majority of these studies were conducted among primarily White, relatively advantaged populations and in well-resourced settings.

Objective: We conducted a qualitative, theory-informed implementation science study to assess the readiness for VR in safety-net settings.

Methods: Using the theoretical lens of the Consolidated Framework for Implementation Research (CFIR) framework, we conducted semistructured interviews with current VR users and nonusers based in safety-net health systems (n=15). We investigated barriers and facilitators to a commercially available, previously validated VR technology platform AppliedVR (Los Angeles, CA, USA). We used deductive qualitative analysis using the overarching domains of the CFIR framework and performed open, inductive coding to identify specific themes within each domain.

Results: Interviewees deemed the VR intervention to be useful, scalable, and an appealing alternative to existing pain management approaches. Both users and nonusers identified a lack of reimbursement for VR as a significant challenge for adoption. Current users cited positive patient feedback, but safety-net stakeholders voiced concern that existing VR content may not be relevant or appealing to diverse patients. All respondents acknowledged the challenge of integrating and maintaining VR in current pain management workflows across a range of clinical settings, and this adoption challenge was particularly acute, given resource and staffing constraints in safety-net settings.

Conclusions: VR for pain management holds interest for frontline pain management clinicians and leadership in safety-net health settings but will require significant tailoring and adaption to address the needs of diverse populations. Integration into complex workflows for pain management is a significant barrier to adoption, and participants cited structural cost and reimbursement concerns as impediments to initial implementation and scaling of VR use.

(J Med Internet Res 2021;23(9):e26623) doi:10.2196/26623

KEYWORDS

virtual reality; medical informatics; information technology; implementation science; qualitative research

Introduction

Immersive virtual reality (VR) has emerged as an efficacious treatment modality for a wide range of medical and neurocognitive conditions including pain. VR is delivered via a headset that displays computer-generated and oftentimes interactive audio and visual content that can be designed to reduce pain through techniques such as distraction, relaxation, and mindfulness. Randomized controlled trials and effectiveness studies demonstrate that VR improves pain scores among inpatients and during medical procedures [1-4]. Additionally, studies indicate that patients can use VR to help manage chronic pain [5-8]. In a prior study among predominantly non-Hispanic White patients, most patients found VR use for pain to be a positive experience [9]. VR is generally considered safe, with the most common adverse effects being dizziness and nausea. Motion sickness and seizure disorders are the most prevalent of the few medical contraindications to VR use.

Clinicians have demonstrated interest in VR therapy as a safe and effective adjunctive or replacement to opioid agents to avoid adverse consequences of opioid use [10]. Opioid misuse represents a growing epidemic in the United States, affecting 10.3 million people, 48,000 of whom died in 2018. Even in acute-care settings, initiation of opioids can lead to long-term use, the consequences of which include dependence and hyperalgesia [11]. There is an urgent need for safer pain management for the millions of chronic pain sufferers worldwide. This represents a public health challenge as well, as racial and ethnic minorities, the elderly, and patients who speak English as a second language are more likely to experience suboptimal treatment for their chronic pain [12].

Despite studies proving the efficacy of and great promise for VR therapy as a nonpharmacologic approach to pain management, VR has not become part of routine chronic pain management [2,13,14]. As with other complex interventions, there is little evidence to guide implementation [15]. Although we know that VR has been effective in clinical trial settings, understanding why and how it has been effective is required to translate this modality to diverse practice settings [16]. Most VR studies have also been conducted in settings that serve ethnically homogenous, relatively advantaged populations with high health literacy and educational attainment.

There are specific considerations in adapting digital innovations to diverse populations and to safety-net health care settings that disproportionately care for them [17]. A systematic review of digital innovations in the safety net found that externally developed interventions face significant challenges, including acceptability to providers, staff, and patients; staffing needs; and implementation costs [18]. We sought to elucidate the implementation climate specifically for VR [19]. Therefore, we conducted a qualitative, theory-informed implementation science study to assess the readiness for VR in safety-net settings.

Methods

This study aims to determine barriers and facilitators to the implementation of VR as a treatment for pain in safety-net settings. We obtained institutional review board approval (#19-29025) for this project from the University of California, San Francisco (UCSF) Human Research Protection Program.

Conceptual Framework and Interview Guide

We developed an interview guide based on the Consolidated Framework for Implementation Research (CFIR) [20,21]. We chose this framework because it accounts for a wide range of contextual factors and is widely used to characterize complex interventions in health care settings [22-26]. The CFIR framework includes 5 overarching domains: (1) the characteristics of the intervention itself; (2) the inner setting, that is, the specific environment in which the intervention will take place; (3) the outer setting, or larger environment, encompassing the broader landscape in which the inner setting sits; (4) the individuals involved in intervention delivery and the target population; and (5) the process by which the intervention is introduced.

Data Collection

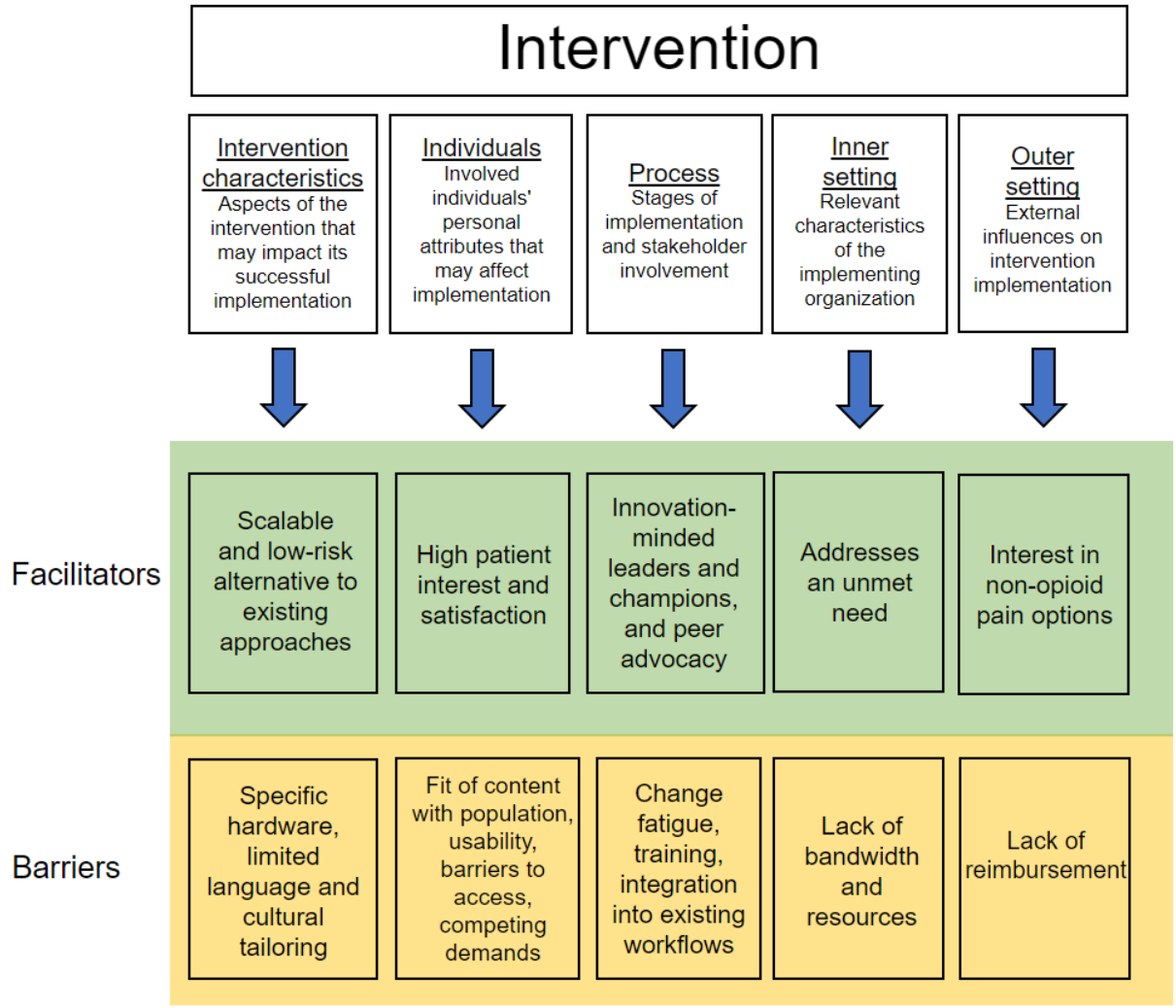
We used the snowball method of sampling to conduct 15 semistructured interviews (Multimedia Appendix 1) of 5 health care providers, leaders, and administrators who currently provide VR to patients (users) and 10 who do not use VR (nonusers). Most but not all nonusers practiced in safety-net settings, defined as public hospitals or federally qualified health centers serving the majority of patients on Medicaid or those who are uninsured [27]. We intentionally sampled nonusers as health care providers or staff without experience in using VR for clinical pain treatment, ensuring that every participant had expertise working with diverse patient populations in their practice (i.e., safety net or academic medical center sites serving Medicaid or racially/ethnically or linguistically diverse patients). We oversampled safety-net sites in particular because they are generally the last to be considered for implementation of innovation programs, and we sought to seek out their perspectives early to improve eventual broadscale implementation of this technology for diverse patient populations [28]. We conducted interviews from October 2019 to September 2020. Two members of the study team conducted the interviews via videoconferencing with screen sharing because of either logistical barriers or public health restrictions due to the COVID-19 pandemic for meeting in person. The team recorded the interviews with participant consent and transcribed the interviews without any identifying information. The participants received an incentive for their participation, with e-gift cards sent via email following interview completion.

Data Analysis

We used the qualitative software Dedoose (SocioCultural Research Consultants, Hermosa Beach, CA, USA) for analysis. Two authors, KN and JL, read all the transcripts in detail. For deductive coding, they assigned excerpts to 1 of the 5 overarching CFIR domains (intervention characteristics, individual characteristics, implementation process, inner setting, and outer setting) and met to discuss codes and resolve inconsistencies (Figure 1). They then independently analyzed 3 interview transcripts inductively to identify themes that arose from the interviews. They subsequently met to reach an

agreement on a comprehensive list of potential themes. One investigator (author JL) coded the remaining (12 total) transcripts under the supervision of KN for any relevant CFIR constructs and new inductive codes that emerged from participant responses. The entire study team met regularly to discuss discrepancies or resolve disputes regarding the coding. All authors iteratively revised and agreed upon the final list of themes and representative quotes. We reached thematic saturation after 11 interviews and elected to complete 4 additional interviews that had already been scheduled.

Figure 1. Overview of Consolidated Framework for Implementation Research (CFIR) domains and examples of barriers and facilitators.



Data Availability

The qualitative data generated and analyzed during the current study are not publicly available due to the risk of re-identification of participants, but codebooks are available upon reasonable request.

Results

We interviewed 15 participants from a variety of roles and organizations (Table 1).

We defined “user” as a health care provider or staff member interviewee with previous research experience implementing the AppliedVR platform for pain treatment at their delivery system; we defined “nonuser” as a health care provider or staff member interviewee without experience with AppliedVR or other VR platforms for clinical pain treatment.

Underlying these results was a strong sense of an unmet need for chronic pain management among users and nonusers. There was agreement among all interviewees that addressing pain management differently represents a high organizational/clinical priority. Building on that universally acknowledged need and prioritization, both users and nonusers of VR identified a range of factors related to the intervention itself and to the milieu in which it would be implemented that they expected would impact VR implementation. We described the major themes within each CFIR domain and provided exemplar quotes.

Table 1. Participant roles, user status, and organization characteristics.

Participant ID	Participant role	User/nonuser	Organization characteristics
1	Chief Medical Information Officer	User	Safety-net health system
2	Research physician	User	Academic medical center
3	Clinical research coordinator	User	Academic medical center
4a	Director for Innovation and Digital Health[a]	User	Academic medical center
4b	Program Manager for Innovation and Digital Health[a]	User	Academic medical center
5	Director of surgical subspecialty service	User	Nonprofit regional tertiary medical center
6	Primary care and addiction services physician	Nonuser	Safety-net health system
7	Associate Director of Research and Health Equity	Nonuser	Academic medical center
8	Chief Medical Officer	Nonuser	Safety-net health system
9	Chief Medical Informatics Officer	Nonuser	Safety-net health system
10	Primary care physician, pain clinic	Nonuser	Safety-net health system
11	Director of Telehealth	Nonuser	Safety-net health system
12	Psychologist, internal medicine	Nonuser	Academic medical center
13	Registered nurse, palliative care	Nonuser	Safety-net health system
14	Associate Chief Medical Informatics Officer	Nonuser	Safety-net health system
15	Psychologist, internal medicine	Nonuser	Safety-net health system

[a]These participants were interviewed simultaneously.

Characteristics of the Intervention

The VR intervention itself has attributes that both facilitate and inhibit implementation (Multimedia Appendix 2, Table S1). A key factor for the intervention is its relative advantage over existing approaches for pain. VR users and nonusers both mentioned the favorable safety of using VR compared to pain medications. Additionally, VR users also mentioned the option for patients to use the technology on their own on an ongoing basis after initial training from staff, as opposed to ongoing, clinician-delivered pain interventions, such as cognitive behavioral therapy. Similarly, nonusers appreciated the potential for scaling pain management activities through VR-enabled self-management approaches. Nonusers of VR commented on the face validity of VR as a chronic pain treatment. In terms of the intervention content, both users and nonusers noted that the platform is available only in English and therefore not usable in limited English–proficiency populations. The difference between users and nonusers was in how they prioritized this limitation. Interviewees based in the safety net described non-English language as a prerequisite, whereas it was considered a limitation of varying importance among those who were already using VR. The participants discussed the ability to culturally tailor VR scenarios in opposing ways. VR nonusers expressed some skepticism about the content's appropriateness for culturally diverse populations and for those with significant trauma histories. VR users noted that the currently available content does not always resonate across cultures, particularly for minoritized populations. However, they did consider the ability of VR to offer many culturally tailored content offerings to be a key attribute of the approach. Both users and nonusers acknowledged the specific contraindication of motion sickness and the overall low risk of the modality.

Individual Characteristics

For this study, we considered the CFIR domain of "individual characteristics" to refer to target patients for VR as a pain intervention (Multimedia Appendix 2, Table S2). A prevalent theme was comfort with technology, including challenges for older adults in navigating technology and digital literacy more broadly. Nonusers, in particular, raised concerns about patient mistrust of new or experimental treatments. Despite this, several interviewees emphasized that they would expect a significant proportion of low-income and diverse patients cared for in safety-net settings to willingly engage with VR. Participants who used VR expressed near-universal high levels of patient satisfaction with VR, and nonusers expected that patients with chronic pain would appreciate having additional treatment options. Nonusers did cite physical opioid dependence and a fear of opioid de-prescribing as expected barriers from patients, and users and nonusers alike referred to patients' more pressing social and health needs as significant barriers to VR implementation for pain. Nonusers specifically mentioned trade-offs for health systems in deciding whether to invest in innovations such as VR versus approaches to address patient needs such as housing and food insecurity.

Implementation Process

Both VR users and nonusers considered VR implementation to be a complex process (Multimedia Appendix 2, Table S3). There was consensus that patients would require specific orientation from staff in order to initiate VR use for pain and that staff support would be required for coaching and troubleshooting on

an ongoing basis. Both users and nonusers pointed to the requirement that frontline staff interact with VR and have a personal buy-in to facilitate successful implementation. Similarly, all participants highlighted the need for champions among clinicians who can share both evidence for VR as a pain treatment and successful treatment experiences. The availability of staff time was a dominant concern, with VR users emphasizing the need to plan for adequate staff effort to implement VR and nonusers expressing doubt about staff availability. Integration of VR into existing pain management workflows, a key aspect of the implementation process, highlighted barriers and facilitators. It emerged that implementation success would depend on the extent to which VR could be integrated into specific clinical settings. Some, but not all, nonusers focused on integration into the electronic health record. The participants cited some especially challenging settings, such as primary care, and some more feasible settings, such as the outpatient pain clinic. More specifically, the intervention requires a multistep process with each patient, which both users and nonusers believe complicates more widespread implementation.

Inner Setting

The inner setting for VR implementation is the specific health system and, within each health system, the specific clinical settings that currently treat pain (Multimedia Appendix 2, Table S4). One overarching inner setting barrier, seen across all participants, is that within health systems, pain management is a cross-cutting issue and therefore requires collaboration among departments and stakeholders, in contrast to other clinical conditions for which one clinical unit is entirely responsible. Both users and nonusers cited pressure to prescribe opioids and the providers' lack of familiarity with VR as barriers to use. Several participants mentioned the default habit and culture of addressing pain with medication. Among the nonusers, participants noted that their peers hold skepticism about corporate approaches to treatment. Both users and nonusers emphasized the importance of leadership attitudes. They described interest in VR as part of a larger mindset around openness to innovation. All participants emphasized the need for dedicated staff time and training and space for VR treatment, which require a leadership commitment. Resource limitations in safety-net settings were seen to preclude these needed actions for implementation. Most participants suggested that an innovative culture is a prerequisite for VR implementation, but safety-net participants described this as pervasive: "We seem to be late adopters." In contrast, VR users described few individuals as having an anti-innovation attitude. Some nonusers described local leaders as not being innovation minded but instead being focused on delivery of long-standing services. Similarly, nonusers described the wide array of stakeholders (eg, frontline staff, information technology [IT], and clinical leaders) within the local environment whose engagement was needed for success. In terms of concrete needs, both users and nonusers cited the need for adequate physical space for VR treatment and up-front investment in the VR hardware as inner-setting requirements. However, VR users experienced these needs as surmountable challenges, while VR nonusers highlighted space and initial costs as formidable obstacles.

Outer Setting

Challenges related to the outer setting of the intervention, such as external policies and incentives, arose as well (Multimedia Appendix 2, Table S5). In the broader health care landscape, use of VR is not a billable service. The participants viewed this lack of insurance reimbursement as a critical barrier to implementation at scale. Both VR users and nonusers mentioned philanthropic and private funding as a current or possible approach for the short term. There was consensus that expecting patients to bear the cost of pain treatment is a significant implementation barrier. Beyond cost, concerns around data privacy and security in digital health surfaced among nonusers. Nonusers also pinpointed the digital health industry's focus on wellness over developing treatments for specific conditions as a barrier to widespread use of digital therapies more broadly.

Discussion

Principal Results

Our results suggest that many of the factors affecting VR implementation, or, indeed, implementation of novel, technology-enabled therapeutic approaches, exist in both safety-net settings and more advantaged, resource-rich health care systems. We did not identify a completely different set of issues relating to VR in the safety net; rather, similar challenges seem to be magnified in safety-net settings. For example, among safety-net stakeholders who do not currently use VR, an English-only intervention would be impossible to implement for their linguistically diverse patient populations, whereas non-safety-net participants saw this limitation as less significant in impacting adoption. Similarly, the cultural relevance and acceptability of the content was a source of concern for safety-net interview participants but appeared to be surmountable to non-safety-net participants who were already testing VR at their sites. In other words, current VR users saw further cultural tailoring as an area for future improvement, whereas our safety-net participants saw it as a prerequisite for use because of the diverse populations served in safety-net settings. Both VR-use and VR-nonuse sites acknowledged that competing demands in patients' lives, relating to housing status, income instability, and other social challenges, would interfere with using a novel therapeutic approach like VR. Again, these challenges seemed more prevalent in the safety net.

Our findings in the CFIR domain of "inner setting" suggest that the process of implementing a complex intervention like VR requires a formal change management strategy encompassing leadership support, frontline champions, and a campaign for staff buy-in. Safety-net respondents reported a lack of capacity for implementing new approaches, particularly with regard to leadership commitment to innovation and resources to conduct change management activities. The outer setting or larger context for VR implementation has facilitators, such as the desire to appear innovative, while reducing the harm from opioid use. However, the lack of insurance reimbursement and available funding for VR interventions is a significant barrier. Competing priorities for leadership can also stymie innovations such as VR implementation, especially in safety-net health care settings

where resource limitations and "keeping the doors open" are perennially in question.

Our findings echo prior work about innovation in safety-net settings. Dating back to the early days of health IT, safety-net settings have lagged in implementation [29]. Prior studies suggest that frontline patient engagement and trust, appropriateness of innovative tools for diverse end users, and dedicated resources for innovation are particularly salient for the implementation of new health IT approaches in the safety net [18,28,30,31]. Implementation science studies in safety-net settings should specifically emphasize these concepts in applying, and, where necessary, adapting, conceptual frameworks such as the CFIR.

Our study lends weight to a previously developed evaluation framework for VR [19]. The VR1 study design phase emphasizes the importance of working directly with end users, and we found that implementation in safety-net health settings will require significant tailoring and adaptation of the VR intervention to address the needs of diverse populations (eg, racial and ethnic minorities, language barriers, and social complexity). Several actions can address the current barriers to VR use for pain management in safety-net health care settings. For example, our results underscore the need for cultural tailoring and translation (eg, VR content that features voices and images of racially and ethnically diverse individuals and is provided in languages other than English). Furthermore, more specific adaptations or considerations addressing structural issues disproportionately facing lower-income and racial/ethnic minorities in the United States (eg, unjust incarceration practices and historical trauma) might require deeper qualitative exploration with patients about their preferences or comfort in using VR headsets. A prior study of predominantly non-Hispanic White hospitalized patients demonstrated that many refuse to try VR [9]; the acceptability of VR to patients should be tested in more diverse patient groups. Following this, usability testing with diverse populations could evaluate concerns about the acceptability and usability of technology, disinfection, and adverse effects such as motion sickness. Any deployment strategy for VR should take into account the resource and workforce constraints of the safety-net environment. Developers of VR for chronic pain management should address implementation considerations in parallel with design and evaluation in order to foster higher uptake of VR and digital tools overall.

Strengths and Limitations

Our study has strengths, such as inclusion of multiple health systems and geographic areas. In addition, we compared and contrasted the experiences of health care system participants who are currently overseeing the use of a single VR platform with those of safety-net and academic providers across multiple sites who are not currently using VR. Furthermore, we used a well-known conceptual framework and reached thematic saturation in content analysis. Despite these strengths, the study does have limitations. The sample size was circumscribed, and we do not have information about the extent of VR use among the participants who are currently using the tool in their health systems.

Although our discussion focused on one technology approach, VR, for a single clinical problem (pain management), many of the themes that emerged are relevant across multiple innovative approaches to clinical care. The spread of innovations to safety-net health settings requires attention to patients' language and culture. Efforts to address resource constraints, such as providing staff training and detailed protocols encompassing workflow integration, can lower barriers to innovation. Inclusion of diverse populations and the health care settings where they disproportionately receive care will address issues that are relevant across other health systems as well.

Conclusion

This study contributes to the literature by revealing barriers and facilitators to implementing VR for pain management in safety-net settings. We demonstrate that frontline pain management clinicians and leaderships are interested in VR, but it will require significant tailoring and adaptation to address the specific needs of the diverse populations they serve. The participants cited integration into complex workflows, structural costs, and reimbursement concerns as major barriers to implementing and scaling VR use. Future studies should augment this approach with direct observation, quantitative measures of VR use, and direct usability assessment of the AppliedVR platform with diverse patient populations.

Acknowledgments

We would like to thank all interviewees for participating.

This publication was supported by the Food and Drug Administration (FDA) of the US Department of Health and Human Services (HHS) as part of a financial assistance award (Center of Excellence in Regulatory Science and Innovation Grant to University of California, San Francisco (UCSF) and Stanford University, U01FD005978) totaling $50,000, with 50% funded by the FDA/HHS, and $50,000, with 50% funded by a nongovernment source (AppliedVR). The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement by, the FDA/HHS or the US government.

Conflicts of Interest

The study team received funding from AppliedVR, but AppliedVR was not involved in data collection, interpretation, or dissemination of findings.

Multimedia Appendix 1

Consolidated Framework for Implementation Research (CFIR) structured interview guides for users and nonusers.
[DOCX File , 37 KB - jmir_v23i9e26623_app1.docx]

Multimedia Appendix 2
Participant quotes.
[DOCX File , 56 KB - jmir_v23i9e26623_app2.docx]

References

1. Carrougher GJ, Hoffman HG, Nakamura D, Lezotte D, Soltani M, Leahy L, et al. The effect of virtual reality on pain and range of motion in adults with burn injuries. J Burn Care Res 2009;30(5):785-791 [FREE Full text] [Medline: 19692911]
2. Dascal J, Reid M, IsHak WW, Spiegel B, Recacho J, Rosen B, et al. Virtual reality and medical inpatients: a systematic review of randomized, controlled trials. Innov Clin Neurosci 2017;14(1-2):14-21 [FREE Full text] [Medline: 28386517]
3. Spiegel B, Fuller G, Lopez M, Dupuy T, Noah B, Howard A, et al. Virtual reality for management of pain in hospitalized patients: a randomized comparative effectiveness trial. PLoS One 2019;14(8):e0219115 [FREE Full text] [doi: 10.1371/journal.pone.0219115] [Medline: 31412029]
4. Wismeijer AAJ, Vingerhoets AJJM. The use of virtual reality and audiovisual eyeglass systems as adjunct analgesic techniques: a review of the literature. Ann Behav Med 2005 Dec;30(3):268-278. [doi: 10.1207/s15324796abm3003_11] [Medline: 16336078]
5. Bani Mohammad E, Ahmad M. Virtual reality as a distraction technique for pain and anxiety among patients with breast cancer: a randomized control trial. Palliat Support Care 2019 Feb;17(1):29-34. [doi: 10.1017/S1478951518000639] [Medline: 30198451]
6. Matheve T, Bogaerts K, Timmermans A. Virtual reality distraction induces hypoalgesia in patients with chronic low back pain: a randomized controlled trial. J Neuroeng Rehabil 2020 Apr 22;17(1):55 [FREE Full text] [doi: 10.1186/s12984-020-00688-0] [Medline: 32321516]
7. Sarig BH, Takasaki H, Chen X, Bet-Or Y, Treleaven J. Cervical kinematic training with and without interactive VR training for chronic neck pain: a randomized clinical trial. Man Ther 2015 Feb;20(1):68-78. [doi: 10.1016/j.math.2014.06.008] [Medline: 25066503]
8. Sarig Bahat H, Croft K, Carter C, Hoddinott A, Sprecher E, Treleaven J. Remote kinematic training for patients with chronic neck pain: a randomised controlled trial. Eur Spine J 2018 Jun;27(6):1309-1323. [doi: 10.1007/s00586-017-5323-0] [Medline: 29018956]
9. Mosadeghi S, Reid MW, Martinez B, Rosen BT, Spiegel BMR. Feasibility of an Immersive Virtual Reality Intervention for Hospitalized Patients: An Observational Cohort Study. JMIR Ment Health 2016 Jun 27;3(2):e28 [FREE Full text] [doi: 10.2196/mental.5801] [Medline: 27349654]
10. FDA identifies harm reported from sudden discontinuation of opioid pain medicines and requires label changes to guide prescribers on gradual, individualized tapering. U.S. Food and Drug Administration. 2019 Apr 09. URL: https://www.fda.gov/drugs/drug-safety-and-availability/fda-identifies-harm-reported-sudden-discontinuation-opioid-pain-medicines-and-requires-label-changes [accessed 2021-08-30]
11. Larochelle MR, Liebschutz JM, Zhang F, Ross-Degnan D, Wharam JF. Opioid prescribing after nonfatal overdose and association with repeated overdose. Ann Intern Med 2016 Sep 06;165(5):376. [doi: 10.7326/l16-0168] [Medline: 27595218]
12. Institute of Medicine (US) Committee on Advancing Pain Research, Care, and Education. Pain as a Public Health Challenge. Washington DC: National Academies Press; 2011.
13. Morris LD, Louw QA, Grimmer-Somers K. The effectiveness of virtual reality on reducing pain and anxiety in burn injury patients. Clin J Pain 2009;25(9):815-826. [doi: 10.1097/ajp.0b013e3181aaa909] [Medline: 19851164]
14. Shahrbanian S, Ma X, Aghaei N, Korner-Bitensky N, Simmonds M. Use of virtual reality (immersive vs. non immersive) for pain management in children and adults: a systematic review of evidence from randomized controlled trials. Eur J Exp Biol 2012 Jan 10;2:1408-1422.
15. Greenhalgh T, Robert G, MacFarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004 Dec;82(4):581-629. [doi: 10.1111/j.0887-378x.2004.00325.x]
16. Dixon-Woods M, Bosk CL, Aveling EL, Goeschel CA, Pronovost PJ. Explaining Michigan: developing an ex post theory of a quality improvement program. Milbank Q 2011 Jun;89(2):167-205. [doi: 10.1111/j.1468-0009.2011.00625.x]
17. Lyles CR, Handley MA, Ackerman SL, Schillinger D, Williams P, Westbrook M, et al. Innovative implementation studies conducted in US safety net health care settings: a systematic review. Am J Med Qual 2019 Jun;34(3):293-306 [FREE Full text] [doi: 10.1177/1062860618798469] [Medline: 30198304]
18. Nouri SS, Adler-Milstein J, Thao C, Acharya P, Barr-Walker J, Sarkar U, et al. Patient characteristics associated with objective measures of digital health tool use in the United States: a literature review. J Am Med Inform Assoc 2020 May 01;27(5):834-841 [FREE Full text] [doi: 10.1093/jamia/ocaa024] [Medline: 32364238]

19. Birckhead B, Khalil C, Liu X, Conovitz S, Rizzo A, Danovitch I, et al. Recommendations for methodology of virtual reality clinical trials in health care by an international working group: iterative study. JMIR Ment Health 2019 Jan 31;6(1):e11973 [FREE Full text] [doi: 10.2196/11973] [Medline: 30702436]
20. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci [Internet] 2009;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
21. Damschroder LJ, Lowery JC. Evaluation of a large-scale weight management program using the consolidated framework for implementation research (CFIR). Implement Sci 2013 May 10;8:51 [FREE Full text] [doi: 10.1186/1748-5908-8-51] [Medline: 23663819]
22. Cilenti D, Brownson RC, Umble K, Erwin PC, Summers R. Information-seeking behaviors and other factors contributing to successful implementation of evidence-based practices in local health departments. J Public Health Manag Pract 2012 Nov;18(6):571-576 [FREE Full text] [doi: 10.1097/PHH.0b013e31825ce8e2] [Medline: 23023282]
23. Garg SK, Lyles CR, Ackerman S, Handley MA, Schillinger D, Gourley G, et al. Qualitative analysis of programmatic initiatives to text patients with mobile devices in resource-limited health systems. BMC Med Inform Decis Mak 2016 Feb;16:16 [FREE Full text] [doi: 10.1186/s12911-016-0258-7] [Medline: 26851941]
24. Khoong EC, Cherian R, Smith DE, Schillinger D, Wolf MS, Sarkar U. Implementation of patient-centered prescription labeling in a safety-net ambulatory care network. Am J Health Syst Pharm 2018 Aug 15;75(16):1227-1238. [doi: 10.2146/ajhp170821] [Medline: 29950392]
25. Kirk MA, Kelley C, Yankey N, Birken SA, Abadie B, Damschroder L. A systematic review of the use of the Consolidated Framework for Implementation Research. Implement Sci 2016 May 17;11:72 [FREE Full text] [doi: 10.1186/s13012-016-0437-z] [Medline: 27189233]
26. Zulman DM, Damschroder LJ, Smith RG, Resnick PJ, Sen A, Krupka EL, et al. Implementation and evaluation of an incentivized Internet-mediated walking program for obese adults. Transl Behav Med 2013 Dec;3(4):357-369. [doi: 10.1007/s13142-013-0211-6] [Medline: 24294324]
27. Lewin ME, Altman S, editors. America's Health Care Safety Net: Intact but Endangered. Washington, D.C: National Academies Press; 2000.
28. Lyles CR, Aulakh V, Jameson W, Schillinger D, Yee H, Sarkar U. Innovation and transformation in California's safety net health care settings: an inside perspective. Am J Med Qual 2014;29(6):538-545 [FREE Full text] [doi: 10.1177/1062860613507474] [Medline: 24170938]
29. Jha AK, DesRoches CM, Shields AE, Miralles PD, Zheng J, Rosenbaum S, et al. Evidence of an emerging digital divide among hospitals that care for the poor. Health Aff (Millwood) 2009 Nov;28(6):w1160-w1170. [doi: 10.1377/hlthaff.28.6.w1160] [Medline: 19858142]
30. Moore SL, Fischer I, Havranek EP. Translating health services research into practice in the safety net. Health Serv Res 2016 Feb;51(1):16-31 [FREE Full text] [doi: 10.1111/1475-6773.12341] [Medline: 26646189]
31. Lyles CR, Nelson EC, Frampton S, Dykes PC, Cemballi AG, Sarkar U. Using electronic health record portals to improve patient engagement: research priorities and best practices. Ann Intern Med 2020 Jun 02;172(11 Suppl):S123-S129 [FREE Full text] [doi: 10.7326/M19-0876] [Medline: 32479176]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
IT: information technology
UCSF: University of California, San Francisco
VR: virtual reality

Edited by R Kukafka, G Eysenbach; submitted 18.12.20; peer-reviewed by B Spiegel, A Tapuria; comments to author 18.01.21; revised version received 26.01.21; accepted 25.05.21; published 22.09.21.

Please cite as:
Sarkar U, Lee JE, Nguyen KH, Lisker S, Lyles CR
Barriers and Facilitators to the Implementation of Virtual Reality as a Pain Management Modality in Academic, Community, and Safety-Net Settings: Qualitative Analysis
J Med Internet Res 2021;23(9):e26623
URL: https://www.jmir.org/2021/9/e26623
doi:10.2196/26623
PMID:34550074

Original Paper

Readiness and Acceptance of eHealth Services for Diabetes Care in the General Population: Cross-sectional Study

PV AshaRani[1], PhD; Lau Jue Hua[1], BSc; Kumarasan Roystonn[1], MSc; Fiona Devi Siva Kumar[1], BSc; Wang Peizhi[1], MPH; Soo Ying Jie[1], BSc; Saleha Shafie[1], BSocSci; Sherilyn Chang[1], BSocSci; Anitha Jeyagurunathan[1], MPhil; Chua Boon Yiang[1], MSc; Edimansyah Abdin[1], PhD; Janhavi Ajit Vaingankar[1], MSc; Chee Fang Sum[2], MBBS; Eng Sing Lee[3], PhD; Siow Ann Chong[1], MBBS, MMed; Mythily Subramaniam[1,4], PhD

[1]Research Division, Institute of Mental Health, Singapore, Singapore

[2]Admiralty Medical Centre, Khoo Teck Puat Hospital, Singapore, Singapore

[3]Clinical Research Unit, National Healthcare Group Polyclinics, Singapore, Singapore

[4]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

Corresponding Author:
PV AshaRani, PhD
Research Division
Institute of Mental Health
10 Buangkok View
Singapore, 539747
Singapore
Phone: 65 63892961
Email: asharani_pezhummoottil_vasudevan_n@imh.com.sg

Abstract

Background: Diabetes management is a growing health care challenge worldwide. eHealth can revolutionize diabetes care, the success of which depends on end user acceptance.

Objective: This study aims to understand the readiness and acceptance of eHealth services for diabetes care among the general population, perceived advantages and disadvantages of eHealth, and factors associated with eHealth readiness and acceptance in a multiethnic Asian country.

Methods: In this cross-sectional epidemiological study, participants (N=2895) were selected through disproportionate stratified random sampling from a population registry. Citizens or permanent residents of Singapore aged >18 years were recruited. The data were captured through computer-assisted personal interviews. An eHealth questionnaire was administered in one of four local languages (English, Chinese, Malay, or Tamil), as preferred by the participant. Bivariate chi-square analyses were performed to compare the sociodemographic characteristics and perception of advantages and disadvantages of eHealth services between the diabetes and nondiabetes groups. Multivariable logistic regression models were used to determine factors associated with eHealth readiness and acceptance. All analyses were weighted using survey weights to account for the complex survey design.

Results: The sample comprised participants with (n=436) and without (n=2459) diabetes. eHealth readiness was low, with 47.3% of the overall sample and 75.7% of the diabetes group endorsing that they were not ready for eHealth (P<.001). The most acceptable eHealth service overall was *booking appointments* (67.4%). There was a significantly higher preference in the diabetes group for face-to-face sessions for *consultation with the clinician* (nondiabetes: 83.5% vs diabetes: 92.6%; P<.001), *receiving prescriptions* (61.9% vs 79.3%; P<.001), *referrals to other doctors* (51.4% vs 72.2%; P<.001), and *receiving health information* (34% vs 63.4%; P<.001). The majority of both groups felt that eHealth requires users to be computer literate (90.5% vs 94.3%), does not build clinician-patient rapport compared with face-to-face sessions (77.5% vs 81%), and might not be credible (56.8% vs 64.2%; P=.03). Age (≥35 years), ethnicity (Indian), and lower education status had lower odds of eHealth readiness. Age (≥35 years), ethnicity (Indian), lower education status (primary school), BMI (being underweight), and marital status (being single) were associated with a lower likelihood of eHealth acceptance. Among only those with diabetes, a longer duration of diabetes (4-18 years), higher education (degree or above), and younger age (23-49 years) were associated with eHealth readiness, whereas younger age and income (SGD 2000-3999 [US $1481-$2961]) were associated with acceptance.

Conclusions: Overall, an unfavorable attitude toward eHealth was observed, with a significantly higher number of participants with diabetes reporting their unwillingness to use these services for their diabetes care. Sociodemographic factors associated with

acceptance and readiness identified a group of people who were unlikely to accept the technology and thus need to be targeted for eHealth literacy programs to avoid health care disparity.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2020-037125

(J Med Internet Res 2021;23(9):e26881) doi:10.2196/26881

KEYWORDS

eHealth; diabetes; general population; acceptance; readiness

Introduction

Background

Diabetes has significant social and economic consequences globally [1]. The prevalence of diabetes increased from 422 million in 2016 [1] to 463 million in 2019 [2], despite innovative disease prevention strategies [1]. This number is estimated to increase by 51% (to 700 million) by 2045 [2], which will lead to higher consumption of health care services [3]. Health care systems are not equipped to manage such a surge in the number of cases [4] and are expected to face additional challenges in the coming years, as a shortage of trained staff is expected because of falling birth rates [5]. The geographical isolation of people living in remote areas is another challenge for health care delivery [6]. This situation calls for a change in the way patients are managed without compromising the quality of health care services.

eHealth is defined by the World Health Organization as a "cost-effective and secure use of information technology in health-related fields: health care services, health surveillance, health literature, and health education, knowledge and research" [7]. Patients with diabetes require regular monitoring of their blood glucose levels, diet, physical activities, and medications to achieve glycemic control [8,9], which will reduce their risk of mortality and cardiovascular complications [10]. eHealth offers diverse treatments and management options in a cost-effective manner to facilitate these requirements [11] and thus allows both clinicians and patients to manage the disease efficiently in a collaborative manner. A meta-analytic review of 18 randomized controlled trials showed a significant reduction in glycated hemoglobin and better glycemic control in participants using eHealth compared with those monitored through routine methods [12]. Pereira et al [13] reviewed the literature on the delivery of health education through eHealth and concluded that the technology delivered better outcomes in terms of efficient management of diabetes. eHealth platforms can also be used for promoting a healthy lifestyle in patients with diabetes by facilitating exercise lessons through video games or other virtual environments [14,15]. Apart from the remote management of diseases, eHealth also offers services such as electronic prescriptions, referrals, teleconsulting, and health education. Thus, eHealth empowers patients to manage their disease and make decisions on their health care remotely, without the need to travel to hospitals.

An individual's decision to use or accept a technology is based on their perception of usefulness and ease of use, which shapes their attitude toward eHealth and intention to use the technology (Technology Acceptance Model). A positive attitude is an antecedent to intention or readiness to use the service. Thus, by modifying the perception of usefulness and ease of use, acceptance can be improved [16,17]. Hossain et al [17] reported that a positive attitude toward eHealth, perceived effectiveness, and access to cell phones were associated with eHealth acceptance in developing countries. The authors also showed that a positive attitude toward technology increased the odds of acceptance by 4.5-fold and reported that sociodemographic factors such as age, gender, and education have a significant influence on this choice. A study of patients with chronic respiratory disease reported higher acceptance of eHealth services for booking appointments, accessing laboratory results, educational purposes, and receiving e-prescriptions, than acceptance of eHealth for treatment-related services such as contact with the health care team and referrals to other clinicians [18]. The likelihood of acceptance was dependent on the duration of the disease, age, and education. A subsequent large-scale study from the author further confirmed that education and perceived usefulness determined eHealth acceptance [19]. Thus, the role of sociodemographic determinants, perception of usefulness, and ease of use in eHealth acceptance is irrefutable.

Although significant efforts have been made to study the efficacy of eHealth programs in cohort studies, there is a dearth of literature on the demographics and attitude-related disparity between members of the general population who have or do not have diabetes. In addition, many of the research studies gathered data from people who had already used specific services to assess the extent of user experience. The expectations and attitudes of those who had never used the technology were not addressed. These perspectives are imperative in the coming era, where the growing demand for health care services will prompt organizations to leverage eHealth for cost-effectiveness and efficient delivery of care. To date, no studies have been conducted at the population level to understand the readiness, acceptability, and attitudes of the general public toward eHealth services for diabetes care. eHealth was one of the most sought-after technologies in the COVID-19 pandemic phase, where many countries tried to deliver diabetes care through eHealth [20]. However, the patients were unwilling to accept the platform even in countries that had an integrated eHealth framework. The most accepted eHealth service is receiving prescriptions for insulin and other medications [20]. Thus, these data clearly show that understanding the readiness and acceptance of eHealth is critical in the future to manage health care services during pandemics.

Objectives

The objective of this study is to understand the (1) readiness and acceptance of eHealth services among the general

population (with and without diabetes) at a national level, (2) perceived advantages and disadvantages of eHealth services, and (3) sociodemographic and clinical factors associated with eHealth readiness and acceptance. This will help organizations to consider various factors that stem from the patient's needs or concerns when implementing eHealth services to avoid technology failure.

Methods

Participants

The detailed methodology of this study has been reported previously [21]. This study was part of a nationwide cross-sectional survey that was intended to understand the knowledge, attitude, and practices of the general public toward diabetes. Participants from the general public were randomly selected from a population registry database of Singapore comprising permanent residents and citizens through disproportionate random sampling (ethnicity and age). Surveys were conducted face-to-face by trained interviewers.

Eligibility

Participants who were aged ≥18 years; who were citizens or permanent residents of Singapore; and who could understand English, Chinese, Malay, or Tamil were recruited. Those who were nonresidents or noncitizens, incapable of doing an interview, living outside the country, or institutionalized during the entire survey period were excluded from the study.

Sample Size

Sample size calculation has been reported previously [21]. Briefly, the power calculations were based on the prevalence rates of knowledge in the general public and included calculations for binary proportions to determine overall sample sizes as well as those for subgroups to produce a margin of error of ≤0.05. A statistical power of 0.8 was targeted with the type 1 error limited to α of .05. The sample size was adjusted to accommodate deviations owing to random sampling. The study used 16 strata: four for ethnicity (Chinese, Malay, Indian, and others) and four for age groups (18-34 years, 35-49 years, 50-64 years, and ≥65 years). The sample was drawn in six replicates (total 5698) released at different intervals starting in February 2019. A target sample of 3000 was estimated to be sufficient to understand the knowledge, attitude, and practices of the general public. The study closed recruitment in September 2020 with a final response rate (total completed interviews / [total number of sample – eligible cases]) of 66.2%. The eligibility rate was 76.8%. The reasons for ineligibility included death, institutionalization (eg, incarceration), residing outside the country during the survey period, uncontactable, severe physical or mental condition that interfered with participation, and language barriers.

Procedure

An invitation letter was sent to all participants 1 to 2 weeks before the intended home visit by a trained interviewer. The trained interviewer approached the households and captured the responses via computer-assisted personal interviews in a language preferred by the participant (English, Chinese, Malay, or Tamil). Regular quality checks were conducted on the data collected [21]. Written consent was obtained from all the participants before the survey. All study procedures were conducted in accordance with the ethical guidelines (Domain Specific Review Board reference 2018/00430). The study was suspended during the *circuit breaker* (heightened safe distancing measures that were implemented to prevent the spread of the virus. This included closure of schools, workplaces, other venues, and avoidance of interaction with those who do not live together) period in response to the pandemic (March 2020 to July 2020) and restarted in August 2020 with interviewers and respondents adhering to safe distancing measures and mask policy.

Questionnaire

Sociodemographic Questionnaire

The information collected included age (as of last birthday), gender, ethnicity, employment status, educational qualification, average monthly personal income (including all allowances over the past 12 months), average household income, height, weight, BMI, and marital status.

Diabetes Questionnaire

The participants were asked two questions to ascertain their diabetes status: (1) "Have you ever been told by a doctor that you have Diabetes?" The response options were "yes," "no," or "I don't know." Those who answered "yes" were further probed with question (2) "What type of diabetes do you have?" The response options were "type 1 diabetes," "type 2 diabetes," "gestational diabetes," "others," and "I don't know." This analysis of the diabetes group included those who endorsed type 1 or type 2 diabetes.

eHealth Questionnaire

The eHealth questionnaire was administered after introducing the concept as "eHealth services refers to health care services (eg, Health Hub) delivered through internet which includes programmes such as online appointment booking, online prescription, online consultation with nurse/doctor/therapist, internet-based support programmes, online referrals, etc."

Readiness

The readiness of the participants toward eHealth was assessed using a single statement, "I am not ready for eHealth." Participants were asked to indicate the extent to which they agreed or disagreed with this statement. The responses were captured on a 5-point scale from "strongly agree" to "strongly disagree" with an option for "neutral" in the middle. For the regression analysis, the responses were grouped as "not ready" (ie, strongly agree or agree) versus "ready" (ie, neutral, strongly disagree, or disagree).

Acceptability

Acceptability of eHealth was assessed using two items: (1) "If you seek treatment for diabetes would you seek treatment delivered via the internet?" The responses were captured as five different options ranging from "definitely would" and "possibly would" to "definitely wouldn't," "definitely wouldn't," and "not sure." For the logistic regression analysis,

"definitely/possibly would" was combined into one category, "definitely/possibly wouldn't" was subsumed into another group, and those who indicated "not sure" (n=95) were removed from the analyses. (2) The acceptance of specific health services was assessed using a stem question, which asked, "Overall, which type of service would you prefer to use if you experience diabetes-related health problems?" The services listed included consultation, receiving prescriptions, booking appointments, receiving health information, and referrals to other clinicians. The participants were asked to indicate their preferred methods to each listed service with one of the three response options "face-to-face," "eHealth," or "both eHealth and face-to-face."

Perceived Advantages and Disadvantages of eHealth

This construct was captured using a series of statements with response options ranging from "strongly agree" to "strongly disagree" with an option for "neutral" in the middle (5-point scale). The items included "internet-based treatments" (1) are more convenient (do not have to travel to the clinic and can access anywhere), (2) save time, (3) are cost-saving, (4) ensure privacy and anonymity (personal information is kept confidential), (5) save from embarrassment related to face-to-face consultation, (6) might not be helpful for my health condition, (7) do not build clinician-patient rapport of face-to-face session, (8) require users to be computer literate, and (9) might not be credible. For bivariate chi-square analyses, the responses were grouped into three groups: strongly agree or agree, neutral, and strongly disagree or disagree.

Chronic Conditions Checklist

The chronic conditions checklist [22] captured 18 different chronic conditions (physical illnesses such as asthma, diabetes, stroke or major paralysis, high blood pressure, hyperlipidemia, arthritis or rheumatism, cancer, neurological conditions such as epilepsy or convulsions, Parkinson disease, congestive heart failure, heart diseases, back problems including disk or spine, stomach ulcer, chronic inflamed bowel, enteritis or colitis, thyroid diseases, kidney failure, migraine headaches, and chronic lung diseases [chronic bronchitis and emphysema]). The participants were asked to respond "yes" or "no" if they have ever been diagnosed with the condition. Diabetes, although captured in the list, was excluded when the number of chronic conditions was tabulated. The total number of chronic conditions was then grouped into *no chronic diseases*, *one chronic disease*, and *two or more chronic diseases*. The duration of each disease could be calculated through an additional question that asks, "How old were you when you were diagnosed?" However, only the duration of diabetes was used in this study. The quartile values were used for grouping the duration of diabetes into four categories (<4 years, 4-9 years, 10-18 years, and ≥19 years).

Analysis

To ensure the representativeness of the data to the general population, the following weights were used in the analysis: design weights to account for oversampling, poststratification weights for adjusting age and ethnicity distributions, and nonresponse weights. All statistical analyses were performed using Stata MP version 15 (StataCorp LLC), and all descriptive, chi-square, and regression analyses were weighted using survey weights to account for the complex survey design. Descriptive analysis was performed for the variables, and the data were represented as the frequency and weighted percentage of the events. First, bivariate chi-square analyses were performed to compare the sociodemographic characteristics of the diabetes and nondiabetes groups. Second, bivariate chi-square analyses were also conducted to examine how the diabetes and nondiabetes groups differed from each other in terms of the perceived advantages or disadvantages of eHealth. Third, within the full sample, multivariable logistic regression analyses (multiple predictor variables and a single binary outcome variable) were then conducted to determine significant sociodemographic factors (ie, age, gender, ethnicity, education, marital status, monthly personal income, employment, BMI, diabetes diagnosis, and chronic physical conditions) associated with readiness for eHealth (ready vs not ready) and acceptance of eHealth (would not vs would accept). The estimated odds ratios (ORs) for each predictor variable of the multivariable regression models were adjusted for other observable variables or potential confounders entered within the model. Finally, to examine the sociodemographic variables associated with readiness or acceptance within individuals with diabetes, a series of bivariate chi-square or Mann-Whitney U analyses were conducted before the estimation of multivariable logistic regression models. Owing to the limited sample size of individuals who endorsed having diabetes, only variables that showed a significant association in the aforementioned bivariate chi-square analyses were included in the final multivariable logistic models. Missing, refused, or do not know responses were removed listwise, as is the default in multivariable logistic regression models and bivariate chi-square analyses.

Results

Sociodemographic Characteristics of the Sample

In total, 2895 participants were recruited from the general population (screened=5698; response rate 66.2%; eligibility rate 76.8%), of which 15.06% (436/2895) had diabetes, whereas the 84.94% (2459/2895) did not have diabetes. An approximately equal number of all age groups, gender, and ethnicity (Multimedia Appendix 1, Table S1) were recruited. Most participants were married (61.7%) and employed (70.5%). Of the participants, 47.6% had no other chronic illnesses (excluding diabetes), 27.3% had at least one chronic disease, and 24.9% had multiple comorbidities (two or more chronic diseases). Detailed information is presented in Multimedia Appendix 1, Table S1. The diabetes and nondiabetes groups differed significantly in most of the characteristics analyzed (P<.001).

Readiness and Acceptance Toward eHealth

Nearly half of the participants acknowledged that they were not ready for eHealth (47.3%). Readiness varied significantly between the nondiabetes and diabetes samples, with the latter showing significantly lower readiness (Table 1; 54.9% vs 24%; P<.001). A lower acceptance level toward eHealth was observed overall (28%), with the majority (68.4%) unwilling to use eHealth for diabetes care. The diabetes group showed significantly lower acceptance (12.1%) than the nondiabetes group (29.6%; P<.001).

Table 1. Readiness toward eHealth and acceptability of eHealth services for diabetes care.

Response categories	Overall (N=2895)	Nondiabetes (n=2459)	Diabetes (n=436)	*P* value
Readiness toward eHealth technology ("I am not ready for eHealth"), n (%)[a]				<.001
Not ready	1460 (47.3)	1132 (44.4)	328 (75.7)	
Ready	1419 (52.1)	1314 (54.9)	105 (24)	
Do not know or refused[b]	16 (0.6)	13 (0.7)	3 (0.3)	
Acceptance of eHealth for diabetes care ("If you seek treatment for diabetes, would you seek treatment delivered via internet?"), n (%)				<.001
Definitely or possibly would	692 (28)	642 (29.6)	50 (12.1)	
Definitely or possibly wouldn't	2100 (68.4)	1733 (67.1)	367 (81.8)	
Not sure[b]	95 (3.3)	78 (3)	17 (5.9)	
Do not know or refused	8 (0.3)	6 (0.3)	2 (0.2)	

[a]Weighted percentage.

[b]Do not know or refused or not sure options were not included in the bivariate chi-square analysis.

Acceptance Toward Specific eHealth Services

A strong preference for face-to-face sessions over eHealth was observed, especially among the diabetes group for most of the services studied. eHealth services were acceptable for the majority of the participants (overall sample, Table 2) for booking appointments (eHealth only: 33.3%; both eHealth or face-to-face: 34.1%) and for receiving health information (eHealth only: 23%; both: 39.8%). Although the nondiabetes group had roughly the same general acceptance level as the overall sample for eHealth, the diabetes group differed significantly in their acceptance rates, preferring face-to-face services (eg, 55.7% for booking appointments and 63.4% for receiving health information; *P*<.001). Acceptance was lower for other services in the overall sample, such as receiving prescriptions for medications (eHealth: 11.3%; both: 24.8%), consultation with clinicians (eHealth only: 1.3%; both: 13.9%), and receiving referral to other doctors (eHealth only: 13.2%; both: 33%). The diabetes group reported significantly lower acceptance rates for all of the previously mentioned services than the nondiabetes group (*P*<.001). A strong preference for face-to-face services was observed for consultation with clinicians, with 92.6% of the diabetes group preferring face-to-face consultation sessions, compared with 83.5% of the nondiabetes group (*P*<.001).

Table 2. Comparison of acceptance of various eHealth services between diabetes and nondiabetes groups.

Response categories	Overall	Nondiabetes	Diabetes	*P* value
Booking appointment, n (%)				<.001
Face-to-face only	1076 (32.1)	794 (29.7)	282 (55.7)	
eHealth only	854 (33.3)	800 (35.6)	54 (10.8)	
Both	953 (34.1)	858 (34.3)	95 (31.4)	
Do not know, refused, or missing[a]	12 (0.5)	7 (0.3)	5 (2.1)	
Getting prescriptions, n (%)				<.001
Face-to-face only	1957 (63.5)	1599 (61.9)	358 (79.3)	
eHealth only	280 (11.3)	259 (11.7)	21 (6.4)	
Both	648 (24.8)	595 (26.0)	53(13.1)	
Do not know, refused, or missing	10 (0.4)	6 (0.3)	4 (1.2)	
Consultation, n (%)				<.001
Face-to-face only	2521 (84.3)	2118 (83.5)	403 (92.6)	
eHealth only	34 (1.3)	33 (1.5)	1 (0.1)	
Both	329 (13.9)	301 (14.7)	28 (6.1)	
Do not know, refused, or missing	11 (0.4)	7 (0.3)	4 (1.2)	
Health information, n (%)				<.001
Face-to-face only	1262 (36.7)	955 (34)	307 (63.4)	
eHealth only	588 (23)	555 (24.4)	33 (8.6)	
Both	1033 (39.8)	942 (41.2)	91 (25.9)	
Do not know, refused, or missing	12 (0.5)	7 (0.3)	5 (2.1)	
Referrals to other doctors, n (%)				<.001
Face-to-face only	1704 (53.3)	1359 (51.4)	345 (72.2)	
eHealth only	377 (13.2)	356 (14.2)	21 (3.6)	
Both	801 (33.0)	735 (34)	66 (23)	
Do not know, refused, or missing	13 (0.5)	9 (0.5)	4 (1.2)	

[a]Do not know or refused options and missing data were not included in the bivariate chi-square analysis.

Perceived Advantages and Disadvantages of eHealth

The most common disadvantages cited were the requirement of computer literacy to use the eHealth service (90.8%; Figure 1), followed by the perception that eHealth did not build clinician-patient rapport in comparison with the face-to-face sessions (77.8%), lack of confidence about the credibility of the eHealth services (57.5%), and the appropriateness of the eHealth services for their specific medical condition (48.7%). A significantly higher proportion of the diabetes group perceived the credibility of the eHealth services (64.2%) and appropriateness of the app for their specific medical condition (62.3%) as disadvantages compared with the nondiabetes group (56.8% and 47.4%, respectively). As a whole, the advantages cited included (1) eHealth services save time (79.8%; Figure 2), (2) it was more convenient than face-to-face sessions (62.5%), and (3) it was cost-saving (61.8%). Overall, the diabetes group perceived significantly lower advantages than the nondiabetes group, except that eHealth ensured anonymity and privacy. Nearly half of the participants in the diabetes (49.8%) and nondiabetes groups (47.6%) reported that eHealth ensures privacy and anonymity. Participants who showed acceptance (endorsed definitely or possibly would use the service) perceived the advantages significantly higher than those with lower acceptance (*P*<.001; Multimedia Appendix 1, Table S2).

Figure 1. Disadvantages of eHealth reported by the participants. The asterisk represents statistically significant values (*P*<.05).

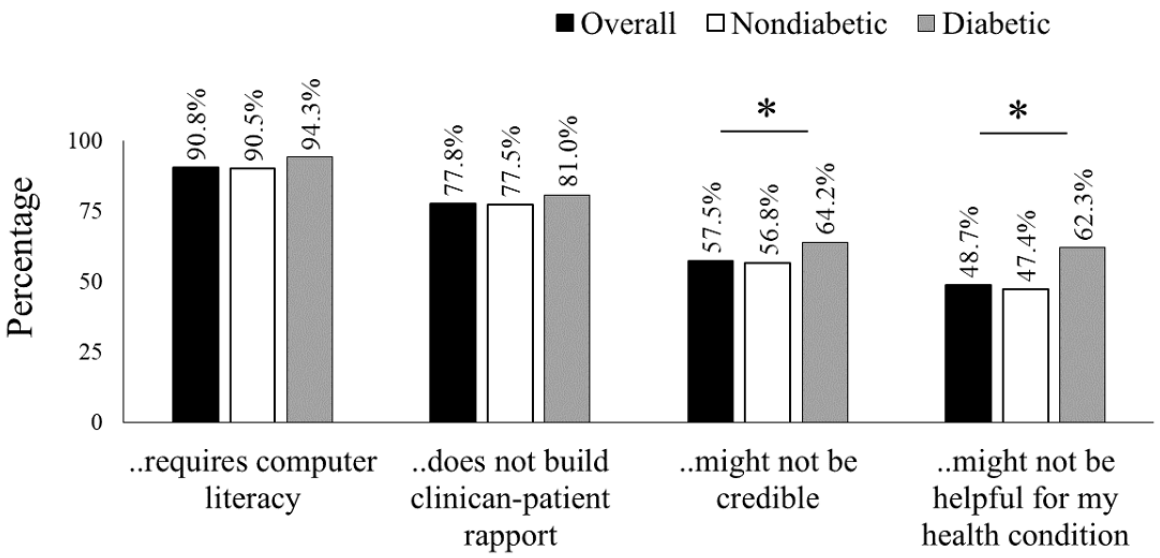
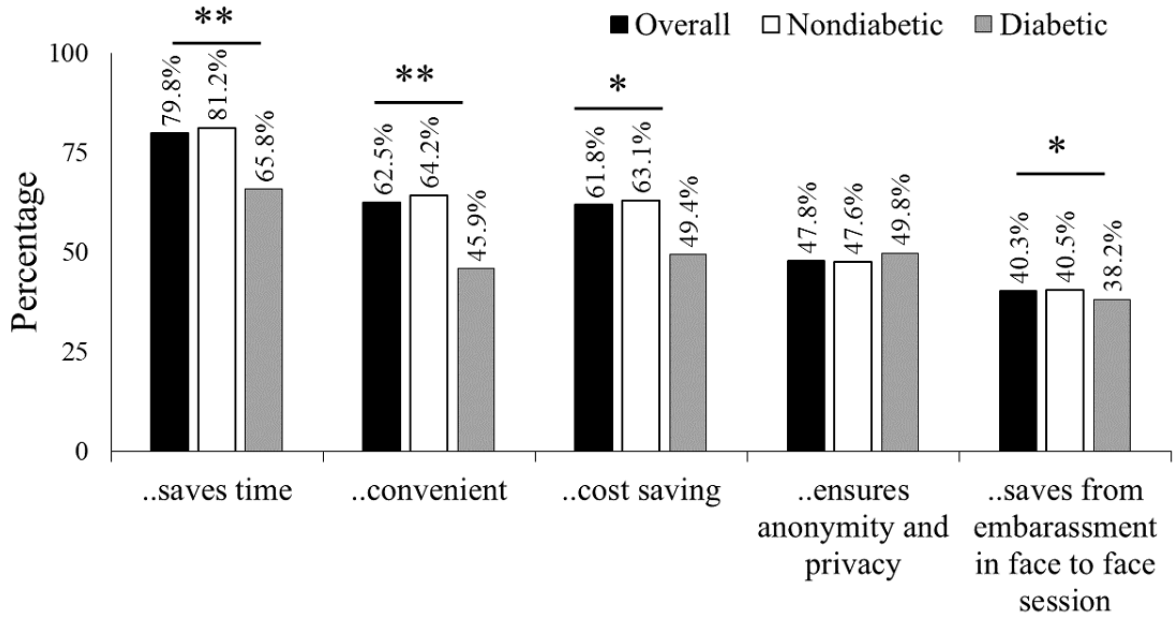


Figure 2. Advantages of eHealth as reported by the participants. Statistical significance is marked with asterisk (**P*<.05; ***P*<.001).



Factors Associated With eHealth Readiness and Acceptance

Factors Associated With eHealth Readiness

Frequencies and weighted percentage counts between sociodemographic variables and readiness toward eHealth are presented in Multimedia Appendix 1, Table S3. The results of the logistic regression analyses are presented in Table 3. The results indicated that age, ethnicity, and education status were associated with eHealth readiness (Multimedia Appendix 1, Table S3 for bivariate analysis). Older adults (aged ≥35 years) had lower odds of endorsing readiness for eHealth than those aged 18-34 years (*P*<.001). With regard to ethnicity, Indians had lower odds of eHealth readiness (OR 0.71, 95% CI 0.54-0.94; *P*=.02) than individuals of Chinese ethnicity. Those with primary school education or below (OR 0.24, 95% CI 0.14-0.44; *P*<.001) and secondary school education (OR 0.48, 95% CI 0.31-0.75; *P*=.001) had lower odds of eHealth readiness than those who had completed a degree or above.

Table 3. Factors affecting eHealth readiness.

Sociodemographic factors	Odds ratio[a] (95% CI)	*P* value
Age group (years)		
18-34 (reference)	__[b]	—
35-49	0.29 (0.19-0.45)	<.001
50-64	0.14 (0.08-0.22)	<.001
≥65	0.05 (0.03-0.11)	<.001
Gender		
Female (reference)	—	—
Male	1.24 (0.93-1.66)	.14
Ethnicity		
Chinese (reference)	—	—
Malay	1.27 (0.95-1.69)	.10
Indian	0.71 (0.54-0.94)	.02
Others	1.40 (0.89-2.21)	.15
Education		
Primary and below	0.24 (0.14-0.44)	<.001
Secondary school	0.48 (0.31-0.75)	.001
Preuniversity or junior college	1.04 (0.54-2.02)	.90
Vocational institute or Institute of Technical Education	0.84 (0.47-1.50)	.56
Diploma	0.98 (0.64-1.49)	.92
Degree, professional certification, and above (reference)	—	—
Marital status		
Single	0.92 (0.60-1.39)	.68
Married or cohabiting (reference)	—	—
Divorced or separated	0.71 (0.34-1.47)	.35
Widowed	1.30 (0.55-3.07)	.55
Employment		
Employed (reference)	—	—
Economically inactive	0.85 (0.55-1.34)	.49
Unemployed	0.98 (0.48-2.01)	.95
Monthly income in SGD (US $)		
Below 2000 (1481; reference)	—	—
2000-3999 (1481-2960)	0.97 (0.66-1.42)	.86
4000-5999 (2961-4441)	1.07 (0.64-1.77)	.80
6000-9999 (4442-7402)	1.48 (0.80-2.74)	.21
≥10,000 (7403)	1.98 (0.97-4.04)	.06
No income	1.07 (0.57-2.01)	.82
BMI		
Underweight	1.06 (0.56-2.00)	.85
Normal range (reference)	—	—
Overweight	0.75 (0.55-1.04)	.08
Obese	1.16 (0.76-1.78)	.50

Sociodemographic factors	Odds ratio[a] (95% CI)	*P* value
Diabetes diagnosis		
No diabetes (reference)	—	—
Has diabetes	0.73 (0.46-1.16)	.18
Number of chronic conditions (excluding diabetes)		
No chronic diseases (reference)	—	—
One chronic disease	0.80 (0.58-1.13)	.20
Two or more chronic diseases	1.11 (0.77-1.59)	.57

[a]Odds ratio >1 indicates higher likelihoods of endorsing readiness for eHealth.

[b]Values not estimated for reference groups.

Factors Associated With eHealth Acceptance

Compared with younger adults (aged 18-34 years), adults in older age groups (aged ≥35 years) were less likely to accept eHealth in the overall sample (Table 4). Indians had lower odds of accepting eHealth (OR 0.54, 95% CI 0.41-0.72; *P*<.001) than Chinese participants. Those who were single (OR 0.62, 95% CI 0.42-0.91; *P*=.01) were less likely to accept eHealth than those who were married or cohabiting. Those with a preuniversity or junior college education (OR 1.90, 95% CI 1.03-3.51; *P*=.04) were more likely to accept eHealth than those with degrees and above. In contrast, those with primary and lower education levels (OR 0.39, 95% CI 0.21-0.75; *P*=.01) were less likely to accept eHealth. People with no income were more likely to accept eHealth than those with income below SGD 2000 (US $1481; OR 1.88, 95% CI 1.06-3.34; *P*=.03). Compared with those with normal BMI, those underweight were less likely to accept eHealth (OR 0.47, 95% CI 0.27-0.83; *P*=.01). A detailed table with results from the bivariate analysis can be found in Multimedia Appendix 1, Table S4.

Table 4. Multivariable logistic regression model examining factors associated with eHealth acceptance within the overall sample.

Sociodemographic variables	Odds ratio[a] (95% CI)	*P* value
Age group (years)		
18-34 (reference)	___[b]	—
35-49	0.56 (0.38-0.85)	.01
50-64	0.31 (0.19-0.51)	<.001
≥65	0.10 (0.05-0.22)	<.001
Gender		
Female (reference)	—	—
Male	1.24 (0.93-1.65)	.15
Ethnicity		
Chinese (reference)	—	—
Malay	1.06 (0.79-1.42)	.70
Indian	0.54 (0.41-0.72)	<.001
Others	0.73 (0.47-1.14)	.17
Education		
Primary and below	0.39 (0.21-0.75)	.01
Secondary school	0.69 (0.43-1.09)	.11
Preuniversity or junior college	1.90 (1.03-3.51)	.04
Vocational institute or Institute of Technical Education	0.57 (0.31-1.05)	.07
Diploma	1.11 (0.73-1.68)	.62
Degree, professional certification, and above (reference)	—	—
Marital status		
Single	0.62 (0.42-0.91)	.01
Married or cohabiting (reference)	—	—
Divorced or separated	0.66 (0.30-1.42)	.28
Widowed	0.98 (0.33-2.93)	.97
Employment		
Employed (reference)	—	—
Economically inactive	0.79 (0.51-1.24)	.31
Unemployed	1.10 (0.58-2.10)	.77
Monthly income in SGD (US $)		
Below 2000 (1481; reference)	—	—
2000-3999 (1481-2960)	0.87 (0.58-1.31)	.51
4000-5999 (2961-4441)	0.73 (0.43-1.24)	.24
6000-9999 (4442-7402)	0.57 (0.31-1.08)	.08
≥10,000 (7403)	1.88 (0.92-3.86)	.08
No income	1.88 (1.06-3.34)	.03
BMI		
Underweight	0.47 (0.27-0.83)	.01
Normal range (reference)	—	—
Overweight	0.75 (0.54-1.04)	.09
Obese	0.67 (0.41-1.07)	.10

Sociodemographic variables	Odds ratio[a] (95% CI)	*P* value
Diabetes diagnosis		
No diabetes (reference)	—	—
Has diabetes	0.65 (0.36-1.18)	.16
Number of chronic conditions (excluding diabetes)		
No chronic diseases (reference)	—	—
One chronic disease	0.76 (0.55-1.06)	.11
Two or more chronic diseases	1.30 (0.88-1.90)	.19

[a]Odds ratio >1 indicates higher likelihood of accepting eHealth.

[b]Values not estimated for reference groups.

Factors Associated With eHealth Readiness and Acceptance in Diabetes Subgroup

The frequencies of diabetes and sociodemographic variables and bivariate chi-square analyses are included in Tables S5 and S6 of Multimedia Appendix 1. Only significant correlates were entered into the final multivariate logistic regression model. Age, education status, and duration of diabetes were significantly associated with eHealth readiness in the diabetes group (Table 5). Those aged 50-64 years (OR 0.08, 95% CI 0.03-0.19; *P*<.001) and ≥65 years (OR 0.06, 95% CI 0.02-0.17; *P*<.001) showed lower odds of eHealth readiness compared with those aged 23-49 years. Participants with a diploma had lower odds of eHealth readiness (OR 0.21, 95% CI 0.06-0.76; *P*=.02) than those with an education level of degree or above. Those with a longer duration of diabetes had higher odds of readiness (4-9 years: OR 2.94, 95% CI 1.07-8.05, *P*=.04; 10-18 years: OR 4.03, 95% CI 1.46-11.18, *P*=.007) than those with <4 years of disease duration.

Table 5. Multivariable logistic regression model examining factors associated with eHealth readiness within the diabetes sample.

Sociodemographic factors	Odds ratio[a] (95% CI)	*P* value
Age (years)		
23-49 (reference)	__[b]	—
50-64	0.08 (0.03-0.19)	<.001
≥65	0.06 (0.02-0.17)	<.001
Ethnicity		
Chinese (reference)	—	—
Malay	0.98 (0.38-2.51)	.96
Indian	0.88 (0.39-1.98)	.76
Others	1.64 (0.39-6.90)	.50
Education		
Primary and below	0.28 (0.07-1.09)	.07
Secondary	0.40 (0.10-1.60)	.19
Preuniversity or junior college	0.62 (0.11-3.52)	.59
Vocational institute or Institute of Technical Education	1.44 (0.31-6.62)	.64
Diploma	0.21 (0.06-0.76)	.02
Degree, professional qualification, and above (reference)	—	—
Marital status		
Single	2.18 (0.74-6.37)	.16
Married or cohabiting (reference)	—	—
Separated or widowed or divorced	0.74 (0.18-3.01)	.68
Employment		
Employed (reference)	—	—
Economically inactive	0.51 (0.15-1.77)	.29
Unemployed	0.57 (0.17-1.91)	.36
Duration of diabetes (years)		
<4 (reference)	—	—
4-9	2.94 (1.07-8.05)	.04
10-18	4.03 (1.46-11.18)	.01
≥19	2.45 (0.73-8.27)	.15

[a]Odds ratio >1 indicates higher likelihood of endorsing readiness for eHealth.

[b]Values not estimated for reference groups.

Age and income were associated with eHealth acceptance and rejection in the diabetes sample (Table 6). Those aged 50-64 years (OR 0.24, 95% CI 0.08-0.75; *P*=.02) had higher odds of rejecting eHealth than those aged 23-49 years. Compared with the participants with income below SGD 2000 (US $1481), those with incomes between SGD 4000 (US $2961) and SGD 3999 (US $2960) were more likely to accept eHealth (OR 7.17, 95% CI 1.61-31.95; *P*=.01). The detailed data from the bivariate analysis are included in Multimedia Appendix 1, Tables S6 and S7.

Table 6. Factors associated with eHealth acceptance in the diabetes sample.

Sociodemographic factors	Odds ratio[a] (95% CI)	*P* value
Age (years)		
23-49 (reference)	—[b]	—
50-64	0.24 (0.08-0.75)	.02
≥65	0.24 (0.05-1.25)	.09
Employment		
Employed (reference)	—	—
Economically inactive	1.00 (0.24-4.23)	.99
Unemployed	0.69 (0.10-4.81)	.70
Personal income in SGD (US $)		
Below 2000 (1481; reference)	1.25 (0.43-3.61)	.68
2000-3999 (1481-2960)	7.17 (1.61-31.95)	.01
4000-5999 (2961-4441)	0.35 (0.05-2.50)	.30
6000-9999 (4442-7402)	1.72 (0.30-10.02)	.54
≥10,000 (7403)	0.24 (0.04-1.41)	.11
No income	1.25 (0.43-3.61)	.68

[a]Odds ratio >1 indicates higher likelihood of accepting eHealth.

[b]Values not estimated for reference groups.

Discussion

Principal Findings and Implications

This study sheds light on the readiness, acceptance, and attitudes of the population, with and without diabetes, toward eHealth services. Nearly half of the population was not ready for the technology, which was more evident in people with diabetes (75.7%). The acceptance level was very low (12% in people with diabetes vs 29.6% in those without diabetes). Age, ethnicity, education, marital status, income, and duration of disease were associated with readiness or acceptance.

Overall, the participants favored face-to-face sessions over eHealth for most of the services surveyed. A strong preference for face-to-face sessions for consultation with clinicians was observed in 84% of the participants and 92% of the diabetes sample. Similar results were noted in other studies in patients with diabetes [23] and in patients with mental illness [24]. It is possible that patients with diabetes feel that diabetes is a complex disease with chances for further health complications; thus, periodic face-to-face sessions with their care providers are required. These sessions answer their queries and help them make important care decisions through discussions with their clinicians. Such interactions and rapport are important in the adaptation to medical conditions and engagement in treatment [25]. In agreement with this, 62.3% of the diabetes group in this study felt that eHealth might not be suitable for their condition, and the majority of the participants in both groups (diabetes and nondiabetes) felt that eHealth did not build clinician-patient rapport to the same extent as face-to-face sessions. Previous studies have identified potential challenges in patient-clinician interaction within the eHealth milieu, wherein 88% of the participants felt that eHealth does not build the same rapport as face-to-face sessions and sharing doctor's phone number or personal email could improve the clinician-patient rapport in the eHealth framework [26]. However, only 2% of the doctors were willing to share their phone number or email address [27] with the patients. Thus, in the absence of additional measures, individuals who value clinician-patient rapport are unlikely to accept eHealth as a replacement for face-to-face sessions [23].

A lower perception of ease of use was also observed in the sample. The most cited disadvantage was the requirement of computer skills, which was identified as a factor that affects the behavioral outcome (perceived ease of use and intention to use the technology) in the Technology Acceptance Model [16]. This observation has also been highlighted in previous studies [24,28]. The usage and navigation of the various functions of apps require training and support, which can adversely affect user acceptance, especially in older adults. This can be addressed by constant training and support by health care staff [29]. Perceived benefits are another factor that determines the acceptance of eHealth services [16,28], especially among people with diabetes [29]. Although the majority of the nondiabetes group agreed that eHealth was convenient and saves time and cost, the diabetes sample showed a significantly lower perception of advantages that could add to their decision to reject the services. Apart from these factors, a user's positive experience with eHealth services and eHealth literacy are important factors that can improve acceptance rates [29]. Thus far, none of the diabetes-related services in Singapore have been offered through eHealth platforms, thereby leaving the participants with no previous experience or knowledge of the technology. This could be the reason for the lower perception of benefits.

Age, education status, income, ethnicity, disease duration, BMI, and marital status were identified as factors associated with readiness and acceptance. We observed that a longer duration of diabetes was associated with higher odds of readiness but not acceptance in patients with diabetes, confirming previous reports. It is postulated that patients with a longer duration of diabetes prefer more personalized care, which involves regular face-to-face discussions with their health care team. In general, they lack trust in new technologies and are reluctant to embrace the new technology that has a lot of uncertainty to replace the comfort of face-to-face sessions [23]. However, as they tend to show readiness toward eHealth, the technology could be introduced slowly alongside their face-to-face session to give a positive experience with the technology that will influence their attitude toward eHealth. Jiang et al [30] studied the sociodemographic factors associated with the acceptance of eHealth in the management of chronic diseases (cancer), where the patients used eHealth for self-management. The study identified similar sociodemographic factors as determinants of eHealth acceptance. Younger adults and those with higher education exhibited higher eHealth literacy and thus had better chances of acceptance and outcome from eHealth use [31]. Gordon et al [32] and Eszes et al [33] also observed social determinants, such as age and ethnicity, to affect eHealth usage and acceptance. Thus, such social determinants should be given attention to avoid health care disparities in underprivileged groups. Overall, the present population in this study had an unfavorable attitude toward eHealth for diabetes care, and this was stronger among patients with diabetes. This is different from the reports from cohort studies where moderate or high levels of acceptance of specific eHealth apps were noted [24,34]. However, all these studies captured acceptance after implementing a specific eHealth program under the care of the attending health care professional, where the patients experienced remote glucose monitoring, diet, and/or physical activity monitoring. It is unclear whether these sessions supplemented their routine face-to-face sessions or replaced them. The participants in this study were technology naïve, as most of the services they encountered were offered face-to-face in the diabetes clinics with no remote monitoring in place. Hence, the lack of previous experience and literacy in eHealth might affect their perceptions and attitudes toward technology.

Overall, the diabetes sample showed a significantly lower readiness and acceptance of eHealth compared with the nondiabetes group. This can be explained based on the differences in sociodemographic characteristics between the two groups (Multimedia Appendix 1, Table S1). Most of the diabetes sample was older adults (aged >34 years), and only 1.8% of the participants were in the 21-34 years age group compared with 32.7% in the nondiabetes group. Nearly half of the diabetes sample (49.1% vs 24.5%) were aged between 50 and 64 years. Our results showed that adults who were aged ≥34 years had higher odds of not favoring eHealth, and the odds increased with age. Thus, the sociodemographic profile of the diabetes sample matches the archetype of individuals who tend not to favor eHealth. The diabetes sample also had a higher proportion of participants from the Indian ethnicity (15.1% vs 7.9%) and those reporting no income (10.9% vs 6.6%), all of which were associated with lower acceptance or readiness toward eHealth. It is also possible that those with diabetes had built a stronger therapeutic relationship with the clinical care team and thus were unwilling to accept new technology, as they feared that it would disrupt their relationship with the clinical team [23]. An in-depth qualitative study that captures the barriers to and facilitators of eHealth acceptance in this group is desirable to understand the needs of this group before introducing eHealth into their routine care.

It is possible that some participants were unaware of existing eHealth platforms in Singapore when answering the questions, and the survey did not capture this information, which is a limitation of this study. The questionnaire used in the study was developed by the study team and has not been validated previously. The diagnosis and duration of the disease were captured through a self-reported measure, which is a limitation of this study. For the duration of diabetes, the participants reported the age at which they were first diagnosed with diabetes, which was subjected to recall bias. It is currently unclear why a significantly higher proportion of patients with diabetes were not in favor of this technology. Age could play an important role, as older participants tend to reject eHealth. With the aging population in Singapore, it is essential to understand the barriers to and facilitators of eHealth acceptance to avoid the underutilization of services by this group. Future research should focus on an in-depth qualitative analysis of the population with and without diabetes with specific age groups to understand the reasons for the lack of acceptance of the technology.

It is also possible that people's perceptions and attitudes toward eHealth evolved during the pandemic period, as many services were disrupted as a consequence of the COVID-19 pandemic. We recruited only a small proportion of the participants (n=16) during this period, and therefore, it was unlikely to have an impact on the results. Globally, patients with diabetes have multiple complications and hyperglycemic episodes with no or limited access to medications or other health care platforms [20]. Although eHealth is a viable option, it is not available to the majority of people worldwide. On the basis of these experiences, the international consensus calls for a transformation of diabetes care to steer toward eHealth rather than going back to the pre–COVID-19 era to prepare better for future disasters [20].

Lower eHealth readiness and acceptance can be a significant barrier to the digitalization of health care services. The general public's eHealth literacy needs to be improved through education and communication before implementing any eHealth services to avoid patient and clinician distress. As patients value clinician-patient rapport, an organizational-level effort is required to improve the clinician-patient rapport in the eHealth framework and to assure patients that comparable care would be delivered under both platforms. Thus, a carefully planned deployment of eHealth to supplement face-to-face sessions rather than replace them would be ideal for improving acceptance and reducing patient dissatisfaction. A gradual transition based on an individual's preferences, capabilities, and needs would result in people feeling comfortable with the technology.

Conclusions

This study showed a negative attitude toward eHealth, with the majority unwilling to use eHealth for their diabetes care. The highest acceptance was noted for booking appointments on the web and receiving health information, whereas face-to-face sessions were preferred for the rest of the services. Participants were aware that eHealth saves time and is convenient; nonetheless, the lower perception of the benefits in patients with diabetes and higher perception of disadvantages is a challenge in accepting the technology for diabetes care. A strong preference for face-to-face sessions was observed with a larger proportion of participants with diabetes, citing the reason that eHealth might not be suitable for their health condition. The lack of clinician-patient rapport, requirement of computer skills, and privacy were highlighted by most of the participants, which needs to be addressed through awareness programs to improve acceptance. Younger age, higher education, marital status, BMI, higher income, and ethnicity were associated with eHealth readiness or acceptance. Attention must be given to the socioeconomic group who are unlikely to use the technology so that they are not affected by health and health care disparities.

Acknowledgments

This study was funded by the National Medical Research Council of Singapore (NMRC NMRC/HSRG/0085/2018).

Authors' Contributions

MS, SAC, ESL, PVAR, and CFS were involved in the conceptualization of the study. PVAR wrote and revised the manuscript. JAV, KR, FDSK, AJ, PVAR, MS, WP, SC, SAC, ESL, CFS, and SS were involved in the questionnaire design. MS, PVAR, FDSK, and KR conducted the cognitive testing of the survey questionnaire. MS, PVAR, FDSK, SS, WP, SC, CBY, and KR were involved in data collection. EA and LJH provided the statistical design and sampling strategy. CBY designed and supported the database. PVAR wrote the manuscript. All authors have reviewed and approved the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary information with detailed analysis.
[DOCX File , 68 KB - jmir_v23i9e26881_app1.docx]

References

1. World Health Organization. Global Report on Diabetes. Geneva: World Health Organization; 2016:1-86.
2. IDF Diabetes Atlas, 9th edition. International Diabetes Federation. 2019. URL: https://www.diabetesatlas.org [accessed 2020-11-12]
3. Tarre K. Applied ICT in the health care industry in Sweden. Invest in Sweden Agency's IT Sweden Project. 2003. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.197.2978&rep=rep1&type=pdf [accessed 2020-11-11]
4. Epping-Jordan J, Bengoa R, Kawar R, Sabaté E. The challenge of chronic conditions: WHO responds. Br Med J 2001 Oct 27;323(7319):947-948 [FREE Full text] [doi: 10.1136/bmj.323.7319.947] [Medline: 11679371]
5. Global health workforce shortage to reach 12.9 million in coming decades. World Health Organization. 2013. URL: https://apps.who.int/mediacentre/news/releases/2013/health-workforce-shortage/en/index.html [accessed 2020-11-11]
6. Alvarez RC. The promise of e-Health - a Canadian perspective. eHealth Int 2002 Sep 17;1(1):4 [FREE Full text] [doi: 10.1186/1476-3591-1-4] [Medline: 12459044]
7. Resolution WHA58.33: Sustainable health financing, universal coverage and social health insurance. World Health Organization - The Fifty-Eighth World Health Assembly, Geneva - Resolutions and Decisions. 2005. URL: https://apps.who.int/gb/ebwha/pdf_files/WHA58-REC1/english/A58_2005_REC1-en.pdf [accessed 2020-11-11]
8. Powers MA, Bardsley J, Cypress M, Duker P, Funnell MM, Hess Fischl A, et al. Diabetes self-management education and support in type 2 diabetes: A joint position statement of the American Diabetes Association, the American Association of Diabetes Educators, and the Academy of Nutrition and Dietetics. J Acad Nutrit Dietet 2015 Aug;115(8):1323-1334. [doi: 10.1016/j.jand.2015.05.012] [Medline: 26054423]
9. Beck J, Greenwood DA, Blanton L, Bollinger ST, Butcher MK, Condon JE, et al. 2017 National standards for diabetes self-management education and support. Diabetes Educ 2020 Feb 24;46(1):46-61. [doi: 10.1177/0145721719897952] [Medline: 31874594]
10. Strelitz J, Ahern AL, Long GH, Hare MJ, Irving G, Boothby CE, et al. Moderate weight change following diabetes diagnosis and 10 year incidence of cardiovascular disease and mortality. Diabetologia 2019 Aug 7;62(8):1391-1402 [FREE Full text] [doi: 10.1007/s00125-019-4886-1] [Medline: 31062041]

11. Castelnuovo G, Mauri G, Simpson S, Colantonio A, Goss S. New technologies for the management and rehabilitation of chronic diseases and conditions. Biomed Res Int 2015;2015:180436 [FREE Full text] [doi: 10.1155/2015/180436] [Medline: 26199935]
12. Huang Z, Tao H, Meng Q, Jing L. Management of endocrine disease. Effects of telecare intervention on glycemic control in type 2 diabetes: a systematic review and meta-analysis of randomized controlled trials. Eur J Endocrinol 2015 Mar;172(3):93-101. [doi: 10.1530/EJE-14-0441] [Medline: 25227131]
13. Pereira K, Phillips B, Johnson C, Vorderstrasse A. Internet delivered diabetes self-management education: a review. Diabetes Technol Ther 2015 Jan;17(1):55-63. [doi: 10.1089/dia.2014.0155] [Medline: 25238257]
14. Foley L, Jiang Y, Ni Mhurchu C, Jull A, Prapavessis H, Rodgers A, et al. The effect of active video games by ethnicity, sex and fitness: subgroup analysis from a randomised controlled trial. Int J Behav Nutr Phys Act 2014 Apr 03;11(1):46 [FREE Full text] [doi: 10.1186/1479-5868-11-46] [Medline: 24694082]
15. Theng Y, Lee JW, Patinadan PV, Foo SS. The use of videogames, gamification, and virtual environments in the self-management of diabetes: a systematic review of evidence. Games Health J 2015 Oct;4(5):352-361. [doi: 10.1089/g4h.2014.0114] [Medline: 26287926]
16. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
17. Hossain N, Yokota F, Sultana N, Ahmed A. Factors influencing rural end-users' acceptance of e-health in developing countries: a study on portable health clinic in Bangladesh. Telemed J E Health 2019 Mar;25(3):221-229 [FREE Full text] [doi: 10.1089/tmj.2018.0039] [Medline: 29664328]
18. Duplaga M. The acceptance of e-health solutions among patients with chronic respiratory conditions. Telemed J E Health 2013 Sep;19(9):683-691 [FREE Full text] [doi: 10.1089/tmj.2012.0306] [Medline: 23734700]
19. Duplaga M. A cross-sectional study assessing determinants of the attitude to the introduction of eHealth services among patients suffering from chronic conditions. BMC Med Inform Decis Mak 2015 Apr 19;15(1):33 [FREE Full text] [doi: 10.1186/s12911-015-0157-3] [Medline: 25927312]
20. Caballero A, Ceriello A, Misra A, Aschner P, McDonnell M, Hassanein M, et al. COVID-19 in people living with diabetes: an international consensus. J Diabetes Complications 2020 Sep;34(9):107671 [FREE Full text] [doi: 10.1016/j.jdiacomp.2020.107671] [Medline: 32651031]
21. AshaRani P, Abdin E, Kumarasan R, Kumar FD, Shafie S, Jeyagurunathan A, et al. Study protocol for a nationwide Knowledge, Attitudes and Practices (KAP) survey on diabetes in Singapore's general population. BMJ Open 2020 Jun 15;10(6):e037125 [FREE Full text] [doi: 10.1136/bmjopen-2020-037125] [Medline: 32540891]
22. Abdin E, Chong SA, Vaingankar JA, Shafie S, Verma S, Luo N, et al. Impact of mental disorders and chronic physical conditions on quality-adjusted life years in Singapore. Sci Rep 2020 Feb 14;10(1):2695 [FREE Full text] [doi: 10.1038/s41598-020-59604-0] [Medline: 32060390]
23. Valikodath NG, Leveque TK, Wang SY, Lee PP, Newman-Casey PA, Hansen SO, et al. Patient attitudes toward telemedicine for diabetic retinopathy. Telemed J E Health 2017 Mar;23(3):205-212 [FREE Full text] [doi: 10.1089/tmj.2016.0108] [Medline: 27336678]
24. Wootton BM, Titov N, Dear BF, Spence J, Kemp A. The acceptability of internet-based treatment and characteristics of an adult sample with obsessive compulsive disorder: an internet survey. PLoS One 2011 Jun;6(6):e20548 [FREE Full text] [doi: 10.1371/journal.pone.0020548] [Medline: 21673987]
25. Hibbard J, Greene J. What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. Health Aff (Millwood) 2013 Feb;32(2):207-214 [FREE Full text] [doi: 10.1377/hlthaff.2012.1061] [Medline: 23381511]
26. Peleg R, Nazarenko E. Providing cell phone numbers and e-mail addresses to patients: The patient's perspective, a cross sectional study. Isr J Health Policy Res 2012 Aug 28;1(1):32 [FREE Full text] [doi: 10.1186/2045-4015-1-32] [Medline: 22929801]
27. Peleg R, Avdalimov A, Freud T. Providing cell phone numbers and email addresses to patients: the physician's perspective. BMC Res Notes 2011 Mar 23;4(1):76 [FREE Full text] [doi: 10.1186/1756-0500-4-76] [Medline: 21426591]
28. AlBar AM, Hoque MR. Patient acceptance of e-health services in Saudi Arabia: an integrative perspective. Telemed J E Health 2019 Sep 01;25(9):847-852. [doi: 10.1089/tmj.2018.0107] [Medline: 30452333]
29. Torbjørnsen A, Småstuen MC, Jenum AK, Årsand E, Ribu L. Acceptability of an mHealth app intervention for persons with type 2 diabetes and its associations with initial self-management: randomized controlled trial. JMIR Mhealth Uhealth 2018 May 21;6(5):e125 [FREE Full text] [doi: 10.2196/mhealth.8824] [Medline: 29784635]
30. Jiang Y, West BT, Barton DL, Harris MR. Acceptance and use of eHealth/mHealth applications for self-management among cancer survivors. Stud Health Technol Inform 2017;245:131-135 [FREE Full text] [Medline: 29295067]
31. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
32. Gordon NP, Hornbrook MC. Differences in access to and preferences for using patient portals and other ehealth technologies based on race, ethnicity, and age: a database and survey study of seniors in a large health plan. J Med Internet Res 2016 Mar 04;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]

33. Eszes D, Szabó DJ, Russell G, Kirby P, Paulik E, Nagymajtényi L, et al. Diabetic retinopathy screening using telemedicine tools: pilot study in Hungary. J Diabetes Res 2016;2016:4529824 [FREE Full text] [doi: 10.1155/2016/4529824] [Medline: 28078306]
34. Banbury A, Nancarrow S, Dart J, Gray L, Parkinson L. Telehealth interventions delivering home-based support group videoconferencing: systematic review. J Med Internet Res 2018 Feb 02;20(2):e25 [FREE Full text] [doi: 10.2196/jmir.8090] [Medline: 29396387]

Abbreviations

OR: odds ratio

Edited by R Kukafka; submitted 01.01.21; peer-reviewed by J Strelitz, TV Duong; comments to author 07.03.21; revised version received 16.03.21; accepted 13.07.21; published 02.09.21.

Please cite as:
AshaRani PV, Jue Hua L, Roystonn K, Siva Kumar FD, Peizhi W, Ying Jie S, Shafie S, Chang S, Jeyagurunathan A, Boon Yiang C, Abdin E, Ajit Vaingankar J, Sum CF, Lee ES, Chong SA, Subramaniam M
Readiness and Acceptance of eHealth Services for Diabetes Care in the General Population: Cross-sectional Study
J Med Internet Res 2021;23(9):e26881
URL: https://www.jmir.org/2021/9/e26881
doi:10.2196/26881
PMID:34473062

Original Paper

Association Between eHealth Literacy in Online Health Communities and Patient Adherence: Cross-sectional Questionnaire Study

Xinyi Lu[1], PhD; Runtong Zhang[2], PhD

[1]School of Management and E-business, Zhejiang Gongshang University, Hangzhou, China

[2]School of Economics and Management, Beijing Jiaotong University, Beijing, China

Corresponding Author:
Xinyi Lu, PhD
School of Management and E-business
Zhejiang Gongshang University
18 Xuezheng Street, Qiantang District
Hangzhou, 310018
China
Phone: 86 18801329327
Email: xinyilu@bjtu.edu.cn

Abstract

Background: eHealth literacy is significantly associated with patients' online information behavior, physician-patient relationship, patient adherence, and health outcomes. As an important product of the internet, online health communities (OHCs) can help redistribute idle medical resources, increase medical resource utilization, and improve patient adherence. However, studies on eHealth literacy in OHCs are limited. Therefore, this study examined patients' eHealth literacy regarding health information–seeking behavior and physician-patient communication in OHCs.

Objective: This study aimed to investigate the association between eHealth literacy in OHCs and patient adherence by employing social cognitive theory.

Methods: This was an empirical study, in which a research model consisting of 1 independent variable (patients' eHealth literacy), 3 mediators (physician-patient communication in OHCs, patient health information–seeking behavior in OHCs, and patients' perceived quality of health information in OHCs), 1 dependent variable (patient adherence), and 4 control variables (age, gender, living area, and education level) was established to examine the associations. Multi-item scales were used to measure variables. An anonymous online survey involving 560 participants was conducted through Chinese OHCs in July 2018 to collect data. Partial least squares and structural equation modeling were adopted to analyze data and test hypotheses.

Results: The survey response rate was 79.6% (446/560). The reliability, convergent validity, and discriminant validity were acceptable. Age, gender, living area, and education level were positively associated with patient adherence, and gender was positively associated with physician-patient communication and patients' perceived quality of internet health information in OHCs. Patients' eHealth literacy was positively associated with patient adherence through the mediations of physician-patient communication, internet health information–seeking behavior, and perceived quality of internet health information in OHCs.

Conclusions: Results indicate that physician-patient communication, internet health information–seeking behavior, and the perceived quality of internet health information are significantly associated with improving patient adherence via a guiding of eHealth literacy in OHCs. These findings suggest that physicians can understand and guide their patients' eHealth literacy to improve treatment efficiency; OHCs' operators should this strengthen the management of information quality, develop user-friendly features, and minimize the gap between the actual and perceived information quality.

(J Med Internet Res 2021;23(9):e14908) doi:10.2196/14908

KEYWORDS

online health communities; OHCs; eHealth literacy; patient adherence; health information; physician-patient communication

Introduction

Background

Health literacy is defined as "the degree to which an individual has the capacity to obtain, communicate, process, and understand basic health information and services to make appropriate health decisions," [1] but individual's skill to obtain and use health information needs to be redefined with the development of medical internet services [2]. Accordingly, eHealth literacy has emerged [3], which is defined by Norman and Skinner [4] as individuals' ability to "seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem." eHealth literacy is related to health literacy, and both of them involve the skills of seeking, appraising, and applying health information [5]. Gilstad [6] added the communication of health information into the definition of eHealth literacy, while Neter and Brainin [5] and Paige et al [7] suggested that eHealth literacy is associated with health-related behaviors, communication skills with physicians, self-management, and health outcomes.

In recent years, online health communities (OHCs) have provided the public with new platforms to obtain health information, share medical experiences, and communicate with physicians. OHCs can help redistribute idle medical resources, improve medical resource utilization, and enhance the physician-patient relationship [8-10]. Patients' eHealth literacy in OHCs has become an important topic to be studied. Previous studies show that patients with adequate eHealth literacy have a higher level of ability to seek internet health information and are more likely to obtain reliable and high-quality health information online than are patients with inadequate eHealth literacy [11,12]. In addition, a high level of eHealth literacy can enable patients to communicate with physicians and improve communication efficiency [11,13].

Physician-patient communication and health information–seeking behavior in OHCs are related to patient adherence, which is defined as "the extent to which a person's behavior (in terms of taking medications, following diets, or executing lifestyle changes) coincides with medical or health advice" [14-16]. Given that physicians cannot constantly participate in patients' daily life, the ability of self-management and self-monitoring is important for patients to maintain a healthy lifestyle. Treatments, medical regimens, and therapies are more likely to be effective if patients take medication in accordance with prescriptions and physicians' advice [17]. Therefore, we speculate that patients' eHealth literacy is likely to be associated with their adherence. Previous studies have focused on the relationship between health literacy and patient adherence or between OHCs and patient adherence [18,19]. However, studies on the association between eHealth literacy in OHCs and patient adherence are limited because of the short development time of medical internet services, OHCs, and eHealth literacy. Given the low level of medical resource utilization, uneven medical resource distribution, and serious hospital congestion in China, this study aims to identify the association between patients' eHealth literacy in OHCs and patient adherence through empirical study and by considering physician-patient communication, internet health information–seeking behavior, and the perceived quality of internet health information as mediators. We hope we can propose productive ideas for improving patient adherence that can ultimately alleviate these aforementioned problems. This study can enrich the theoretical study of OHCs, eHealth literacy, and patient adherence, and may have practical implications for managing OHCs and improving patient adherence through eHealth literacy in OHCs.

Patient Adherence

Patient adherence plays a vital role in health management and health care, especially for patients with chronic diseases [20]. Horwitz et al [20] proved that patient adherence is positively associated with health outcomes and that patients who adhere strictly to medical regimens and physicians' suggestions are relatively healthier than those who do not. Previous studies have confirmed that low adherence may lead to serious consequences for patients, the economy, and society. For patients with low adherence, diseases may be less likely to be controlled or cured, thereby increasing the morbidity and mortality [21]. For the economy, if patients do not adhere to medical regimens or physicians' advice, therapies cannot achieve the intended outcomes, which may waste medical resources [22,23]. In terms of social effect, the unbalanced medical resource allocation may be serious. For example, medical resources may be wasted in some hospitals in economically developed regions, while hospitals in some regions may lack sufficient resources. Moreover, some genuinely beneficial drugs may be mistaken as useless and then be discontinued because patients do not take medications in accordance with prescriptions [23].

Given that patient adherence is a dynamic parameter [24], patient adherence can be enhanced to improve patients' outcomes, increase the utilization of medical resources, advance the clinical testing of drugs, and promote the stable development of society.

Online Health Communities

This study mainly discusses the OHCs in which patients and physicians participate. Patients are increasingly using OHCs to seek health information and communicate with physicians about medical and emotional support because of certain benefits brought by OHCs [25,26]. First, patients can ask questions and communicate with physicians by creating posts and sending online messages anytime and anywhere without visiting hospitals [27]. Moreover, users can conveniently access OHCs' corresponding apps using smartphones. Second, patients' privacy can be protected as meeting face-to-face is unnecessary, and OHCs generally allow patients to ask questions anonymously [28]. Third, OHCs can help alleviate hospital congestion and integrate idle medical resources [9,10,15]. Specifically, patients can diagnose a few basic and simple symptoms by themselves based on the support from OHCs and do not need to visit hospitals frequently or wait extensively. In turn, physicians can spend their time communicating with patients in OHCs if there are only a few patients present in hospitals. In this way, medical resource utilization can be improved.

However, several drawbacks of OHCs must be considered. The observation of patients' breath, sound, and facial expressions is important for physicians in diagnosing illness accurately [29], which is difficult to achieve though OHCs. In this type of situation, physicians may misunderstand patients' conditions and provide inaccurate advice, which may decrease patients' satisfaction. In addition, some patients may worry that they cannot receive answers from actual physicians. Given OHCs' advantages and persistent development, studies on OHCs should be conducted to decrease or avoid these shortcomings.

eHealth Literacy

eHealth literacy is nested in the context of a digital environment. Based on Norman and Skinner [4], Gilstad [6] extended the definition of eHealth literacy to include an individual's skill in identifying, defining, and solving a health problem by communicating, seeking, understanding, appraising, and applying health information and digital technologies under a given cultural, social, and situational background [30]. Paige et al [31] further proposed a transactional model to update the definition of eHealth literacy to include a patient's skill in interacting with the external environment and in counteracting the negative effects of environmental factors. eHealth literacy is a central skill [31] and is significantly associated with health-related behavior [11,32], self-management skills, patients' health-related decision-making, and health outcomes [33,34]. For example, adequate eHealth literacy can help patients understand their conditions accurately, obtain health-related information and knowledge, and improve the ability to self-manage their health and communicate with physicians [11]. In addition, eHealth literacy is associated with health responsibility, self-actualization, and the relationship between patients and health care providers [35].

Previous studies have verified that patients with a high level of eHealth literacy are good at seeking, selecting, and assessing health information from many sources using additional search strategies [5,32], whereas patients with limited eHealth literacy may find the use of online health-related resources difficult. Freemann et al [36] asserted that eHealth literacy plays an important role in adolescents' daily life. Mitsutake et al [32] found that eHealth literacy can promote individuals' health behaviors, such as physical exercise and balanced nutrition. Bodie and Dutta [37] suggested that eHealth literacy can help develop patients' ability to solve specific health problems by themselves using information obtained from the internet, thereby possibly encouraging patients' health behavior. In this study, we focused on patients' eHealth literacy in OHCs. Specifically, we concentrated on patient behavior in seeking, obtaining, understanding, and evaluating health-related information in OHCs and the association between eHealth literacy and patient adherence.

Internet Health Information

In terms of content, health information can be divided into 2 categories: health care information and healthy lifestyle information [5,38,39]. The internet has become the main source for individuals to seek health information [40], and various institutions, such as governments, medical institutions, and business corporations, have established health-related websites to provide information-seeking platforms for the public [41]. According to Wilson [42], combined with the context of the internet, internet health information behavior can be categorized as internet health information–seeking behavior, internet health information use behavior, online communication behavior, and passive receiving of internet health information behavior. According to Neter et al [43], individual's online behaviors in the context of Web 1.0 are mainly consumption activities, such as reading others' experiences, watching videos, and sending or receiving emails, while behaviors in the context of Web 2.0 are mainly production activities, such as rating physicians and hospitals, communicating with others, and sharing their experiences. The aim of individual's use of OHCs is obtaining health information, which indicates that this behavior is active. This study intends to identify the association between eHealth literacy in OHCs and patient adherence. eHealth literacy requires the skill of seeking health information from digital sources. However, having a capacity to seek information online does not mean that the individual would like to perform health information–seeking behavior. Rather, eHealth literacy may be associated with an individual's internet health information–seeking behavior. Therefore, we mainly focused on the internet health information–seeking behavior and online communication behavior in OHCs, which includes Web 1.0 and Web 2.0 health-related activities.

Given the zero gatekeeping and zero-cost publishing of the internet, information can be published and spread quickly and in timely fashion [44], and patients can obtain health information conveniently through the internet, which may help improve their ability in health-related problem-solving [45]. However, in consideration of the universal accessibility of the internet, internet health information has several shortcomings, such as the quality of information [46]. For one, each person is allowed to publish health information online regardless of its correctness, as, for example, most websites in China lack a strict review mechanism for health information. For another, individuals may have inadequate eHealth literacy to appraise the quality of health information or select reliable facts [5]. With limited eHealth literacy, individuals are unable to accurately appraise health information and may at times regard actual high-quality information as low quality or regard actual low-quality information as high quality. For example, a person who does not have adequate eHealth literacy may not be able to find reliable health information from online sources, but sometimes he or she may consider the information obtained from an official health portal as high quality. Therefore, patients' eHealth literacy may be associated with their perceived quality of internet health information, and thus this study concentrated on patients' perception of internet health information quality in OHCs.

Physician-Patient Communication

Originally, the communication between physicians and patients could only occur in hospitals. A new form of physician-patient communication has emerged with the internet, and OHCs serve as important platforms. Because of the advantages of OHCs, patients are willing to communicate with physicians online before or after visiting hospitals [25,47], which is beneficial for improving the efficiency of offline treatments and saving time.

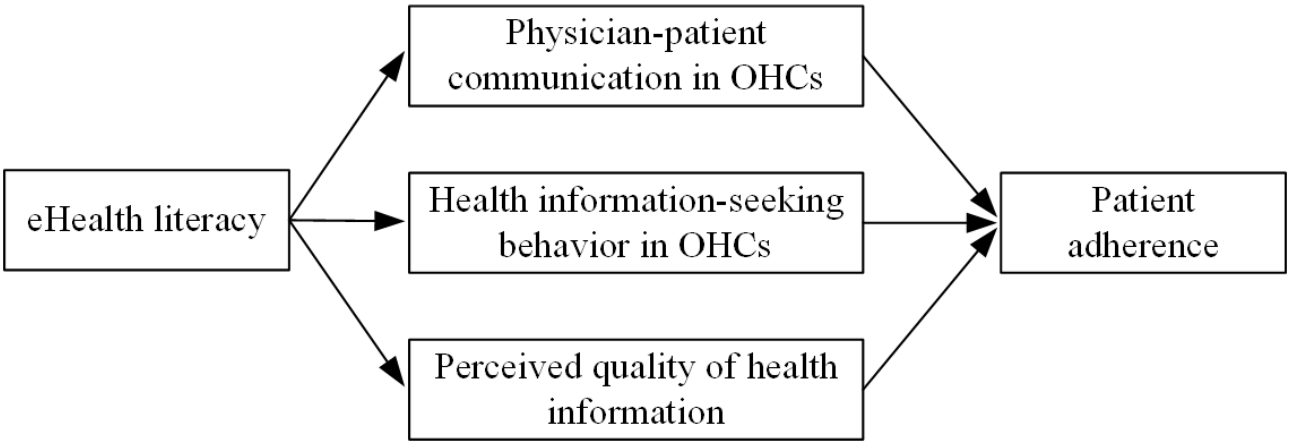
Patients can initially assess their conditions through communication with physicians in OHCs before visiting hospitals, which may help them select the relevant health care department and minimize unnecessary checks [48]. In addition, communicating with physicians in advance can help patients gain information of physicians and decrease the uncertainty in offline treatments [10].

The internet can provide more opportunities to communicate [49]. It is better to practice individuals' communication skills through online channels than on offline channels. Therefore, online communication may be more effective than offline communication. To specify, without a face-to-face meeting, patients and physicians do not need to answer immediately. Therefore, they have adequate time to reconsider their responses, and patients' sense of unfamiliarity and nervousness can be reduced. In addition, conflicts between physicians and patients may be avoided in OHCs, thereby improving communication efficiency and quality and promoting patients' satisfaction with physicians [45].

Model and Hypotheses

Human behavior can be associated with social information obtained from their environment [50]. Specifically, individuals can obtain information from their social environment when they lack knowledge, which may influence their opinions, beliefs, and behavior [51]. In the context of this study, patients actively obtain health information from OHCs by communicating with physicians and searching for information to satisfy their demand for health information and to improve their self-efficacy, which may be associated with patient health-related behavior. According to the review by Neter and Brainin [5], eHealth literacy is associated with individuals' health behaviors. Therefore, we established a research model (Figure 1), which involves 1 independent variable, 3 mediators, and 1 dependent variable to identify the association between eHealth literacy in OHCs and patient adherence through the mediations of physician-patient communication, patient health information–seeking behavior, and patients' perceived quality of health information by employing social cognitive theory. Social cognitive theory can be used to predict health-related behaviors [52,53]. Self-efficacy is a critical factor in social cognitive theory [54], and it can influence an individual's way of thinking, thereby self-regulating motivations of behavior. Specifically, a high level of self-efficacy can encourage individuals to change their behavior, while a low level of self-efficacy may make individuals feel frustrated and refuse to change.

Figure 1. Research model. OHC: online health community.



Communication is an important skill related to eHealth literacy [5,6], and patients' communication behaviors may be associated with individuals' eHealth literacy. Patients with limited eHealth literacy may not have favorable information communication skills and may not be able discuss their illnesses and conditions with physicians clearly [55]. The lack of eHealth literacy may decrease patients' self-efficacy in communication with physicians in OHCs. According to social cognitive theory [54], patients' intention to communicate with physicians may be reduced and they may not want to communicate with physicians through OHCs. We thus assumed that eHealth literacy is associated with communication between physicians and patients in OHCs and proposed the following hypothesis: H1*:* Patients' eHealth literacy is positively associated with physician-patient communication in OHCs.

Seeking health information is another way for patients to obtain health information in OHCs, and it may be associated with eHealth literacy. First, patients with limited eHealth literacy may not be familiar with OHCs. Even if these patients realize the existence of OHCs, they may not be aware that health information can be searched for through these portals [56]. Second, patients with low eHealth literacy may not be good at evaluating the quality of information obtained from OHCs [12], which may decrease patients' intention to seek health information online. Furthermore, patients with inadequate eHealth literacy may find it difficult to obtain high-quality health information. Third, inadequate eHealth literacy makes patients lack adequate confidence in seeking suitable health information in OHCs [11]. Accordingly, under the guidance of social cognitive theory [54], patients with limited eHealth literacy may have low a level of self-efficacy in seeking health information through OHCs, which may inhibit their information-seeking behavior in OHCs. We thus proposed the following hypotheses: H2*:* Patients' eHealth literacy is positively associated with patients' internet health information–seeking behavior in OHCs; H3*:* Patients' eHealth literacy is positively associated with

patients' perceived quality of internet health information in OHCs.

Patients' attitudes and behaviors can be associated with health information obtained from OHCs [50]. Therefore, we assume that, as two ways for patients to obtain health information in OHCs, communicating with physicians and seeking health information are associated with patient adherence. The physician-patient relationship can be enhanced through effective communication, which may promote information processing and decision-making [57,58]. Therefore, patients are likely to perceive high-quality health information and consider their physicians to be professional [59], ultimately improving their self-efficacy. A high level of self-efficacy can encourage patients to adopt health-related behavior [54], such as adherence with physicians' advice. Indeed, Roberts et al [60] confirmed that communication between patients and physicians can improve patient adherence. In terms of health information–seeking in OHCs, generally, the information obtained by patients in OHCs is often basic health-related knowledge, such as medical terminologies and diagnoses, which has been already grasped by physicians [47]. When patients are aware of that fact, they are more likely to trust in their physicians' professional competence. Thus, patients may be increasingly willing to follow the regimens and advice given by physicians. These considerations produce the following hypotheses: H4*:* Physician-patient communication in OHCs is positively associated with patient adherence; H5: Patients' internet health information–seeking behavior in OHCs is positively associated with patient adherence; H6: Patients' perception of internet health information quality is positively associated with patient adherence.

In addition, control variables (age, gender, living area, and education level) were included in this research model to examine and adjust the effects of demographic factors on the research model.

Methods

Instrument Development

The survey instrument (see Multimedia Appendix 1) was developed on the basis of scales validated by previous studies. Specifically, eHealth literacy was measured using a 10-item scale consisting of 8 main items adopted from Norman and Skinner [61] and 2 supplementary items adopted from Park and Lee [62]. Physician-patient communication was measured using a 14-item scale adopted from Makoul et al [63]. A 4-item health information–seeking scale reflects the means by which people search and apply health information actively and passively [64]. Laugesen et al [47] adopted a 16-item scale to measure the perceived quality of internet health information and a 5-item scale to measure patient adherence.

We conducted a formal investigation of Chinese individuals who had communicated with physicians and sought health information in Chinese OHCs. Therefore, the questionnaire was translated into Chinese. The translation process was divided into 3 stages [65,66]. First, 3 native Chinese speakers who had at least a master's degree in English and were skilled in scientific research translation were recruited to translate the instrument into Chinese. Second, we conducted a pretest to acquire advice for improving the comprehensibility, conciseness, readability, and cross-cultural adaptation of the instrument. Third, we recruited English professionals to translate the Chinese questionnaire back to English and to compare the final English version with the original English version to guarantee the conceptual consistency between the 2 versions.

Data Analysis

Our participants were those who had obtained health information by communicating with physicians and seeking health information in OHCs within the previous month. Thus, they could recall relevant experiences. The formal survey was anonymously conducted through several Chinese OHCs in July 2018, and we guaranteed that the privacy of participants would be strictly protected.

As a comprehensive and nonparametric structural equation modelling (SEM) approach, partial least squares (PLS)-SEM is effective in analyzing complex models, providing robust model estimations, and evaluating the quality of predicting results [67-69], especially in the case of complex models or small sample sizes. In addition, PLS-SEM can be used to conduct mediation analysis [68]. Therefore, this study used PLS-SEM to test hypotheses and analyze the research model [47,68] and adopted SmartPLS software version 3.2.8 to analyze data. First, we re-evaluated the reliability and validity of scales since the research context and participants in this study were different from previous works. Specifically, we calculated Cronbach α to assess the reliability and adopted confirmatory factor analysis to assess the convergent validity and discriminant validity. Second, we calculated the magnitude and significance of path coefficients to examine the effects of control variables and test hypotheses, and used multivariate coefficient of determination (R^2) to calculate Cohen f^2 and used goodness of fit (GoF) to analyze the effect sizes of the research model and the fit between the research model and observed data [70]. Third, we further examined the mediations of physician-patient communication, patient health information–seeking behavior, and the perceived quality of health information using the bootstrapping method (n=5000; 95% CI).

Results

We sent the questionnaire to 560 participants and received 446 responses, 381 of which were valid. Accordingly, the response rate was 79.6% (446/560), and the validity rate was 85.4% (381/446). As shown in Table 1, 58.3% (222/381) of the participants were 20 to 40 years old, 53% (202/381) were female, and 51.4% (196/381) had at least a bachelor's degree, all which accounted for more than half of the sample. Given that the participants were OHC users who tended to be young, female, and highly educated, our sample could be considered representative [71,72].

Table 1. Sample demographics (N=381).

Demographic characteristics	Value, n (%)
Age (years)	
<20	18 (4.7)
20-29	107 (28.1)
30-39	115 (30.2)
40-49	89 (23.4)
50-59	48 (12.6)
60 and above	4 (1.0)
Gender	
Male	179 (47.0)
Female	202 (53.0)
Living area	
Urban	212 (55.6)
Rural	169 (44.4)
Education	
Junior middle school and below	19 (5.0)
High school	50 (13.1)
Junior college	116 (30.4)
Bachelor's degree	144 (37.8)
Master's degree	41 (10.8%)
PhD	11 (2.9%)

The Cronbach α of eHealth literacy, physician-patient communication, internet health information–seeking behavior, perceived quality of internet health information, and patient adherence were .849, .898, .709, .905, and .771, respectively. Therefore, the reliability of the scales was acceptable [73]. The Kaiser-Meyer-Olkin value was .964, so the collected data could be used to conduct factor analysis [74]. This study adopted the composite reliability and the average variance extracted (AVE) to evaluate the validity of scales. As shown in Table 2, for each construct, the composite reliability was above 0.700 and the AVE was above 0.500, indicating an acceptable convergent validity [75]. Table 3 shows the correlations between constructs, and the discriminant validity was acceptable because the square root of AVE was greater than the correlations between other constructs and themselves [75].

Table 2. Composite reliability and average variance extracted.

Construct	CR[a]	AVE[b]	Sqrt[c] AVE
eHealth literacy	0.916	0.522	0.723
Physician-patient communication in OHCs[d]	0.937	0.514	0.717
Internet health information–seeking behavior in OHCs	0.834	0.556	0.746
Perceived quality of internet health information in OHCs	0.943	0.506	0.712
Patient adherence	0.845	0.523	0.723

[a]CR: composite reliability.

[b]AVE: average variance extracted.

[c]Sqrt: square root.

[d]OHC: online health community.

Table 3. Correlations between constructs.

Construct	EHL[a]	PPC[b]	IHISB[c]	PQIHI[d]	PA[e]
EHL	1.000	—[f]	—	—	—
PPC	0.674	1.000	—	—	—
IHISB	0.632	0.690	1.000	—	—
PQIHI	0.665	0.681	0.681	1.000	—
PA	0.673	0.692	0.669	0.650	1.000

[a]EHL: eHealth literacy.

[b]PPC: physician-patient communication.

[c]IHISB: internet health information–seeking behavior.

[d]PQIHI: perceived quality of internet health information.

[e]PA: patient adherence.

[f]Not applicable.

In terms of the effects of demographic factors, results indicated that 4 control variables (age, gender, living area, and education level) were positively associated with patient adherence, and gender was positively associated with physician-patient communication and patients' perceived quality of internet health information in OHCs. Specifically, older patients were more willing to take medication and maintain a healthy lifestyle in accordance with medical regimens and physicians' advice than were younger patients; female patients were more likely to communicate with physicians through OHCs, perceive a higher quality of information, and be more willing to adhere to treatments and physicians than were male patients; patients who lived in rural areas were more compliant with physicians than those who lived in urban areas; highly educated patients were more willing to adhere to physicians than were patients with a low level of education. However, all path coefficients of relationships related to control variables were nearly zero, indicating weak effects. The multivariate coefficient of determination (R^2) was used to calculate Cohen f^2 to further evaluate the size of the control variables' effects [76] as listed in Table 4. We confirmed that the effect sizes of the control variables were so small that they could be considered insignificant.

Table 4. Multivariate coefficient of determination (R^2) results.

Variables	R^2		Control variable effects		
	In	Out	ΔR^{2a}	f^2[b]	Effects
Physician-patient communication					
Control variables	0.722	0.720	0.002	0.007	Insignificant
eHealth literacy	0.677	0.004	0.673	2.084	Large
Internet health information–seeking behavior					
Control variables	0.537	0.534	0.003	0.006	Insignificant
eHealth literacy	0.537	0.003	0.534	1.153	Large
Perceived quality of internet health information					
Control variables	0.677	0.676	0.001	0.003	Insignificant
eHealth literacy	0.677	0.002	0.675	2.090	Large
Patient adherence					
Control variables	0.617	0.613	0.004	0.010	Insignificant
Physician-patient communication	0.617	0.571	0.046	0.120	Small
Internet health information–seeking behavior	0.617	0.608	0.009	0.023	Small
Perceived quality of internet health information	0.617	0.604	0.013	0.034	Small

[a]ΔR^2: $R^2_{In} - R^2_{Out}$.

[b]f^2: Cohen f^2.

GoF was used to evaluate the fit between the research model and observed data [70,77] and was calculated as follows:

In this study, the GoF value was 0.578, which indicated that the fit between the data and the research model was good [78].

The magnitude and significance of path coefficients revealed that all 6 hypotheses were supported (Table 5). In particular, eHealth literacy was positively associated with physician-patient communication, internet health information–seeking behavior, and perceived quality of internet health information in OHCs. Physician-patient communication, internet health information–seeking behavior, and perceived quality of internet health information in OHCs were positively associated with patient adherence. We further analyzed the effect sizes of the independent variable and mediators, the results of which are presented in Table 4. We found that the effects of physician-patient communication, internet health information–seeking behavior, and the perceived quality of internet health information in OHCs on patient adherence were weak with small effect sizes, while the effects of eHealth literacy on physician-patient communication, internet health information–seeking behavior, and the perceived quality of internet health information in OHCs were strong with large effect sizes.

Table 5. Results of hypothesis testing.

Hypothesis	Path coefficient	*t*	*P* value
H1: Patients' eHealth literacy is positively associated with physician-patient communication in OHCs.	0.849	46.484	<.001
H2: Patients' eHealth literacy is positively associated with patients' internet health information–seeking behavior in OHCs.	0.732	26.479	<.001
H3: Patients' eHealth literacy is positively associated with patients' perceived quality of internet health information in OHCs.	0.822	33.227	<.001
H4: Physician-patient communication in OHCs is positively associated with patient adherence.	0.429	4.423	<.001
H5: Patients' internet health information–seeking behavior in OHCs is positively associated with patient adherence.	0.156	2.408	0.02
H6: Patients' perception of internet health information quality is positively associated with patient adherence.	0.247	2.918	0.004

According to Baron and Kenny [79], we adopted a bootstrapping method (n=5000, 95% CI) to further examine the mediations. The results of *a*, *b*, *c*, and *c*' are shown in Table 6, and the CIs (bias corrected) of *a* and *b* are shown in Table 7. Under the guidance of Wen and Ye [80], we could conclude that physician-patient communication and patient internet health information–seeking behavior in OHCs played partially mediating roles between eHealth literacy and patient adherence, and the mediating effect accounted for 30.07% and 13.25% of the total effect, respectively.

Table 6. Parameters of mediating effects.

M	*X*	*Y*	*a* (*P*)	*b* (*P*)	*c* (*P*)	c' (*P*)
PPC[a]	EHL[b]	PA[c]	0.849(<0.001)	0.291 (.005)	0.747 (<.001)	0.271 (.001)
IHISB[d]	EHL	PA	0.733(<0.001)	0.135 (.03)	0.747 (<.001)	0.271 (.001)
PQIHI[e]	EHL	PA	0.823(<0.001)	0.157 (.06)	0.747 (<.001)	0.271 (.001)

[a]PPC: physician-patient communication.

[b]EHL: eHealth literacy.

[c]PA: patient adherence.

[d]IHISB: internet health information–seeking behavior.

[e]PQIHI: perceived quality of internet health information.

Table 7. Confidence intervals (bias corrected).

M	*X*	*Y*	CIs (bias corrected) of *a*		CIs (bias corrected) of *b*	
			2.5%	97.5%	2.5%	97.5%
PPC[a]	EHL[b]	PA[c]	0.805	0.880	0.082	0.490
IHISB[d]	EHL	PA	0.668	0.776	0.013	0.260
PQIHI[e]	EHL	PA	0.770	0.863	–0.005	0.313

[a]PPC: physician-patient communication.

[b]EHL: eHealth literacy.

[c]PA: patient adherence.

[d]IHISB: internet health information–seeking behavior.

[e]PQIHI: perceived quality of internet health information.

Discussion

Principal Results

This study explored the association between patients' eHealth literacy in OHCs and their adherence to treatment regimens and physicians' suggestions, and it has theoretical contributions and practical implications for studies on OHCs, eHealth literacy, and patient adherence. We constructed a research model which clarified that patients' eHealth literacy in OHCs can help improve their adherence by guiding their communication behavior with physicians, strengthening their health information–seeking behavior, and increasing their perceived quality of information in OHCs. This study enriches the theoretical research on OHCs, eHealth literacy, and patient adherence, and reduces the deficit in research related to improving patient adherence by strengthening patients' eHealth literacy in OHCs in the Chinese context. In addition, this study adopted social cognitive theory in developing the research model for examining the relationship between eHealth literacy in OHCs and patient adherence, which enriches the application of social cognitive theory in the field of health behavioral psychology.

In terms of practical implications, our findings suggest that physician-patient communication, internet health information–seeking behavior, and patients' perceived quality of internet health information are perspectives from which patient adherence can be enhanced by developing patients' eHealth literacy in OHCs. First, among the 4 mediators, patients' eHealth literacy was the strongest associated with physician-patient communication in OHCs, and physician-patient communication had the strongest association with patient adherence. Physician-patient communication plays a significant mediating role between eHealth literacy in OHCs and patient adherence, which suggests that physicians can encourage patients to communicate with them through OHCs in addition to offline treatments. OHCs can adopt some incentives to improve the possibility of patients' participation in communication with physicians.

Second, in OHCs, the association between patients' eHealth literacy and their perceived quality of internet health information was slightly weaker than that between eHealth literacy and physician-patient communication, but the effect size remained large. eHealth literacy can improve patients' ability to appraise information quality accurately, ultimately encouraging them to adhere to physicians and the recommended treatments. This finding suggests that physicians and OHCs should focus not only on the actual information quality but also the perceived information quality. Evidence from Silver [81] reveals that some patients prefer to trust online sources rather than health care providers, but the limited level of eHealth literacy may make the treatment difficult. Previous studies suggest that eHealth literacy can be enhanced by considering several factors, such as education, computer skills, physical exercise [82], and medical experiences [81]. Therefore, physicians should aim to understand the level of patients' eHealth literacy and their perception of quality during the communication process so as to better understand the possibility and extent of patients obtaining unsuitable health information. Physicians can then help patients improve the efficiency of treatment. OHCs should be developed to be convenient and easy to use, and to provide detailed guidelines that can aid patients' usage. This may bring benefits to improving patients' eHealth literacy and their perception of information quality in OHCs. In addition, to further improve OHCs, operators of OHCs should strengthen the management of information quality and acquire patients' feedback on the perceived quality of health information. Third, eHealth literacy is associated with patient adherence through the mediation of internet health information–seeking behavior although the associations between eHealth literacy and internet health information–seeking behavior and between internet health information–seeking behavior and patient adherence were relatively weaker compared with the other 2 mediators. Schulz et al [12] found that patients with adequate eHealth literacy are more willing to seek health information online than those with inadequate eHealth literacy. The positive association between patients' health information–seeking behavior and patient adherence is supported by Zhang et al [14]. However, some patients with limited eHealth literacy find it difficult to use or find suitable and reliable information through OHCs [56], thereby decreasing their intention to seek health information through OHCs. Given the advantages of OHCs, physicians can discuss the benefits and drawbacks of seeking information online with patients and encourage patients to obtain information by themselves. In addition, when patients find that their physicians have grasped the health information that the patients had themselves already acquired, they will consider their physicians to be professional and thus be more willing to adhere to their

advice. Although patients may be unwilling to seek heath information online, they can perceive that their physicians are sincere and indeed concerned about them. Accordingly, discussion between physicians and patients is beneficial for strengthening patients' trust in their physicians and improving the efficiency of treatment. Moreover, some medical professionals (eg, nurses) can actively discuss with patients the information obtained by patients through OHCs, thus possibly promoting patients' additional use. Furthermore, the popularity of OHCs can be improved to increase the possibility of patients seeking health information through OHCs.

Limitations

This study has several limitations and future directions. First, other mediators can be examined in future studies in addition to physician-patient communication, internet health information–seeking behavior, and perceived quality of internet health information. Second, the sample was a fairly specific one, and the results therefore cannot be generalized to a general or global context. To make these results more universal, further surveys in other countries could be conducted by comparing the similarities and differences between China and these countries. Third, our sampling did not capture the characteristics of the Chinese population because of the limitation of the research conditions, and the inclusion of Chinese census data can be considered in future studies. Fourth, in the research model, the 2 mediators, eHealth literacy and health information–seeking behaviour, may overlap. eHealth literacy includes the skill of seeking health information through OHCs, and eHealth literacy and health information–seeking behavior in OHCs may be related to each other. Fifth, the observed data were self-reported data, and we did not track the participants' use of the internet. Sixth, we conducted the cross-sectional questionnaire investigation only once, so we failed to capture the dynamic changes in participants' attitudes toward variables. Finally, the adherence measure we used has no recorded previous use. These issues can be addressed in future studies.

Conclusions

This study aimed to identify the association between patients' eHealth literacy in OHCs and patient adherence. The results revealed that eHealth literacy in OHCs is positively associated with patient adherence through the mediations of physician-patient communication, patient internet health information–seeking behavior, and the perceived quality of internet health information in OHCs. All 3 mediators are crucial for improving patient adherence. These findings suggest the following: physicians should encourage patients to seek health information and communicate with them through OHCs, understand the level of patients' eHealth literacy and their perception of information quality, and help patients during the treatment to compensate for the lack of eHealth literacy; OHC operators should strengthen the management of information quality, make OHCs user friendly by providing detailed guidelines, and increase their popularity; OHC operators should improve the reliability of high-quality information provided to patients to decrease the gap between the actual and the perceived quality of health information.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant #62173025).

Authors' Contributions

XL conceived and designed the study, developed the research model, designed the questionnaire, conducted data collection and analysis, and drafted and modified the manuscript. RT drafted and modified the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Measurement instruments.
[DOC File , 122 KB - jmir_v23i9e14908_app1.doc]

References

1. Ratzan SC, Parker RM. Introduction. In: Selden CR, Zorn M, Ratzan SC, Parker RM, editors. Current Bibliographies in Medicine: Health Literacy. Bethesda,Md: National Library of Medicine; 2000:v-vii.
2. Ritterband LM, Borowitz S, Cox DJ, Kovatchev B, Walker LS, Lucas V, et al. Using the internet to provide information prescriptions. Pediatrics 2005 Nov;116(5):e643-e647. [doi: 10.1542/peds.2005-0404] [Medline: 16263978]
3. Levin-Zamir D, Bertschi I. Media health literacy, eHealth literacy, and the role of the social environment in context. Int J Environ Res Public Health 2018 Aug;15(8):1643 [FREE Full text] [doi: 10.3390/ijerph15081643] [Medline: 30081465]
4. Norman CD, Skinner HA. eHealth Literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
5. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]

6. Gilstad H. Toward a comprehensive model of eHealth literacy. In: Proceedings of the 2nd European Workshop on Practical Aspects of Health Information. 2014 May 19 Presented at: 2nd European Workshop on Practical Aspects of Health Information; 2014 May 19; Trondheim, Norway p. 19-20 URL: http://europepmc.org/abstract/MED/28001489
7. Paige SR, Krieger JL, Stellefson ML. The influence of eHealth literacy on perceived trust in online health communication channels and sources. J Health Commun 2017 Jan;22(1):53-65 [FREE Full text] [doi: 10.1080/10810730.2016.1250846] [Medline: 28001489]
8. Yan L, Tan Y. Feeling blue? Go online: An empirical study of social support among patients. Inform Syst Res 2014 Dec;25(4):690-709. [doi: 10.1287/isre.2014.0538]
9. Yang H, Guo X, Wu T. Exploring the influence of the online physician service delivery process on patient satisfaction. Decis Support Syst 2015 Oct;78:113-121. [doi: 10.1016/j.dss.2015.05.006]
10. Wu H, Lu N. Online written consultation, telephone consultation and offline appointment: an examination of the channel effect in online health communities. Int J Med Inform 2017 Nov;107:107-119. [doi: 10.1016/j.ijmedinf.2017.08.009] [Medline: 29029686]
11. Chang F, Chiu C, Chen P, Miao N, Lee C, Chiang J, et al. Relationship between parental and adolescent eHealth literacy and online health information seeking in Taiwan. Cyberpsychol Behav Soc Netw 2015 Oct;18(10):618-624. [doi: 10.1089/cyber.2015.0110] [Medline: 26375050]
12. Schulz PJ, Fitzpatrick MA, Hess A, Sudbury-Riley L, Hartung U. Effects of eHealth literacy on general practitioner consultations: a mediation analysis. J Med Internet Res 2017 May 16;19(5):e166 [FREE Full text] [doi: 10.2196/jmir.6317] [Medline: 28512081]
13. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. J Med Internet Res 2014 Apr;16(4):e104 [FREE Full text] [doi: 10.2196/jmir.2968] [Medline: 24718852]
14. Zhang R, Lu X, Wu W, Shang X. Why do patients follow physicians' advice? The influence of patients' regulatory focus on adherence: an empirical study in China. BMC Health Serv Res 2019 May 10;19(1):301 [FREE Full text] [doi: 10.1186/s12913-019-4127-9] [Medline: 31077196]
15. Lu X, Zhang R. Impact of physician-patient communication in online health communities on patient compliance: cross-sectional questionnaire study. J Med Internet Res 2019 May 13;21(5):e12891 [FREE Full text] [doi: 10.2196/12891] [Medline: 31094342]
16. Haynes RB, Taylor DW, Sackett DL, editors. Compliance in Health Care. Baltimore, Maryland: Johns Hopkins University Press; 1979.
17. Tustin N. The role of patient satisfaction in online health information seeking. J Health Commun 2010 Jan;15(1):3-17. [doi: 10.1080/10810730903465491] [Medline: 20390974]
18. Fernandez DM, Larson JL, Zikmund-Fisher BJ. Associations between health literacy and preventive health behaviors among older adults: findings from the health and retirement study. BMC Public Health 2016 Dec 19;16(1):596 [FREE Full text] [doi: 10.1186/s12889-016-3267-7] [Medline: 27430477]
19. Keim-Malpass J, Doede A, Kennedy C, Showalter SL. Health literacy assessment: feasibility in a breast surgical oncology clinic . Clin J Oncol Nurs 2017 Jun 01;21(3):384-386. [doi: 10.1188/17.CJON.384-386] [Medline: 28524902]
20. Horwitz RI, Horwitz SM. Adherence to treatment and health outcomes. Arch Intern Med 1993 Aug;153(16):1863-1868. [Medline: 8250647]
21. Hensher DA, Rose JM, Greene WH. Applied Choice Analysis. 2nd edition. Cambridge: Cambridge University Press; 2015.
22. Varleta P, Acevedo M, Akel C, Salinas C, Navarrete C, García A, et al. Mobile phone text messaging improves antihypertensive drug adherence in the community. J Clin Hypertens (Greenwich) 2017 Dec;19(12):1276-1284 [FREE Full text] [doi: 10.1111/jch.13098] [Medline: 28941056]
23. Brown MT, Bussell J, Dutta S, Davis K, Strong S, Mathew S. Medication adherence: truth and consequences. Am J Med Sci 2016 Apr;351(4):387-399. [doi: 10.1016/j.amjms.2016.01.010] [Medline: 27079345]
24. Umaki TM, Umaki MR, Cobb CM. The psychology of patient compliance: a focused review of the literature. J Periodontol 2012 Apr;83(4):395-400. [doi: 10.1902/jop.2011.110344] [Medline: 21819248]
25. Lu Y, Wu Y, Liu J, Li J, Zhang P. Understanding health care social media use from different stakeholder perspectives: a content analysis of an online health community. J Med Internet Res 2017 Apr 07;19(4):e109 [FREE Full text] [doi: 10.2196/jmir.7087] [Medline: 28389418]
26. Nambisan P. Information seeking and social support in online health communities: impact on patients' perceived empathy. J Am Med Inform Assoc 2011 May 1;18(3):298-304 [FREE Full text] [doi: 10.1136/amiajnl-2010-000058] [Medline: 21486888]
27. Yan Z, Wang T, Chen Y, Zhang H. Knowledge sharing in online health communities: A social exchange theory perspective. Inform Manage 2016 Jul;53(5):643-653. [doi: 10.1016/j.im.2016.02.001]
28. Wu B. Patient continued use of online health care communities: web mining of patient-doctor communication. J Med Internet Res 2018 Apr 16;20(4):e126 [FREE Full text] [doi: 10.2196/jmir.9127] [Medline: 29661747]
29. Huang L, Hsu C, Ko H, Kuo F. Influential factors on physicians' behaviour towards the medical web messaging board: a qualitative exploration. Int J Electron Healthc 2007 Feb;3(2):220-231. [doi: 10.1504/IJEH.2007.013102] [Medline: 18048271]

30. Karnoe A, Furstrand D, Christensen KB, Norgaard O, Kayser L. Assessing competencies needed to engage with digital health services: Development of the eHealth literacy assessment toolkit. J Med Internet Res 2018 May;20(5):e178 [FREE Full text] [doi: 10.2196/jmir.8347] [Medline: 29748163]
31. Paige SR, Krieger JL, Stellefson ML. The influence of eHealth literacy on perceived trust in online health communication channels and sources. J Health Commun 2017 Jan;22(1):53-65. [doi: 10.1080/10810730.2016.1250846] [Medline: 28001489]
32. Mitsutake S, Shibata A, Ishii K, Oka K. Associations of eHealth literacy with health behavior among adult internet users. J Med Internet Res 2016 Jul;18(7):e192 [FREE Full text] [doi: 10.2196/jmir.5413] [Medline: 27432783]
33. Hayat TZ, Brainin E, Neter E. With some help from my network: Supplementing eHealth literacy with social ties. J Med Internet Res 2017 Mar 30;19(3):e98 [FREE Full text] [doi: 10.2196/jmir.6472] [Medline: 28360024]
34. Park H, Cormier E, Glenna G. Health consumers eHealth literacy to decrease disparities in accessing eHealth information. Stud Health Technol Inform 2016;225:895-896. [Medline: 27332397]
35. Luo YF, Yang SC, Chen A, Chiang C. Associations of eHealth literacy with health services utilization among college students: cross-sectional study. J Med Internet Res 2018 Oct 25;20(10):e283 [FREE Full text] [doi: 10.2196/jmir.8897] [Medline: 30361201]
36. Freeman JL, Caldwell PHY, Bennett PA, Scott KM. How adolescents search for and appraise online health information: a systematic review. J Pediatr 2018 Apr;195:244-255.e1. [doi: 10.1016/j.jpeds.2017.11.031] [Medline: 29398062]
37. Bodie GD, Dutta MJ. Understanding health literacy for strategic health marketing: eHealth literacy, health disparities, and the digital divide. Health Mark Q 2008 Jan;25(1-2):175-203. [doi: 10.1080/07359680802126301] [Medline: 18935884]
38. Dobransky K, Hargittai E. Inquiring minds acquiring wellness: uses of online and offline sources for health information. Health Commun 2012 Sep;27(4):331-343. [doi: 10.1080/10410236.2011.585451] [Medline: 21932982]
39. Suri VR, Majid S, Chang Y, Foo S. Assessing the influence of health literacy on health information behaviors: A multi-domain skills-based approach. Patient Educ Couns 2016 Dec;99(6):1038-1045. [doi: 10.1016/j.pec.2015.12.017] [Medline: 26794667]
40. Zhang Y, Sun Y, Kim Y. The influence of individual differences on consumer's selection of online sources for health information. Comput Hum Behav 2017 Feb;67:303-312. [doi: 10.1016/j.chb.2016.11.008]
41. Li F, Li M, Guan P, Ma S, Cui L. Mapping publication trends and identifying hot spots of research on Internet health information seeking behavior: a quantitative and co-word biclustering analysis. J Med Internet Res 2015 Mar 25;17(3):e81 [FREE Full text] [doi: 10.2196/jmir.3326] [Medline: 25830358]
42. Wilson DT. Human Information Behavior. Informing Sci J 2000;3(2):49-56. [doi: 10.28945/576]
43. Neter E, Brainin E, Baron-Epel O. The third digital divide in the health domain: Is Internet use for health purposes associated with health benefits? In: eHealth: Current Evidence, Promises, Perils and Future Directions. Bingley: Emerald Publishing Limited; Jun 14, 2018:153-175.
44. Zhu J. Antecedents, covariates, and outcomes of media credibility in China. Mass Commun Rev 1997;24:4-17.
45. Jiang S, Street RL. Pathway linking internet health information seeking to better health: a moderated mediation study. Health Commun 2017 Dec;32(8):1024-1031. [doi: 10.1080/10410236.2016.1196514] [Medline: 27464036]
46. Maloney EK, D'Agostino TA, Heerdt A, Dickler M, Li Y, Ostroff JS, et al. Sources and types of online information that breast cancer patients read and discuss with their doctors. Palliat Support Care 2015 Apr;13(2):107-114 [FREE Full text] [doi: 10.1017/S1478951513000862] [Medline: 24182945]
47. Laugesen J, Hassanein K, Yuan Y. The impact of internet health information on patient compliance: a research model and an empirical study. J Med Internet Res 2015;17(6):e143 [FREE Full text] [doi: 10.2196/jmir.4333] [Medline: 26068214]
48. Jiang S, Street RL. Factors influencing communication with doctors via the internet: a cross-sectional analysis of 2014 HINTS survey. Health Commun 2017 Feb;32(2):180-188. [doi: 10.1080/10410236.2015.1110867] [Medline: 27196037]
49. Adler KG. Web portals in primary care: an evaluation of patient readiness and willingness to pay for online services. J Med Internet Res 2006 Oct 26;8(4):e26 [FREE Full text] [doi: 10.2196/jmir.8.4.e26] [Medline: 17213045]
50. Salancik GR, Pfeffer J. A social information processing approach to job attitudes and task design. Adm Sci Q 1978 Jun;23(2):224-253. [Medline: 10307892]
51. Young GJ, Meterko MM, Mohr D, Shwartz M, Lin H. Congruence in the assessment of service quality between employees and customers: A study of a public health care delivery system. J Bus Res 2009 Nov;62(11):1127-1135. [doi: 10.1016/j.jbusres.2008.08.004]
52. Bandura A. Health promotion from the perspective of social cognitive theory. Psychol Health 1998 Jul;13(4):623-649. [doi: 10.1080/08870449808407422]
53. Myrick JG. The role of emotions and social cognitive variables in online health information seeking processes and effects. Comput Hum Behav 2017 Mar;68:422-433. [doi: 10.1016/j.chb.2016.11.071]
54. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
55. Liang C, Wang K, Hwang S, Lin K, Pan H. Factors affecting the physician-patient relationship of older veterans with inadequate health literacy: an observational study. Br J Gen Pract 2013 May;63(610):e354-e360 [FREE Full text] [doi: 10.3399/bjgp13X667222] [Medline: 23643234]
56. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, Lopez A, et al. The literacy divide: health literacy and the use of an internet-based patient portal in an integrated health system-results from the diabetes study of northern California

(DISTANCE). J Health Commun 2010 Aug;15(Suppl 2):183-196 [FREE Full text] [doi: 10.1080/10810730.2010.499988] [Medline: 20845203]
57. Matusitz J, Spear J. Effective doctor-patient communication: an updated examination. Soc Work Public Health 2014 Apr;29(3):252-266. [doi: 10.1080/19371918.2013.776416] [Medline: 24802220]
58. Orom H, Underwood W, Cheng Z, Homish DL, Scott I. Relationships as medicine: quality of the physician-patient relationship determines physician influence on treatment recommendation adherence. Health Serv Res 2018 Dec;53(1):580-596. [doi: 10.1111/1475-6773.12629] [Medline: 27981559]
59. Bultman DC, Svarstad BL. Effects of physician communication style on client medication beliefs and adherence with antidepressant treatment. Patient Educ Couns 2000 May;40(2):173-185. [Medline: 10771371]
60. Roberts L, Cornell C, Bostrom M, Goldsmith S, Ologhobo T, Roberts T, et al. Communication skills training for surgical residents: Learning to relate to the needs of older adults. J Surg Educ 2018 Sep;75(5):1180-1187. [doi: 10.1016/j.jsurg.2018.02.005] [Medline: 29609892]
61. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
62. Park H, Lee E. Self-reported eHealth literacy among undergraduate nursing students in South Korea: a pilot study. Nurse Educ Today 2015 Feb;35(2):408-413. [doi: 10.1016/j.nedt.2014.10.022] [Medline: 25466791]
63. Makoul G, Krupat E, Chang C. Measuring patient views of physician communication skills: development and testing of the Communication Assessment Tool. Patient Educ Couns 2007 Aug;67(3):333-342. [doi: 10.1016/j.pec.2007.05.005] [Medline: 17574367]
64. Khazaee-Pool M, Majlessi F, Montazeri A, Pashaei T, Gholami A, Ponnet K. Development and psychometric testing of a new instrument to measure factors influencing women's breast cancer prevention behaviors (ASSISTS). BMC Womens Health 2016 Jul;16(1):40 [FREE Full text] [doi: 10.1186/s12905-016-0318-2] [Medline: 27444675]
65. Bu X, You L, Li Y, Liu K, Zheng J, Yan T, et al. Psychometric properties of the Kessler 10 Scale in Chinese parents of children with cancer. Cancer Nurs 2017 Jul;40(4):297-304. [doi: 10.1097/NCC.0000000000000392] [Medline: 27257800]
66. Xiao Y, Li T, Xiao L, Wang S, Wang S, Wang H, et al. The Chinese version of instrument of professional attitude for student nurses (IPASN): Assessment of reliability and validity. Nurse Educ Today 2017 Feb;49:79-83. [doi: 10.1016/j.nedt.2016.11.013] [Medline: 27889582]
67. Albertsen L, Wiedmann K, Schmidt S. The impact of innovation-related perception on consumer acceptance of food innovations – Development of an integrated framework of the consumer acceptance process. Food Qual Prefer 2020 Sep;84:103958. [doi: 10.1016/j.foodqual.2020.103958]
68. Hair JF, Ringle CM, Sarstedt M. Partial least squares structural equation modeling: Rigorous applications, better results and higher acceptance. Long Range Plann 2013 Feb;46(1-2):1-12. [doi: 10.1016/j.lrp.2013.01.001]
69. Bari MW, Mahmood F, Qurrah-tul-ain, Bashir M, Usman M. The role of instrumental guanxi in the relation between entrepreneurs' social competence and firms' financial performance: A comparative study. Economic Research-Ekonomska Istraživanja 2020 Aug 28;34(1):1-23. [doi: 10.1080/1331677x.2020.1782244]
70. Hussain S, FangWei Z, Ali Z. Examining influence of construction projects' quality factors on client satisfaction using partial least squares structural equation modeling. J Constr Eng Manage 2019 May;145(5):05019006. [doi: 10.1061/(asce)co.1943-7862.0001655]
71. Audrain-Pontevia A, Menvielle L. Do online health communities enhance patient-physician relationship? An assessment of the impact of social support and patient empowerment. Health Serv Manage Res 2018 Aug;31(3):154-162. [doi: 10.1177/0951484817748462] [Medline: 29280679]
72. Huh J, Kwon BC, Kim S, Lee S, Choo J, Kim J, et al. Personas in online health communities. J Biomed Inform 2016 Oct;63:212-225. [doi: 10.1016/j.jbi.2016.08.019] [Medline: 27568913]
73. Buysse HEC, Coorevits P, Van Maele G, Hutse A, Kaufman J, Ruige J, et al. Introducing telemonitoring for diabetic patients: development of a telemonitoring 'Health Effect and Readiness' Questionnaire. Int J Med Inform 2010 Aug;79(8):576-584. [doi: 10.1016/j.ijmedinf.2010.05.005] [Medline: 20599161]
74. Biasutti M, Frate S. A validity and reliability study of the Attitudes toward Sustainable Development scale. Environ Educ Res 2017 Feb;23(2):214-230. [doi: 10.1080/13504622.2016.1146660]
75. Sobral MP, Costa ME, Schmidt L, Martins MV. COMPI Fertility Problem Stress Scales is a brief, valid and reliable tool for assessing stress in patients seeking treatment. Hum. Reprod 2017 Feb;32(2):375-382. [doi: 10.1093/humrep/dew315]
76. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York: Routledge; 1988.
77. Tenenhaus M, Vinzi VE, Chatelin Y, Lauro C. PLS path modeling. Comput Stat Data An 2005 Jan;48(1):159-205. [doi: 10.1016/j.csda.2004.03.005]
78. Wetzels M, Odekerken-Schröder G, van Oppen C. Using PLS path modeling for assessing hierarchical construct models: Guidelines and empirical illustration. MIS Quart 2009 Mar;33(1):177-195. [doi: 10.2307/20650284]
79. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 1986 Dec;51(6):1173-1182. [Medline: 3806354]
80. Wen Z, Ye B. Analyses of mediating effects: the development of methods and models. Advances in Psychological Science 2014 Jan;22(5):731-745. [doi: 10.3724/sp.j.1042.2014.00731]

81. Silver MP. Patient perspectives on online health information and communication with doctors: a qualitative study of patients 50 years old and over. J Med Internet Res 2015 Jan;17(1):e19 [FREE Full text] [doi: 10.2196/jmir.3588] [Medline: 25586865]
82. Xesfingi S, Vozikis A. eHealth literacy: In the quest of the contributing factors. Interact J Med Res 2016 May;5(2):e16 [FREE Full text] [doi: 10.2196/ijmr.4749] [Medline: 27226146]

Abbreviations

AVE: average variance extracted
GoF: goodness of fit
OHC: online health community
PLS: partial least squares
SEM: structural equation modelling

Edited by G Eysenbach, L Buis; submitted 02.06.19; peer-reviewed by S Paige, E Neter, D Nault; comments to author 15.05.20; revised version received 31.10.20; accepted 29.01.21; published 13.09.21.

Please cite as:
Lu X, Zhang R
Association Between eHealth Literacy in Online Health Communities and Patient Adherence: Cross-sectional Questionnaire Study
J Med Internet Res 2021;23(9):e14908
URL: https://www.jmir.org/2021/9/e14908
doi:10.2196/14908
PMID:34515638

Original Paper

Measuring a Broad Spectrum of eHealth Skills in the Web 3.0 Context Using an eHealth Literacy Scale: Development and Validation Study

Hua-Xuan Liu[1*], MA; Bik-Chu Chow[1*], PhD, Prof Dr; Wei Liang[1,2], PhD; Holger Hassel[3], PhD, Prof Dr; YaJun Wendy Huang[1], PhD

[1]Department of Sport, Physical Education and Health, Hong Kong Baptist University, Hong Kong, China (Hong Kong)

[2]Center for Health and Exercise Science Research, Hong Kong Baptist University, Hong Kong, China (Hong Kong)

[3]Institute of Applied Health Sciences, Coburg University of Applied Science and Arts, Coburg, Germany

[*]these authors contributed equally

Corresponding Author:
Hua-Xuan Liu, MA
Department of Sport, Physical Education and Health
Hong Kong Baptist University
Room 926, 9/F, Academic and Administration Building
Kowloon
Hong Kong
China (Hong Kong)
Phone: 852 1 305 157 5809
Email: xuan0402@life.hkbu.edu.hk

Abstract

Background: eHealth literacy (EHL) refers to a variety of capabilities that enable individuals to obtain health information from electronic resources and apply it to solve health problems. With the digitization of health care and the wide availability of health apps, a more diverse range of eHealth skills is required to properly use such health facilities. Existing EHL measurements focus mainly on the skill of obtaining health information (Web 1.0), whereas skills for web-based interactions (Web 2.0) and self-managing health data and applying information (Web 3.0) have not been well measured.

Objective: This study aims to develop an EHL scale (eHLS) termed eHLS-Web3.0 comprising a comprehensive spectrum of Web 1.0, 2.0, and 3.0 skills to measure EHL, and evaluate its validity and reliability along with the measurement invariance among college students.

Methods: In study 1, 421 Chinese college students (mean age 20.5, SD 1.4 years; 51.8% female) and 8 health experts (mean age 38.3, SD 5.9 years; 87.5% female) were involved to develop the eHLS-Web3.0. The scale development included three steps: item pool generation, content validation, and exploratory factor analysis. In study 2, 741 college students (mean age 21.3, SD 1.4 years; 52.2% female) were recruited from 4 Chinese cities to validate the newly developed eHLS-Web3.0. The construct validity, convergent validity, concurrent validity, internal consistency reliability, test-retest reliability, and measurement invariance across genders, majors, and regions were examined by a series of statistical analyses, including confirmatory factor analysis (CFA) and multigroup CFAs using SPSS and Mplus software packages.

Results: Based on the item pool of 374 statements collected during the conceptual development, 24 items (4-10 items per subscale) were generated and adjusted after cognitive testing and content validity examination. Through exploratory factor analysis, a 3-factor eHLS-Web3.0 was finally developed, and it included acquisition (8 items), verification (6 items), and application (10 items). In study 2, CFAs supported the construct validity of the 24-item 3D eHLS-Web3.0 (χ^2_{244}=903.076, χ^2_{244}=3.701, comparative fit index=0.924, Tucker-Lewis index=0.914, root mean square error of approximation [RMSEA]=0.06, and standardized root mean residual [SRMR]=0.051). The average variance extracted (AVE) value of 0.58 and high correlation between eHLS-Web3.0 subscales and the eHealth Literacy Scale (r=0.725-0.880, P<.001) indicated the convergent validity and concurrent validity of the eHLS-Web3.0. The results also indicated satisfactory internal consistency reliability (α=.976, ρ=0.934-0.956) and test-retest reliability (r=0.858, P<.001) of the scale. Multigroup CFA demonstrated the 24-item eHLS-Web3.0 to be invariant at

all configural, metric, strength, and structural levels across genders (female and male), majors (sport-related, medical, and general), and regions (Yinchuan, Kunming, Xiamen, and Beijing).

Conclusions: The 24-item 3D eHLS-Web3.0 proved to be a reliable and valid measurement tool for EHL in the Web 3.0 context among Chinese college students.

(J Med Internet Res 2021;23(9):e31627) doi:10.2196/31627

KEYWORDS

eHealth literacy; scale development; validation; college students

Introduction

With the digitization of health care and the wide availability of health applications, personal eHealth has become important for health management. The World Health Organization defines eHealth as "the use of information and communication technology for health" [1], and researchers have called for a broader understanding of the capabilities and skills required for individuals to use and benefit from eHealth services [2]. In 2006, Norman and Skinner labeled this ability as "eHealth literacy" (EHL) [3].

The definition of EHL was grounded in the health promotion theory, which referred to EHL as the "the ability to seek, find, understand, and appraise health information from electronic resources and apply that knowledge to solve a health problem or make a health-related decision" [3]. Based on this concept, an 8-item measurement tool named the eHealth Literacy Scale (eHEALS) was developed. The eHEALS is the most well-known and widely accepted tool for measuring EHL. Till 2015, there have been 45 studies that used the eHEALS, and it is the only tool used to measure EHL in more than 1 study [4].

However, with advances in technology, some people identified the need for updating the content of EHL to better fit it with the new internet environment. They also doubted on how accurately the eHEALS measured the use of new technologies to find and evaluate health information [5]. In 2011, the first critique of eHEALS was made, finding a weak correlation between the scale and eHealth behaviors beyond web-based information-researching skills, suggesting the need to revise the tool [6]. These observations were reasonable. According to the most widely accepted generation division of internet evolution, the current internet environment has gone through three stages. The first one, Web 1.0, refers to the read-only web, whereas Web 2.0 refers to a read-write mode, providing a "participative social web" with greater collaboration and interactivity between consumers, programmers, service providers, and organizations. Web 3.0 is the current environment, the so-called "semantic web," accessible in a "read-write-execute" mode, providing digital, personal, and intelligent services [7-9]. The eHEALS was developed 15 years ago for measuring the abilities related to page views; it is necessary to enrich and update it for better scaling of current eHealth usage [4,5].

To fill this gap, other measures of EHL have been developed. For example, van der Vaart et al [6] designed 9 assignments to test the "actual performance of eHealth literacy." However, those assignments were more for testing health-related internet skills rather than actual EHL. Moreover, the practical test approach would limit the number of participants. In 2017, another tool called the Digital Health Literacy Instrument was developed [10], which included 21 self-assessed items supplemented with 7 performance-related tasks that focused on handling digital information. They were mainly related to navigating the internet and messaging health professionals. The assignments developed by van der Vaart et al [6,10] were not the only ones to assess individuals' EHL. The Readiness Self-Assessment-Health Professions version [11] also measured participants' perceived ability and their actual ability to obtain and evaluate eHealth information. It is worthy to mention that the developers of the Readiness Self-Assessment-Health Professions version believed that the readiness to use internet-based resources should be seen as a component of EHL. The Patient Readiness to Engage in Health Internet Technology [12] measures readiness as well. It was developed based upon the groundwork of eHEALS but expanded the understanding of EHL to include two meta-factors: facilitators and barriers. Beyond task performance and readiness evaluations, several other measurements focused on evaluating people's Web 2.0 skills, specifically their literacy on web-based social activities. One of them was the 19-item electronic Health Literacy Scale (eHLS) [13]. It examined the behavioral, communicational, and attitudinal components of health literacy among eHealth information seekers, which expanded the conceptualization beyond the traditional document-based measures to include interactive and communicative aspects of literacy. Another scale for Web 2.0 is the eHealth Literacy Scale (eHEALS) [14], suggesting that the underlying social structure affects an individual's health status, computer literacy, intrinsic interest in health, and perceived ability to use the internet for health purposes. In addition, there were two other measurements that provided a different understanding of EHL. The eHealth Literacy Questionnaire (eHLQ) [15] comprehending EHL under a user-need perspective characterizing eHealth users paid attention to eHealth users' understanding, attitudes, and motives. The eHealth Literacy Assessment Toolkit (eHLA) [16] was a mix of existing and newly developed scales that viewed EHL as a combination of health literacy, computer and digital literacy, and information literacy. A recent study translated the eHEALS [3] into Chinese and examined its psychometric properties among the Chinese population [17]. However, the referenced study has not measured individuals' EHL in the context of Web 2.0 and 3.0, and the stability (eg, test-retest reliability and measurement invariance) of the EHL scale has not been examined.

Although previous studies have provided new ways to assess EHL with some of them scaling web communication abilities, all the tools mentioned above have not yet measured the skills of Web 3.0 technologies. Insufficient knowledge existed in the newly required competencies of EHL (eg, capacity for using mobile services, sense of information safety, and abilities of screening, communicating, and sharing). Furthermore, except for eHEALS, these scales have not been used by other researchers, which means there is still no well-developed scale for evaluating individuals' EHL worldwide.

The COVID-19 pandemic led to a decrease in interpersonal social activities, but increased individuals' EHL [18,19]. According to the annual report provided by a famous clinical virtual community in China publishing health care–related and clinical insights and findings [19], the visitor volume increased by 4.6 times compared to what it was earlier. Predictably, increased use of health information technology would change people's health care styles, shifting partially from offline to internet-based services. In such an environment, it is essential to develop a new instrument and provide a valid tool to measure current EHL levels. As college students are the major internet users and frequently turn to it for eHealth information [20], the current research aimed to develop a scale to measure EHL in the Web 3.0 context (eHLS-Web3.0) and examine its reliability and validity among Chinese college students. Accordingly, two studies were implemented; the first study aimed to develop an eHealth literacy scale (eHLS-Web3.0), including three steps: item pool generation, content validation, and exploratory factor analysis (EFA); the second study aimed to examine the validity (construct, convergent, and concurrent), reliability (internal consistency and test-retest reliability) and measurement invariance across genders (female and male), majors (sport-related, medical, and general), and regions (Yinchuan, Kunming, Xiamen, and Beijing) in Chinese college students. The current research was approved by the Research Ethics Committee of Hong Kong Baptist University.

Methods

Study 1

Participants

Step 1: Item Pool Generation

This step involved 28 college students, including 18 interviewees for item pool generation and 10 students for readability and cultural sensitivity examination. Snowball sampling [21] was applied in 2019 (before the COVID-19 pandemic) to recruit interviewees from three different Chinese cities (Beijing, Wuhan, and Putian). These three cities were selected based on the cultural geographic (north, middle, and south) and economic statuses (high, medium, and low) of these Chinese cities [22-24].

All eligible participants met the following inclusion criteria: (1) experience in using eHealth websites and tools, (2) having sufficient Chinese language skills, and (3) willingly consent to participate in the interview. Regarding the sample size, although there is no definite criterion for qualitative research, the number of participants should be in accordance with the creed of theoretical saturation, which means no new or relevant data emerge [25]. To achieve theoretical saturation, based on the "rule of thumb," the suggested number of participants for interview studies is approximately 12 to 15 [26]. In the current study, 18 Chinese college students participated in the semistructured interviews via telephone; they included 3 males and 3 females from sports, medical, and ordinary nonhealth-related majors. All the telephonic interviews were conducted in Chinese (Mandarin) language and recorded digitally with the interviewees' consent. An initial item pool was built based on the interview data and the existing literature. Next, 10 college students were randomly recruited to examine accessibility (eg, readability) and cultural sensitivity of those candidate items, and the participants' ability to complete the self-administered survey. An informed consent form was delivered to each student before data collection.

Step 2: Content Validation

Based on the guideline proposed by Lynn [27], 8 Chinese health experts were invited to validate the content of the selected items [27]. The selection of the health experts was based on the following criteria: (1) published at least one paper related to health literacy or eHealth, (2) have sufficient Chinese language proficiency, and (3) confirmed their willingness and consent for participating in the interview.

Step 3: EFA Process

The participants involved in this step constituted an independent sample with the recruitment criteria similar to those in the interviewees of step 1, except that the target cities for participant recruitment were changed. This is because step 3 was performed after the outbreak of COVID-19. The previous target city Wuhan was avoided to enhance the generalizability of the new scale. In step 3, four different representative Chinese cities were chosen, including Beijing, Xiamen, Kunming, and Yinchuan. These four cities were selected based on their geographic locations (north, southeast, southwest, and northwest), political status (country capital, provincial capital, and prefecture-level city), and economic status (high, medium, and low) [22-24]. Furthermore, these four cities were chosen considering the issues of "convenience and feasibility" [28].

The questionnaire was distributed via a survey distribution website, starting with prior informed consent, followed by the main content of the survey. The participation was voluntary, and all participants were allowed to withdraw from the study at any time. University lecturers from the target cities helped distribute the link to the questionnaire in the classes. Based on the suggestion of Bryman [25,29,30], at least 240 observations were required for EFA (1:10 item-to-response ratio) in the current study. Finally, 393 students responded to the questionnaire and provided information regarding their age, gender, major, and experience of using health-related electronic tools or websites, which led to an adequate sample size. It took approximately 5 to 10 minutes for the participants to complete the questionnaire. The characteristics of the participants in study 1 are presented in Table 1.

Table 1. Characteristics of participants in study 1.

Demographic information	Values
Interviewees (n=18)	
Age (years), mean (SD)	22.1 (1.02)
Age range	21-25
Gender (female), n (%)	9 (50)
Readability examiners (n=10)	
Age (years), mean (SD)	20.3 (0.95)
Age range	19-22
Gender (female), n (%)	7 (70)
Experts for content validation (n=8)	
Age (years), mean (SD)	38.3 (5.92)
Age range	30-46
Gender (female), n (%)	7 (87.5)
Sample for EFA[a] (n=393)	
Age (years), mean (SD)	20.5 (1.36)
Age range	18-25
Gender (female), n (%)	202 (51.4)

[a]EFA: exploratory factor analysis.

Procedure

Step 1: Item Pool Generation

There were three types of sources for forming the item pool. The first comprised well-established EHL instruments (eg, eHEALS, eHLQ, and e-HLS). The feasible items were selected and adapted to form the initial item pool. Second, items from well-known health or computer literacy scales were deliberately selected because they were measuring important elements of EHL. Third, items were generated from the previous interview results. For the first and second types of sources, some items were developed by western scholars. Those items were translated by two PhD students specializing in health promotion, and then back-translated by a senior Chinese English teacher unfamiliar with eHealth-literacy–related indices. Subsequently, 10 college students helped assess the comprehensibility, clarity, and length of the items via a dichotomous scale (applicable vs inapplicable). The inapplicable items were deleted or refined.

Step 2: Content Validation

The generated item pool was sent to the panel of experts. The experts assessed the relevance of each item to the understanding of EHL, using a 4-point Likert scale ranging from 1 (irrelevant) to 4 (extremely relevant). They were asked to provide ratings and suggestions for alternative items. The ratings were used to calculate the content validity index [27] of each item (I-CVI). The I-CVI was calculated by dividing the number of judges providing a rating above 3 by the total number of judges. The acceptable I-CVI score had to be above 0.83 [27,31]. The invalid items were either eliminated or revised based on the experts' suggestions.

Step 3: EFA Process

The items generated in step 2 were placed in the format of a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). The distribution pattern of the items was examined. The ones with non-normal distributions were eliminated. EFA was performed for the remaining items. Principal component factoring with direct oblimin rotation was performed. Factor extraction was based on an eigenvalue higher than 1 and a confirmatory inspection of the scree plot. Items with primary factor loadings of 0.4 and above were considered interpretable [32,33]. Items with cross-loading were deleted, which refers to those having a second highest factor loading of 0.3 or higher or having a small gap (less than 0.2) between the primary and secondary loadings [32,33]. Bivariate correlations between the subscales were also tested.

Study 2

Participants

The participants were independent samples with the recruitment criteria similar to those used in the EFA step. Considering the item-to-response ratios of at least 1:10 [29] and the recommendation for a minimum sample size of 200 in confirmatory factor analysis (CFA) [34], 741 college students were invited to participate in the current study. After 1.5 months, a follow-up retest was distributed to the same group of people, and 306 of them responded. A summary of the CFA participants is presented in Table 2.

Table 2. Sample characteristics for confirmatory factor analysis (N=741).

Demographic information	Value
Age (years), mean (SD)	21.3 (1.39)
Age range	17-25
Gender (female), n (%)	387 (52.2)
Major	
Medical, n (%)	41 (5.5)
Sport, n (%)	128 (17.3)
Nonhealth-related, n (%)	572 (77.2)
Region	
Beijing, n (%)	124 (16.7)
Xiamen, n (%)	287 (38.7)
Kunming, n (%)	216 (29.1)
Yinchuan, n (%)	114 (15.4)

Measures

The eHLS-Web3.0 instrument was developed in study 1, consisting of 24 items classified based on 3 factors: acquisition (8 items), verification (6 items), and application (10 items). The answers were indicated on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The internal consistency (Cronbach α) for the total scale was .971 and those for the subscales were 0.913-0.962 in Chinese college students.

The eHEALS was initially developed by Norman and Skinner [3] as the first and most widely used tool for measuring EHL. It is an 8-item unidimensional scale whose validation was performed in a population of adolescents (aged 13-21 years) from 14 secondary schools in a large Canadian city. The Chinese version of eHEALS was tested in 110 senior high school students by Guo et al [35]. The internal consistency (Cronbach α) for the total scale was .913, and the factor loading coefficients were between 0.692 and 0.869 [35] in Chinese senior high school students.

Procedure

The procedures for data collection were identical to those of the EFA step. Participants were selected from four target cities, Beijing, Xiamen, Kunming, and Yinchuan, by snowball sampling. The eHLS-Web3.0 was distributed to participants who were independent of those participating in the EFA step via the internet. University lecturers from the target cities independent of those in the EFA helped distribute the link to the questionnaire in the classes they taught; they consciously avoided the classes that had participants in the EFA step. Informed consent forms were delivered to the participants before the survey. It took approximately 10 minutes for the participants to finish the questionnaire. A retest was delivered to them 1.5 months later.

Data Analysis

To test the construct validity of the current scale, the items were analyzed through CFA with structural equation modeling using Mplus (version 7) [36]. After testing the distribution of the items in the measurement model, the maximum likelihood robust estimator was employed [37]. The chi-square statistic (χ^2) was used to test the model's overall goodness of fit [38]. Multiple model fit indices were examined further, including the comparative fit index [39] and Tucker-Lewis index [40], with a cutoff value of approximately 0.9 and above indicating a satisfactory fit between the CFI and TLI [41]; the standardized root mean residual (SRMR) [39,41] with values near 0.08 indicates adequate model fit; the root mean square error of approximation (RMSEA) and its 90% CI [42] indicate good fit at values less than 0.08. The standardized factor loadings and standardized residuals were examined. Items with factor loadings below 0.4 and large standardized residuals (≥2) were removed. The concurrent validity was examined by calculating the bivariate correlations between the subscale of the eHLS-Web3.0 and eHEALS using SPSS (version 24.0, SPSS Inc). The convergent validity was tested by examining the average variance extracted (AVE), for which values above 0.5 indicated an acceptable measurement. Internal consistency reliability was examined by analyzing the Cronbach α and the composite reliability coefficient (CR), for which the CR values above 0.7 [43] were considered acceptable. The test-retest reliability was measured by examining the bivariate correlation between the score of eHLS-Web3.0 determined in the CFA stage and the follow-up stage after 1.5 months. A multiple-group covariance structure analysis approach was employed to examine whether the measurement was invariant across genders, majors, and regions [44]. A ΔCFI value smaller than or equal to 0.01 between incrementally constrained models should be accepted as evidence of adequate fit [45].

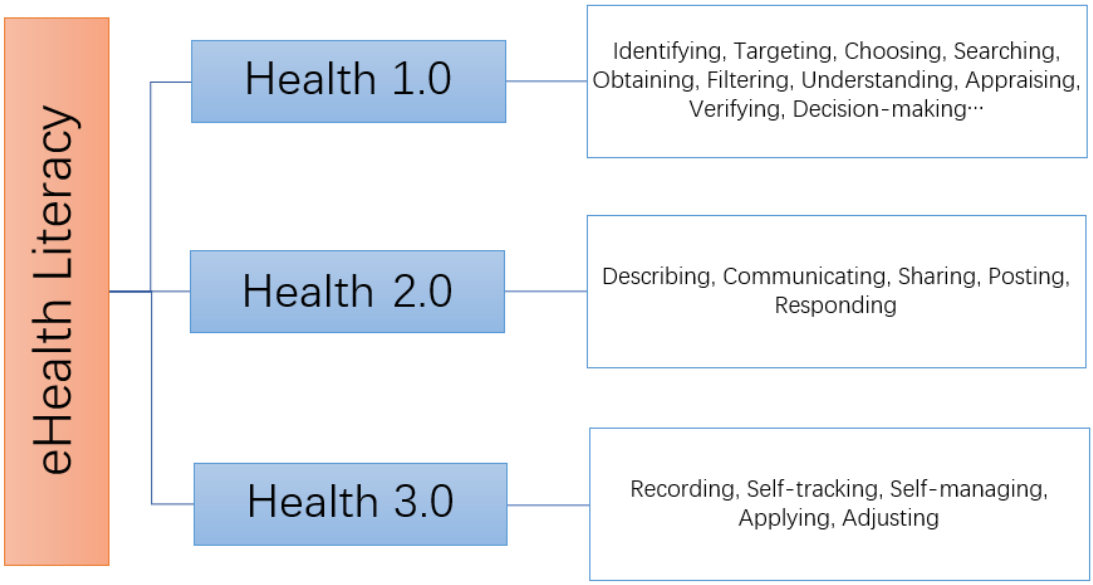
Results

Study 1

Step 1: Item Pool Generation

Previous interviews yielded 374 tags and a conceptual framework for EHL, as shown in Figure 1.

Figure 1. Conceptual framework of eHealth literacy.



The Health 1.0 part was consistent with existing research [3,4,6,10-12]; some Health 2.0 behaviors (describing, communicating, and sharing) [14,15] have also been mentioned by other researchers. In contrast, this study presents a first look at behaviors related to Web 3.0 and delves deeper into the social network services with respect to eHealth, finding that the application of information is currently much more complicated than before; other than researching on the internet or applying the information to make health-related decisions or solve problems, individuals could also create their own health data, use the information to communicate, respond (ie, giving suggestions and advice, and responding to help seekers), socialize with others, and share and post information (ie, forward helpful information or post their own health or fitness data).

Based on the abovementioned tags and existing well-established scales, 266 items were developed, taking the form of eHEALS (eg, I know how to find helpful health resources on the internet) in the Chinese language. Specifically, there were 114 items for Health 1.0 skills, 39 for Health 2.0 ones, 55 for Health 3.0 ones, 42 for readiness-related ones (mostly retrieved from earlier studies) [11,12], and 16 for health literacy ones (as per an earlier study) [10]. To ensure the relevance, readability, comprehensibility, and clarity of the scale to the targeted Chinese population, it was assessed by 10 college students who were conveniently selected. Suggestions were given by the participants to improve the comprehensibility of the preliminary pool. Among the items, 7 duplicated ones were discussed, 3 of them were deleted, and 4 of them were merged into 2 items. Next, 5 items were removed owing to unclear text. Further, 56 items were excluded because they were considered inapplicable for testing EHL by more than 3 students. After that, 200 candidate items were retained for the next step.

Step 2: Content Validation

Following the adjustments of the item pool, the candidate items were forwarded to 8 specialists in the health literacy and eHealth areas to review the content and assess its validity. Based on their quantitative ratings, the I-CVI was calculated for individual items. Items rated “1” or “2” indicating inapplicability according to two reviewers (I-CVI<0.83) were subsequently deleted. Additionally, suggestions regarding refinement of the scenario or item pool were elicited.

Based on the specialists’ quantitative feedback, 128 items were considered irrelevant and excluded. For the remaining 72 items, a discussion among specialists was arranged via a web-based group chat and videoconferencing for obtaining more detailed suggestions. Some mergers and adjustments were made to make the scale the clearest and the most easily understandable one. For instance, in the initial version, the abilities to choose proper forums and apps were measured separately following expert suggestions; these items were merged into one using the term “eHealth tools.” The questions on web-based celebrity and official accounts were also conflated. Some items testing attitudes for eHealth, especially the readiness-related ones, were considered not related to “ability” and eliminated. After the creation and refinement steps, the final number of items for field testing was 50.

Step 3: EFA Process

The 50 refined items were randomized and administered to 393 college students via the internet. Descriptive statistical analysis was performed with SPSS (Version 20) examining the distribution pattern of each item. Items 2, 10, 37, and 50 were non-normally distributed (their kurtosis ranges were beyond –1 and 1). For the 46 remaining items, EFA was performed. The results of the Kaiser-Meyer-Olkin **(**KMO) and Bartlett tests performed during EFA showed that the KMO statistic for these

items was 0.966, which is higher than 0.6, thus proving that the factor analysis model was appropriate and useful to analyze current data sets. Principal component factoring with direct oblimin rotation was used. Factor extraction was based on an eigenvalue higher than 1 and a confirmatory inspection of the scree plot shown in Figure 2. Items with primary factor loadings of 0.4 and above were considered interpretable. We excluded 5 items (#1, #11, #27, #28, and #36) with inadequate factor loadings were excluded. Subsequently, 13 other items were found to have a second highest factor loading of 0.3 or higher, and 4 other items (#22, #29, #33, and #41) had a small gap (less than 0.2) between the primary and secondary loadings. They were identified as items with cross-loading and were deleted as well. Therefore, 5 factors were extracted from initial factoring, and 3 strong factors were identified with the 24 items remaining, as shown in Table 3.

Figure 2. Confirmatory inspection of the scree plot.

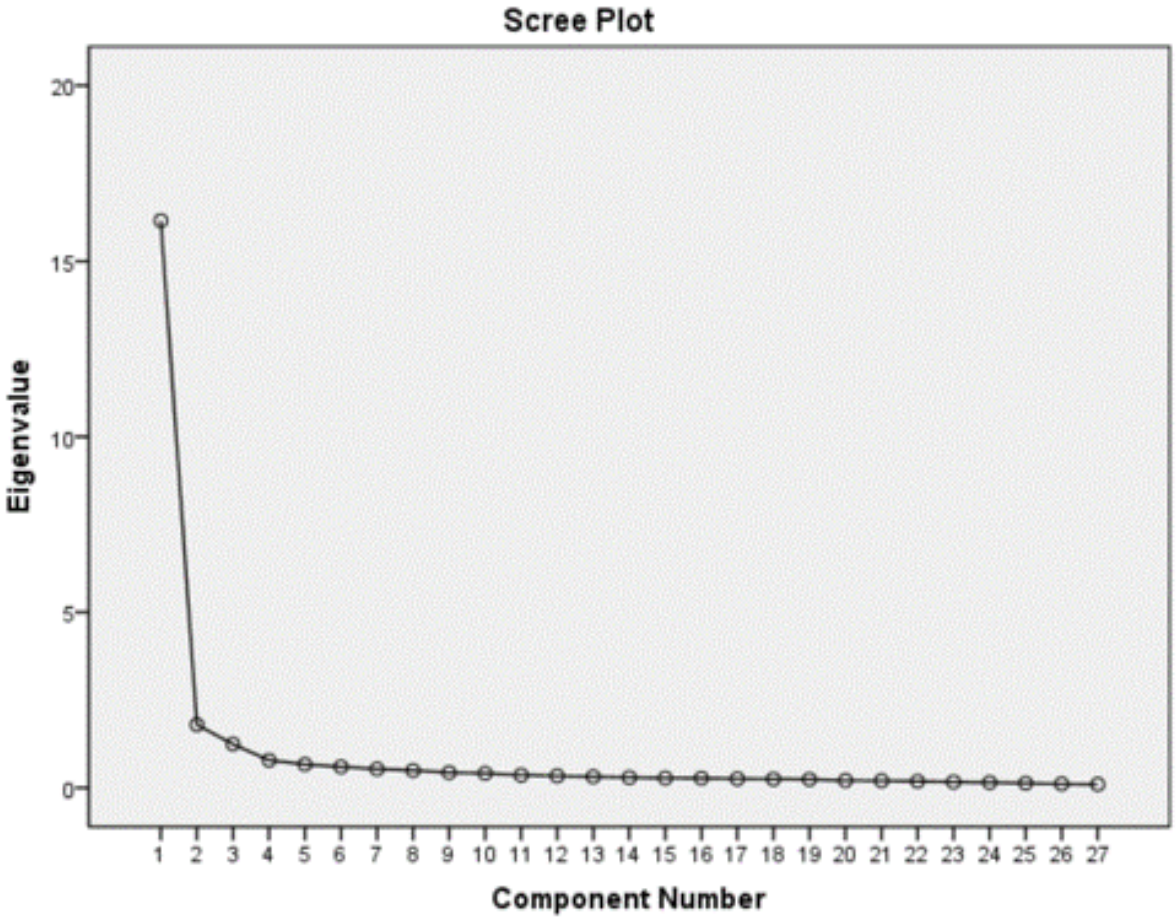


Table 3. Principal component analysis of the electronic Health Literacy Scale-Web3.0.

Item	Component		
	1	2	3
#3: I know which kind of eHealth tool I should choose to fit my health needs (ie, checking drug description, seeking health advice, or making weight loss plan).	__[a]	—	–0.817
#4: I can judge whether the eHealth tool is credible or not.	—	—	–0.659
#5: I will prefer to obtain the health information (ie, information about medical, sport or daily care) via offline channel.	—	—	0.880
#6: I know where to find useful health resources on the internet.	—	—	–0.424
#26: When communicating with others on the internet, I can articulate my health-related concerns clearly.	—	—	–0.530
#30: When replying to others' health-related help-seeking on the internet, I can provide responsible response (which means my answer can neither mislead others, nor)	—	—	–0.547
#31: I can judge whether the health information available on the internet has a commercial interest (ie, the person providing the information is for the sale of a product).	—	—	–0.557
#35: When using the eHealth tools, I will protect the originality of the information (ie, never plagiarize others' original content, report an offense to those infringing ones).	—	—	–0.550
#14: When searching the health information on the internet, I will check the credentials and affiliations of author.	—	–0.979	—
#15: When searching the health information on the internet, I will check who owns the website.	—	–0.938	—
#16: When searching the health information on the internet, I will check the date of its last update.	—	–0.864	—
#17: When searching the health information on the internet, I will check whether other print or web resources had confirmed this information.	—	–0.903	—
#18: I know how to verify the eHealth information from multiple sources.	—	–0.510	—
#20: Even if the health information I obtained is from someone I trust, I will still check it on the internet.	—	–0.444	—
#38: I know how to use the eHealth tools to record my health behaviors.	0.873	—	—
#39: I know how to make use of the records on the eHealth tools to provide reference for my daily health management.	0.852	—	—
#40: I know how to use the eHealth tools to track my health behaviors (ie, acquainting my exercise frequency or the change curve of body fat rate).	0.869	—	—
#42: I can continuously use a certain eHealth tool (ie, apps, intelligent body fat scale, fitness bracelet) for a long time.	0.919	—	—
#43: I can use the eHealth tools with a clear plan.	0.863	—	—
#44: I can adjust my frequency, strength, and usage pattern timely when using the eHealth tool according to the actual condition.	0.879	—	—
#45: I know how to use the eHealth tools to post and share my eHealth behaviors (ie, post my motion trails on health apps or moments on WeChat).	0.922	—	—
#46. I know how to use the sports functions on social network services (such as WeRun on WeChat) to interact with others (eg, thumb up, forward).	0.688	—	—
#47: I will target the advanced players I follow on the eHealth tools, learn from them, and catch up with them.	0.793	—	—
#48: I will try out some health-related suggestions available on the internet and control the risks (ie, get injured or take the wrong medicine).	0.811	—	—
Eigenvalue	14.439	1.658	1.114
Variance (%)	60.163	6.909	4.643
Cumulative (%)	60.163	67.072	71.715

[a]Not available.

Thus far, the current instrument has 24 items that were sorted based on 3 factors: acquisition (8 items), verification (6 items), and application (10 items). The total variance of this scale explained by these 3 factors was 71.715%. The internal consistency of each factor was also tested using the Cronbach α, as shown in Table 4. The result was satisfactory.

Bivariate correlations between the subscales were tested, and they are shown in Table 5. The moderate strength of the correlations indicated that the subscales were measuring related but distinct constructs.

Table 4. Reliability statistics.

Scale	Score, mean (SD)	Cronbach α
Scale in total	86.27 (14.5)	.971
Subscale of Factor 1	36.25 (7.5)	.962
Subscale of Factor 2	21.96 (7.5)	.934
Subscale of Factor 3	28.06 (4.0)	.913

Table 5. Correlations between the subscales.

	Subscale 1	Subscale 2	Subscale 3
Subscale 1			
r	1	—[a]	—
P value	—	—	—
Subscale 2			
r	0.774	1	—
P value	<.001	—	—
Subscale 3			
r	0.783	0.726	1
P value	<.001	<.001	—

[a]Not available.

It was surprising that positive and negative factor loadings existed simultaneously in the principal component analysis. Reviewing the scale showed that only item 5 was asked in a reverse manner, which might have confused participants and led to such a case. After discussion, the authors decided to refine item 5 as "I will obtain the health information (ie, medical, sport, or daily care information) on the internet." In addition, a conditional statement "If needed, I can…" was added to items 18 and 19 in case some participants did not use eHealth tools frequently. As a result, the adjusted 24 items in 3 dimensions comprising 6, 8, or 10 items in each were developed to yield the new eHLS-Web 3.0 instrument (see Multimedia Appendix 1).

Study 2

Construct Validity of eHLS-Web3.0

The primary CFA showed that the model did not fit the data very well, with fit indices of χ^2_{249}=1351.230, χ^2_{249}=5.427, CFI=0.873, TLI=0.860, RMSEA=0.077, and SRMR=0.056. According to the model modification indices, the residual errors of 5 items (#1, #15, #19, #21, and #23) were corrected. After modification, the eHLS-Web3.0 met the criteria for good model fit, with χ^2_{244}=903.076, χ^2_{244}=3.701, CFI=0.924, TLI=0.914, RMSEA=0.06, and SRMR=0.051. The factor loadings (see Multimedia Appendix 2) obtained using the 3-factor model were found to fit well with the data. The standardized loadings were all above 0.6, most of which were greater than 0.8. Correlations between the 3 factors were moderate to high.

Convergent and Concurrent Validities of eHLS-Web3.0

The AVE was calculated as 0.52 using the data collected in the EFA stage, showing an acceptable convergent validity. For the concurrent validity, the bivariate correlations between the subscale of the eHLS-Web3.0 and eHEALS were significant, showing a satisfactory result, as indicated in Table 6.

Table 6. Correlations between subscales of the electronic Health Literacy Scale-Web3.0 and eHealth Literacy Scale.

	eHLS-Web3.0[a]	Subscale 1	Subscale 2	Subscale 3
eHEALS[b]				
r	0.893	—[c]	—	—
P value	<.001	—	—	—
Subscale 1				
r	0.880	—	—	—
P value	<.001	—	—	—
Subscale 2				
r	—	—	0.725	—
P value	—	—	<.001	—
Subscale 3	—	—	—	—
r	—	—	—	0.853
P value	—	—	—	<.001

[a]eHLS-Web3.0: electronic Health Literacy Scale-Web3.0.

[b]eHEALS: eHealth Literacy Scale.

[c]Not applicable.

Internal Consistency and Test-Retest Reliabilities of eHLS-Web3.0

The Cronbach α, composite reliability coefficient, and the item-total correlations were calculated to examine the internal consistency reliability of the new instrument. The specific results are presented in Table 7, indicating satisfactory internal consistency reliability.

An acceptable test-retest reliability was achieved, showing that the eHLS-Web3.0 was relatively stable over time, as shown in Table 8.

Table 7. Internal consistency reliability statistics.

Statistic	Value
Subscale 1	
Range of interitem correlations	0.549-0.832
Minimum corrected item-total correlation	0.726
Composite reliability coefficient	0.937
Cronbach α	.950
Subscale 2	
Range of interitem correlations	0.726-0.807
Minimum corrected item-total correlation	0.828
Composite reliability coefficient	0.934
Cronbach α	.935
Subscale 3	
Range of interitem correlations	0.511-0.898
Minimum corrected item-total correlation	0.673
Composite reliability coefficient	0.956
Cronbach α	.958
eHLS-Web3.0[a]	
Range of interitem correlations	0.470-0.898
Minimum corrected item-total correlation	0.858
Composite reliability coefficient	0.981
Cronbach α	.976

[a]eHLS-Web3.0: electronic Health Literacy Scale-Web3.0.

Table 8. Correlations between test (T1) and retest (T2).

Test	T1	T2
T1		
Mean (SD)	83.97 (14.67)	—[a]
r	—	—
P value	<.001	—
T2		
Mean (SD)	—	86.37 (15.35)
r	—	0.858
P value	—	<.001

[a]Not applicable.

Measurement Invariance of eHLS-Web3.0

Multigroup CFAs were employed to examine whether the measurement was invariant across genders, majors, and regions. Results of the invariance analyses are provided in Tables 9-11.

For the gender invariance analysis presented in Table 9, it was found that the configural model (M0) showed satisfactory fit to the data. The matric model (M1) and strong model (M2) displayed satisfactory fit to the data, and their △CFI was less than 0.01; the △RMSEA was less than 0.01, and the △SRMR was less than 0.025. These indices supported the invariance of factor loadings and intercepts across genders. The strict model (M3) reflected acceptable fit to the data. Although the △SRMR in M3 showed poor fit, the other indices showed a satisfactory goodness of fit (△CFI and △RMSEA<0.01), indicating that the residual errors were equivalently invariant across genders.

For the major invariance analysis, the model fitness tests and comparisons were performed across all three samples, including general (nonhealth-related), sport, and medical major students. The configural model (M0) showed satisfactory fit to the data between general and sport as well as general and medical, whereas the configural model reflected a marginal-to-acceptable

fit (CFI=0.849) between sport and medical majors. This is most likely because the medical group had a much smaller sample size than the sport group. As for the goodness of model fit, the matric model (M1), strong model (M2), and strict model (M3) displayed satisfactory and acceptable fit to the data between general and sport as well as general and medical, and marginal-to-acceptable fit to the data between sport and medical groups. The △CFI and △RMSEA of all these models were less than 0.01, and the △SRMR was less than 0.025. These indices supported the invariance of the factor loadings, intercepts, and residual errors across majors, as observed in Table 10.

Table 9. Invariance analysis of the electronic Health Literacy Scale-Web3.0 across genders (nmale=354 and nfemale=387).

Model	χ^{2a} (*df*)	CFI[b]	△CFI[c]	RMSEA[d]	△RMSEA[e]	SRMR[f]	△SRMR[g]
M0[h]	1215.588 (488)	0.918	—[i]	0.063	—	0.057	—
M1[j]	1264.312 (488)	0.915	–0.003	0.063	0	0.064	0.007
M2[k]	1328.711 (488)	0.91	–0.005	0.063	0	0.067	0.003
M3[l]	1368.771 (539)	0.906	–0.004	0.064	0.001	0.205	0.138

[a]χ^2: chi-square ratio statistic.

[b]CFI: comparative fit index.

[c]△CFI: change in the comparative fit index.

[d]RMSEA: root mean square error of approximation.

[e]△RMSEA: change in the root mean square error of approximation.

[f]SRMR: standardized root mean residual.

[g]△SRMR: change in the standardized root mean residual.

[h]M0: baseline configural invariance model.

[i]Not available.

[j]M1: metric invariance model.

[k]M2: strong invariance model.

[l]M3: strict invariance model.

Table 10. Invariance analysis of the electronic Health Literacy Scale-Web3.0 across majors (ngeneral=572, nmedical=41, and nsport=128).

Model	χ^{2a} (*df*)	CFI[b]	△CFI[c]	RMSEA[d]	△RMSEA[e]	SRMR[f]	△SRMR[g]
General and sport-related majors							
M0[h]	1240.042 (488)	0.922	—[i]	0.066	—	0.05	—
M1[j]	1272.243 (509)	0.92	–0.002	0.065	–0.001	0.053	0.003
M2[k]	1320.117 (533)	0.918	–0.002	0.065	0	0.061	0.008
M3[l]	1326.536 (539)	0.918	0	0.065	0	0.059	–0.002
General and medical majors							
M0	1354.433 (488)	0.912	—	0.076	—	0.056	—
M1	1394.817 (509)	0.91	–0.002	0.075	–0.001	0.064	0.008
M2	1448.693 (533)	0.907	–0.003	0.075	0	0.066	0.002
M3	1451.485 (539)	0.907	0	0.074	–0.001	0.067	0.001
Medical and sport-related majors							
M0	952.825 (488)	0.849	—	0.106	—	0.068	—
M1	991.739 (509)	0.843	–0.006	0.106	0	0.092	0.024
M2	1031.023 (533)	0.838	–0.005	0.105	–0.001	0.095	0.003
M3	1035.032 (539)	0.839	0.001	0.104	–0.001	0.099	0.004

[a]χ^2: chi-square ratio statistic.

[b]CFI: comparative fit index.

[c]△CFI: change in the comparative fit index.

[d]RMSEA: root mean square error of approximation.

[e]△RMSEA: change in the root mean square error of approximation.

[f]SRMR: standardized root mean residual.

[g]△SRMR: change in the standardized root mean residual.

[h]M0: baseline configural invariance model.

[i]Not applicable.

[j]M1: metric invariance model.

[k]M2: strong invariance model.

[l]M3: strict invariance model.

For the region invariance analysis, the model tests and comparisons were conducted across all the four groups of samples, including students from Kunming, Yinchuan, Xiamen, and Beijing. The configural model (M0) showed acceptable fit to the data across all the four regional samples. For the goodness of model fit, the matric model (M1) and strong model (M2) for every group displayed acceptable fit to the data. For every group, the △CFI and △RMSEA of the models were less than 0.01, and the △SRMR was less than 0.025. These indices supported the invariance of the factor loadings and intercepts across regions. The strict model (M3) for each group also reflected acceptable fit to the data. However, between Kunming and Yinchuan, Kunming and Xiamen, Yinchuan and Xiamen, and Xiamen and Beijing, the △SRMR of the strict model (M3) reflected a poor fit, as shown in Table 11. Despite this, the other indices showed a satisfactory goodness of fit (the △CFI and △RMSEA were less than 0.01), proving that the residual errors were equivalently invariant across the regions.

In summary, these findings supported the measurement invariance of the eHLS-Web3.0 model (factor loadings, intercepts, and residual errors) across genders, majors, and regions.

Table 11. Invariance analysis of the electronic Health Literacy Scale-Web3.0 across regions (nKunming=216, nYinchuan=114, nXiamen=287, and nBeijing=124).

Region	χ^{2}[a] (*df*)	CFI[b]	△CFI[c]	RMSEA[d]	△RMSEA[e]	SRMR[f]	△SRMR[g]
Kunming and Yinchuan							
M0[h]	1045.16 (488)	0.875	—[i]	0.083	—	0.068	—
M1[j]	1074.43 (509)	0.873	–0.002	0.082	–0.001	0.076	0.008
M2[k]	1126.886 (533)	0.867	–0.006	0.082	0	0.082	0.006
M3[l]	1137.122 (539)	0.866	–0.001	0.082	0	0.107	0.025
Kunming and Xiamen							
M0	1127.636 (488)	0.908	—	0.072	—	0.058	—
M1	1172.651 (509)	0.904	–0.004	0.072	0	0.071	0.013
M2	1228.577 (533)	0.899	–0.005	0.072	0	0.079	0.008
M3	1249.567 (539)	0.897	–0.002	0.072	0	0.182	0.103
Kunming and Beijing							
M0	1023.786 (488)	0.879	—	0.08	—	0.074	—
M1	1038.028 (509)	0.88	0.001	0.078	–0.002	0.078	0.004
M2	1095.957 (533)	0.872	–0.008	0.079	0.001	0.083	0.005
M3	1102.118 (539)	0.872	0	0.078	–0.001	0.098	0.015
Yinchuan and Xiamen							
M0	1043.404 (488)	0.908	—	0.075	—	0.047	—
M1	1078.271 (509)	0.906	–0.002	0.075	0	0.054	0.007
M2	1118.863 (533)	0.903	–0.003	0.074	–0.001	0.055	0.001
M3	1128.242 (539)	0.903	0	0.074	0	0.109	0.054
Yinchuan and Beijing							
M0	917.12 (488)	0.874	—	0.086	—	0.067	—
M1	936.311 (509)	0.875	0.001	0.084	–0.002	0.076	0.009
M2	1008.445 (533)	0.861	–0.014	0.087	0.003	0.085	0.009
M3	1024.654 (539)	0.858	–0.003	0.087	0	0.094	0.009
Xiamen and Beijing							
M0	1026.47 (488)	0.913	—	0.073	—	0.054	—
M1	1058.117 (509)	0.912	–0.001	0.072	–0.001	0.063	0.009
M2	1126.022 (533)	0.904	–0.008	0.074	0.002	0.069	0.006
M3	1137.64 (539)	0.904	0	0.074	0	0.117	0.048

[a]χ^{2}: chi-square ratio statistic.

[b]CFI: comparative fit index.

[c]△CFI: change in the comparative fit index.

[d]RMSEA: root mean square error of approximation.

[e]△RMSEA: change in the root mean square error of approximation.

[f]SRMR: standardized root mean residual.

[g]△SRMR: change in the standardized root mean residual.

[h]M0: baseline configural invariance model.

[i]Not available.

[j]M1: metric invariance model.

[k]M2: strong invariance model.

[l]M3: strict invariance model.

Discussion

Principal Findings

The current study developed and tested a new measurement tool for EHL in Chinese with the Web 3.0 context named eHLS-Web3.0. A multistage program was applied, generating 24 items that represent the updated content of EHL in the current internet environment. In comparison with other eHealth models [3,5], the new instrument measures not only the skills for searching eHealth information (Health 1.0 skills) or communicating with service providers (Health 2.0 skills), but also the skills for building personal health data sets, self-tracking, and protecting privacy (Health 3.0 skills). Although the items were built upon the groundwork laid by the eHEALS [3], with the insights from the participants' usage of the Web 3.0 technology, eHLS-Web3.0 went beyond the computer skill and media literacy components of the eHEALS. It provided a much deeper understanding of people's interactions with different types of eHealth tools and the application of the information they obtained or created. Construction and validity testing in a broad range of target groups generated clear evidence of construct, convergent, and concurrent validities, as well as composite and internal consistency, and test-retest reliability. Measurement invariance was also found across genders, majors, and regions. This initial validity test indicated that the eHLS-Web3.0 was likely to be valuable for the characterization and understanding of EHL.

Study 1

An effective eHealth literacy scale was developed in this study. When generating the item pool, it was hypothesized that the eHLS-Web3.0 would be categorized into 3 dimensions, namely Health 1.0, Health 2.0, and Health 3.0, for the items were generated based on the use experience of eHealth tools in different information technology (IT) generations. However, the result was different from what was hypothesized. Each item was applicable for evaluating the usage of eHealth tools in all the IT generations. The result revealed that current EHL is a comprehensive constellation of abilities where the 3 factors were correlated, distinguished, and combined, together influencing people's usage of eHealth tools, irrespective of the generations of the tools. Moreover, during the development, the features of EHL were examined by EFA, which revealed that the data supported a 3-factor structure. This 3-factor structure differed from those in other concept-based EHL scales used in previous studies. For instance, the most widely used one, eHEALS, is a unidimensional tool that only focuses on individuals' Health 1.0 abilities [3]. As an updated measurement tool, it is reasonable for eHLS-Web3.0 to be categorized into more factors. The 8-subscale Patient Readiness to Engage in Health Internet Technology [12] and 7-subscale eHLQ [15] considered the internal perceptions (eg, need, motivation, sense of safety, and anxiety) and external environment (relationship with service provider and access to the tools) along with the skills and abilities. In accordance with Norman and Skinner [3], this study defined EHL as the set of capabilities for people to properly understand all types of web-based health information and maintain their health and did not consider other internal and external factors for item generation so that the eHLS-Web3.0 could have a concise and targeted structure. It is worth mentioning that the factor structure of eHLS-Web3.0 is similar to the 3D eHLS [14]. The eHLS was developed following a thorough literature review, including functional, interactive, and critical dimensions of the eHLS. In the eHLS, the functional dimension is about understanding and calculating, the interactive one is about filtering, and the critical one is about cross-checking and evaluating applicability. No Web 2.0- or 3.0-related item was mentioned in the eHLS. Comparing the eHLS and eHLS-Web3.0 showed that the acquisition subscale in eHLS-Web3.0 was similar to the first and second dimensions in the eHLS (functional and interactive), whereas the verification part of eHLS-Web3.0 was close to the critical subscale in the eHLS. The application part is a newly developed dimension in the current EHL scale, which is consistent with daily experiences, as the application of information has become much more complicated than before. Others than apply the information to make health decisions or solve problems; individuals could also create their own health data, use the information to communicate, respond (ie, giving suggestions and advice, and responding to help seekers), socialize with others, and share and post information (ie, forward helpful information or post their own health or fitness data). Therefore, it is essential to enrich and include the abilities of information application into EHL. It was confirmed that the factorial structure of the eHLS-Web3.0 was reliable and valid based on statistics and theories, and in line with the actual situation.

Study 2

The aim of this study was to examine the validity (concurrent, convergent, and construct), reliability (composite, internal consistency, and test-retest reliability) and measurement invariance (gender, major, and region) of eHLS-Web3.0. The construct validity was determined through CFA; the goodness-of-fit indices supported the 3D model structure in study 1 and confirmed the factorial composition of eHLS-Web3.0. The concurrent validity was supported by its high correlation with other well-established and validated measures of EHL (eHEALS) [3]. The convergent validity was confirmed by the AVE values, with values above 0.5 indicating acceptability. Regarding the reliability, the Cronbach α estimation showed adequate internal consistency reliability for eHLS-Web3.0 and its subscales. The item-total correlations were also tested, and they showed relatively high positive values, indicating that the items in eHLS-Web3.0 discriminated effectively between high- and low-performing participants. Given the limitation of the Cronbach α approach [46], the composite reliability coefficients were also calculated for eHLS-Web3.0, providing further positive evidence for the

reliability of eHLS-Web3.0. Furthermore, the measurement invariance was measured for eHLS-Web3.0 across genders, majors, and regions. The establishment of configural, metric, strong, and structural invariances demonstrated that eHLS-Web3.0 is an appropriate and a meaningful instrument to measure EHL across diverse groups and perform comparisons. Compared with previous studies evaluating EHL among Chinese people [17,35,47], the current research had the advantage of using student samples drawn from different regions in China and from different majors. Besides, a follow-up questionnaire was designed to check if eHLS-Web3.0 was stable over time. Most importantly, this study explored the latest usage behavior related to eHealth, based on which new factors were meticulously developed to enrich the contents of EHL.

Limitations, Strengths, and Future Directions

The study has several limitations. First, the generalizability of the findings might be hindered by the snowball sampling. As a result, a stratified sampling approach is desirable in future [48]. In addition, the results are only based on the responses provided by Chinese college students; therefore, the applicability of this tool in other groups needs to be examined. Although a purposive sample was sought, it is possible that those who responded were more interested in eHealth than nonrespondents, and this may have biased the result and the level of EHL. Moreover, self-reporting questionnaires were used, which may result in unreliability and inaccuracy because of the inherent drawbacks of self-reporting (ie, recall bias, over- or under-reporting, and social desirability) [49]. Finally, the nonhealth-related major participants were over-represented in the sample, which reflected certain realities and may also lead to some bias. Consequently, further invariance tests across majors are required with acceptably equalized sample sizes in different groups.

Despite the limitations mentioned above, the current study fills a gap in the literature, as no specific measure of EHL considering the Web 3.0 IT environment has been previously developed. A comprehensive understanding of EHL was achieved in the context of new IT and Chinese culture. By developing this reliable and valid measurement tool, this study provides an up-to-date tool to measure the level of EHL among Chinese college students and evaluate the efficacy of EHL interventions in future experimental studies on EHL. Moreover, the relationships among EHL and other health-related variables could be explored using this new tool. Apart from the previously stated theoretical values, the newly developed EHL scale may also be applied to eHealth-related interventions, such as providing new and valuable information for constructing EHL training methods.

Conclusions

A new 24-item instrument for measuring eHealth literacy termed the eHLS-Web3.0 was developed and initially validated in this study. Three factors were identified in this tool, namely acquisition, verification, and application. This study validated the eHLS-Web3.0 in terms of its construct validity, convergent validity, concurrent validity, internal consistency reliability, and test-retest reliability in Chinese college students. The measurement invariance of the eHLS-Web3.0 was also confirmed across genders, majors, and regions. The 24-item eHLS-Web3.0 is a valid and reliable instrument for assessing eHealth literacy in the Chinese context.

Acknowledgments

HXL is grateful for the immense support provided by Qing-jian Liu, Guang-ze Zhang, Xiao-yi Yang, and Hong Wei in reaching participants and distributing the questionnaire. The authors appreciate Chun Hu and Ming-Wa Kwai for their valuable inputs.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Final items of eHealth Literacy Scale-Web3.0.
[PDF File (Adobe PDF File), 1082 KB - jmir_v23i9e31627_app1.pdf]

Multimedia Appendix 2
Results of factor loading for the construct validity test of eHealth Literacy Scale-Web3.0.
[PDF File (Adobe PDF File), 89 KB - jmir_v23i9e31627_app2.pdf]

References

1. World Health Organization. Atlas of eHealth country profiles: the use of eHealth in support of universal health coverage. WHO Library Cataloguing-in-Publication Data. 2015. URL: http://apps.who.int/iris/bitstream/handle/10665/204523/9789241565219_eng.pdf;jsessionid=E6309DB207ED3C7336432E0184CB1865?sequence= [accessed 2018-06-15]
2. Norgaard O, Furstrand D, Klokker L, Karnoe A, Batterham R, Kayser L, et al. The e-health literacy framework: a conceptual framework for characterizing e-health users and their interaction with e-health systems. Knowl Manag E-Learn 2015 Dec;7(4):522-540. [doi: 10.34105/j.kmel.2015.07.035]
3. Norman CD, Skinner HA. eHEALS: the eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]

4. Karnoe A, Kayser L. How is eHealth literacy measured and what do the measurements tell us? a systematic review. Knowl Manag E-Learn 2015 Dec;7(4):576-600. [doi: 10.34105/j.kmel.2015.07.038]
5. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015 Mar;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
6. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
7. Zeldman J. Web 3.0. A List Apart. 2006 Jan 17. URL: https://alistapart.com/article/web3point0/ [accessed 2018-02-18]
8. Rudman R, Bruwer R. Defining Web 3.0: opportunities and challenges. Electron.Libr 2016 Feb;34(1):132-154. [doi: 10.1108/el-08-2014-0140]
9. Delgado RH. Web 2.0 history, evolution and characteristics. disenowebakus.net. 2019. URL: https://disenowebakus.net/en/web-2 [accessed 2020-06-08]
10. van der Vaart R, Drossaert C. Development of the Digital Health Literacy Instrument: measuring a broad spectrum of Health 1.0 and Health 2.0 skills. J Med Internet Res 2017 Jan;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]
11. Hanik B, Stellefson M. E-Health Literacy Competencies among Undergraduate Health Education Students: A Preliminary Study. Int Electron J Health Educ 2011 Apr;14:46-58 [FREE Full text]
12. Koopman RJ, Petroski GF, Canfield SM, Stuppy JA, Mehr DR. Development of the PRE-HIT instrument: patient readiness to engage in health information technology. BMC Family Practice 2014 Jan 28;15(1):1-9. [doi: 10.1186/1471-2296-15-18]
13. Seçkin G, Yeatts D, Hughes S, Hudson C, Bell V. Being an informed consumer of health information and assessment of electronic health literacy in a national sample of internet users: validity and reliability of the e-HLS instrument. J Med Internet Res 2016 Jul;18(7):e161 [FREE Full text] [doi: 10.2196/jmir.5496] [Medline: 27400726]
14. Hsu W, Chiang C, Yang S. The effect of individual factors on health behaviors among college students: the mediating effects of eHealth literacy. J Med Internet Res 2014 Dec;16(12):e287 [FREE Full text] [doi: 10.2196/jmir.3542] [Medline: 25499086]
15. Kayser L, Karnoe A, Furstrand D, Batterham R, Christensen KB, Elsworth G, et al. A multidimensional tool based on the eHealth literacy framework: development and initial validity testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018 Feb;20(2):e36. [doi: 10.2196/jmir.8371]
16. Karnoe A, Furstrand D, Christensen KB, Norgaard O, Kayser L. Assessing competencies needed to engage with digital health services: development of the eHealth Literacy Assessment Toolkit. J Med Internet Res 2018 May;20(5):e178 [FREE Full text] [doi: 10.2196/jmir.8347] [Medline: 29748163]
17. Ma Z, Wu M. The psychometric properties of the Chinese eHealth Literacy Scale (C-eHEALS) in a Chinese rural population: cross-sectional validation study. J Med Internet Res 2019 Oct;21(10):e15720 [FREE Full text] [doi: 10.2196/15720] [Medline: 31642811]
18. Liang W, Duan Y, Shang B, Hu C, Baker JS, Lin Z, et al. Precautionary behavior and depression in older adults during the COVID-19 pandemic: an online cross-sectional study in Hubei, China. Int J Environ Res Public Health 2021 Feb;18(4):1853 [FREE Full text] [doi: 10.3390/ijerph18041853] [Medline: 33672885]
19. Lv Y, Zhan Y, Ye B. Doctor Dingxiang 2021 National Health Insight Report. 2021 Jan 24. URL: https://wechat.dxy.cn/japi/file/ebfb0761febe8ae093f88e41ee651ff0/18580 [accessed 2021-02-04]
20. Robb M, Shellenbarger T. Influential factors and perceptions of eHealth literacy among undergraduate college students. Online J Nurs Inform 2014 Oct;18(3):1.
21. Goodman LA. Snowball sampling. Ann Math Statist 1961 Mar;32(1):148-170. [doi: 10.1214/aoms/1177705148]
22. Gilstad H. Toward a comprehensive model of eHealth literacy. In: Proceedings of the 2nd European Workshop on Practical Aspects of Health Informatics. 2014 May Presented at: The 2nd European Workshop on Practical Aspects of Health Informatics; May 19, 2014; Trondheim, Norway URL: http://ceur-ws.org [doi: 10.1007/978-0-387-72596-3_16]
23. National Bureau of Statistics of China. China Statistical Yearbook 2018. China Statistics Press. 2018. URL: http://www.stats.gov.cn/tjsj/ndsj/2018/indexeh.htm [accessed 2020-06-12]
24. Koetse M. Behind the Baidu scandal: Baidu and the Putian medical group. What's on Weibo. 2016. URL: https://www.whatsonweibo.com/behind-baidu-scandal-baidu-putian-medical-group/ [accessed 2020-09-18]
25. Bryman A. Social Research Methods. Oxford, England: Oxford University Press; 2016:62.
26. Guest G, Bunce A, Johnson L. How many interviews are enough? Field Methods 2016 Jul;18(1):59-82. [doi: 10.1177/1525822x05279903]
27. Lynn MR. Determination and quantification of content validity. Nurs Res 1986;35(6):382-385. [doi: 10.1097/00006199-198611000-00017]
28. Denscombe M. The Good Research Guide: For Small-Scale Social Research Projects. London, England: McGraw-Hill Education (UK); 2014:41.
29. Schwab DP. Construct validity in organization behavior. In: Staw BM, Cummings LL, editors. Research in Organizational Behavior. Greenwich, CT: JAI press; 1980:3-43.

30. Guadagnoli E, Velicer WF. Relation of sample size to the stability of component patterns. Psychol Bull 1988 Mar;103(2):265-275 [FREE Full text] [doi: 10.1037/0033-2909.103.2.265] [Medline: 3363047]
31. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health 2007 Aug;30(4):459-467 [FREE Full text] [doi: 10.1002/nur.20199] [Medline: 17654487]
32. Park HS, Dailey R, Lemus D. The use of exploratory factor analysis and principal components analysis in communication research. Human Comm Res 2002 Oct;28(4):562-577 [FREE Full text] [doi: 10.1111/j.1468-2958.2002.tb00824.x]
33. Henson RK, Roberts JK. Use of exploratory factor analysis in published research: common errors and some comment on improved practice. Educ Psychol Meas 2016 Jul;66(3):393-416 [FREE Full text] [doi: 10.1177/0013164405282485]
34. Hoelter JW. The analysis of covariance structures: goodness-of-fit indices. Sociol Methods Res 2016 Jun;11(3):325-344 [FREE Full text] [doi: 10.1177/0049124183011003003]
35. Guo SJ, Yu XM, Sun YY, Nie D, Li XM, Wang L. Adaptation and evaluation of Chinese version of eHEALS and its usage among senior high school students. Chin J Health Educ 2013;29(2):106-108 [FREE Full text]
36. Byrne D, Clore GL. Effectance arousal and attraction. J Pers Soc Psychol 1967 Aug;6(4):1-18. [doi: 10.1037/h0024829] [Medline: 5588183]
37. Byrne BM. Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming, Second Edition. New York: Routledge, Taylor & Francis Group; 2013.
38. Jöreskog KG, Sörbom D. LISREL 8: Structural equation modeling with the SIMPLIS command language. United States: Scientific Software International, Inc; 1993.
39. Bentler PM. Comparative fit indexes in structural models. Psychol Bull 1990 Mar;107(2):238-246. [doi: 10.1037/0033-2909.107.2.238] [Medline: 2320703]
40. Bollen KA. Structural Equations with Latent Variables. New York: John Wiley & Sons; 1989.
41. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 1999 Jan;6(1):1-55 [FREE Full text] [doi: 10.1080/10705519909540118]
42. Steiger JH. A note on multiple sample extensions of the RMSEA fit index. Struct Equ Modeling 2009 Nov;5(4):411-419. [doi: 10.1080/10705519809540115]
43. Raykov T. Estimation of composite reliability for congeneric measures. Appl Psychol Meas 2016 Jul;21(2):173-184 [FREE Full text] [doi: 10.1177/01466216970212006]
44. Steenkamp JEM, Baumgartner H. Assessing measurement invariance in cross-national consumer research. J Consum Res 1998 Jun;25(1):78-107 [FREE Full text] [doi: 10.1086/209528]
45. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equ Modeling 2002 Apr;9(2):233-255 [FREE Full text] [doi: 10.1207/s15328007sem0902_5]
46. Bacon DR, Sauer PL, Young M. Composite reliability in structural equations modeling. Educ Psychol Meas 2016 Jul;55(3):394-406. [doi: 10.1177/0013164495055003003]
47. Koo M, Norman CD, Hsiao-Mei C. Psychometric evaluation of a Chinese version of the eHealth literacy scale (eHEALS) in school age children. Glob J Health Educ Promot 2012;15(1):29-36.
48. Liang W, Wang DD, Shang BR, Zhang CQ, Duan YP, Si GY. Further examination of the psychometric properties of the Brief Self-Control Scale: evidence from Chinese athletes and students. Int J Sport Exerc Psychol 2020 Sep:1-20. [doi: 10.1080/1612197x.2020.1827000]
49. Liang W, Duan YP, Shang BR, Wang YP, Hu C, Lippke S. A web-based lifestyle intervention program for Chinese college students: study protocol and baseline characteristics of a randomized placebo-controlled trial. BMC Public Health 2019 Aug;19(1):1-11. [doi: 10.1186/s12889-019-7438-1]

Abbreviations

AVE: average variance extracted
CFA: confirmatory factor analysis
CFI: comparative fit index
EFA: exploratory factor analysis
EHL: eHealth literacy
eHEALS: eHealth Literacy Scale
eHLA: Health Literacy Assessment Toolkit
eHLS: electronic Health Literacy Scale
eHLQ: eHealth Literacy Questionnaire
I-CVI: content validity index of each item
IT: information technology
RMSEA: root mean square error of approximation
SRMR: standardized root mean residual

Edited by G Eysenbach; submitted 28.06.21; peer-reviewed by C Mather, SL Lee; comments to author 20.07.21; revised version received 26.07.21; accepted 01.08.21; published 23.09.21.

Please cite as:
Liu HX, Chow BC, Liang W, Hassel H, Huang YW
Measuring a Broad Spectrum of eHealth Skills in the Web 3.0 Context Using an eHealth Literacy Scale: Development and Validation Study
J Med Internet Res 2021;23(9):e31627
URL: https://www.jmir.org/2021/9/e31627
doi:10.2196/31627
PMID:34554098

Original Paper

The Association Between Patients' eHealth Literacy and Satisfaction With Shared Decision-making and Well-being: Multicenter Cross-sectional Study

Richard Huan Xu[1,2], PhD; Ling-Ming Zhou[3], PhD; Eliza Lai-Yi Wong[2], ProfDr; Dong Wang[3], ProfDr

[1]Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hong Kong, Hong Kong

[2]JC School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, Hong Kong

[3]School of Health Management, Southern Medical University, Guangzhou, China

Corresponding Author:
Dong Wang, ProfDr
School of Health Management
Southern Medical University
No. 1023 Shatai nan Road
Guangzhou, 510515
China
Phone: 86 61648573
Email: dongw96@smu.edu.cn

Abstract

Background: Although previous studies have shown that a high level of health literacy can improve patients' ability to engage in health-related shared decision-making (SDM) and improve their quality of life, few studies have investigated the role of eHealth literacy in improving patient satisfaction with SDM (SSDM) and well-being.

Objective: This study aims to assess the relationship between patients' eHealth literacy and their socioeconomic determinants and to investigate the association between patients' eHealth literacy and their SSDM and well-being.

Methods: The data used in this study were obtained from a multicenter cross-sectional survey in China. The eHealth Literacy Scale (eHEALS) and Investigating Choice Experiments Capability Measure for Adults were used to measure patients' eHealth literacy and capability well-being, respectively. The SSDM was assessed by using a self-administered questionnaire. The Kruskal-Wallis one-way analysis of variance and Wilcoxon signed-rank test were used to compare the differences in the eHEALS, SSDM, and Investigating Choice Experiments Capability Measure for Adults scores of patients with varying background characteristics. Ordinary least square regression models were used to assess the relationship among eHealth literacy, SSDM, and well-being adjusted by patients' background characteristics.

Results: A total of 569 patients completed the questionnaire. Patients who were male, were highly educated, were childless, were fully employed, were without chronic conditions, and indicated no depressive disorder reported a higher mean score on the eHEALS. Younger patients ($SSDM_{\geq 61\ years}$=88.6 vs $SSDM_{16\text{-}30\ years}$=84.2) tended to show higher SSDM. Patients who were rural residents and were well paid were more likely to report good capability well-being. Patients who had a higher SSDM and better capability well-being reported a significantly higher level of eHealth literacy than those who had lower SSDM and poorer capability well-being. The regression models showed a positive relationship between eHealth literacy and both SSDM (β=.22; *P*<.001) and well-being (β=.26; *P*<.001) after adjusting for patients' demographic, socioeconomic status, lifestyle, and health status variables.

Conclusions: This study showed that patients with a high level of eHealth literacy are more likely to experience optimal SDM and improved capability well-being. However, patients' depressive status may alter the relationship between eHealth literacy and SSDM.

(J Med Internet Res 2021;23(9):e26721) doi:10.2196/26721

KEYWORDS

eHealth literacy; shared decision-making; well-being; eHEALS; ICECAP-A

Introduction

Background

eHealth literacy refers to the acquisition and use of web-based information and communication technology to make appropriate health decisions [1]. Unlike collecting health information through traditional methods (eg, hospital pamphlets and medical magazines), acquiring information from the internet requires extended skills [2-4]. For this task, people need to have professional knowledge about specific health issues, computer and mobile phone literacy, knowledge and skills to navigate the internet, and the ability to analyze and digest web-based information [5,6]. In recent decades, the rapid proliferation of web-based information about health and health care has significantly changed individuals' health-seeking behaviors, such as participating in web-based communities or purchasing products on the internet and services to improve health and well-being. The internet increasingly serves as a major source of health information for individuals to understand their health concerns, instead of seeking professional advice [2,7].

The internet provides a convenient way to approach health-related information to the public; however, a low level of eHealth literacy may lead, in contrast, to serious harm [8] and health-related social inequality [9,10]. Previous studies have shown that individuals with a low level of eHealth literacy are more likely to report insufficient use of preventive health services [11], negative health-related attitudes [12], unhealthy behaviors and lifestyles [13], and poor medical adherence [14]. However, globally, the relationship between patients' eHealth literacy and well-being remains insufficiently explored. Nabi et al [15] indicated that seeking information from social networks (eg, Facebook) impacts people's stress levels and, in turn, influences their physical and psychological well-being. Another systematic review found that providing breast cancer patients with access to digital systems or technological devices could improve their health and well-being [16]. Given the complex and fragmented nature of the current health care systems and the high prevalence of chronic conditions, the internet has been increasingly identified as an essential and valuable information source to support patient-centered care and help patients and their families seek cost-effective health care services [17].

Recently, shared decision-making (SDM) has been reported to be an effective way to improve trust in patient–doctor relationships, reduce negative emotions, and promote patients' well-being [18,19]. eHealth literacy, as an important concept rooted in the practice of patient-centered care, is increasingly suggested to be used to improve SDM in clinical practice. For example, Nejati et al [20] found that low levels of eHealth literacy can limit patients' trust in the health care system and their communication patterns and are a barrier to patient participation in the decision-making process. As the global population becomes increasingly reliant on the internet to locate and obtain health information and services [21,22], patients and their caregivers struggle to possess adequate eHealth literacy to engage in the decision-making process. The limited availability of web-based health information restricts patients from participating in their health care decision-making process. For example, Car et al [23] pointed out that poor eHealth literacy limited patients' ability to make decisions in medicine management. Netjati et al [20] also showed that lower levels of eHealth literacy are associated with poorer SDM among patients with multiple myeloma. However, in China, there is no evidence regarding the relationship between eHealth literacy and medical decision-making.

As reported, there are more than 980 million internet users in China, accounting for more than 20% of the users worldwide [24]. In 2018, the State Council of China, jointly with the National Health Commission, released a series of decrees to encourage integration of traditional industries with internet technologies to improve the quality and efficiency of health care services [25]. A new nationwide web-based service system will be developed to provide patients with a novel way to approach quality health care information and facilitate their active engagement in clinical decision-making. This could help health care providers not only in understanding patients' preferences, needs, and satisfaction but also in clarifying their health care situations, treatment options, and likely outcomes [26]. Thus, an individual's level of eHealth literacy is the key to searching and using internet-based health care services to improve their health outcomes and well-being [27]. In 2020, the COVID-19 pandemic further proved that innovative eHealth approaches are vital for delivering health services and supporting patients to prevent contracting COVID-19 and increase their willingness to get vaccinated [28,29]. However, in China, there is a dearth of information regarding patients' level of eHealth literacy and whether eHealth literacy could improve their satisfaction with SDM (SSDM) and well-being in clinical practice. Without adequate information about these associations, there is a risk that internet-based interventions may lead to some negative outcomes, such as producing a digital divide, solidifying current health disparities, and perpetuating inequities, all of which could result in poor health outcomes and well-being [21].

Objectives

This study aims to (1) assess the relationship between patients' eHealth literacy and their socioeconomic status (SES) and (2) investigate the association between patients' eHealth literacy and their SSDM and well-being.

Methods

Study Design and Data Collection

The data used in this study were obtained from a multicenter cross-sectional survey that investigated patients' attitudes toward patient-centered care (PCC) in Guangdong province, China, from November 2019 to January 2020. Patients were recruited from the inpatient departments of 8 hospitals from 5 cities (Guangzhou, Shenzhen, Zhanjiang, Meizhou, and Shaoguan). All patients from the target hospitals were invited to participate in the survey during the survey period. The inclusion criteria were as follows: (1) being aged ≥18 years, (2) being able to read and speak Chinese, (3) having no cognitive impairment, and (4) being able to provide informed consent. With the assistance of ward nurses, all eligible patients were invited to

participate in the survey. The patients who agreed to participate in the survey and provided written informed consent were asked to complete a structured questionnaire that included questions about their demographic characteristics, SES, health conditions, well-being, use of health services, lifestyle, and attitudes toward PCC. A convenience sample of 569 patients (569/800, 71.1% response rate) successfully completed the questionnaire and provided valid responses. The study protocol and informed consent were approved by the institutional review board of the Second Affiliated Hospital of Guangzhou Medical University (reference ID: 2019-ks-28).

Measures

eHealth Literacy

The eHealth Literacy Scale (eHEALS) was used to measure consumers' combined knowledge, comfort, and perceived skills at finding, evaluating, and applying eHealth information to manage health problems [30]. It was developed based on a framework that comprised six dimensions to understand and use eHealth information [31]. The eHEALS has eight items that are rated on a 5-point Likert scale (including "strongly disagree," "disagree," "neutral," "agree," and "strongly agree"). The sum score ranges from 8 to 40, where a higher score indicates greater perceived eHealth literacy. To compare it with the results of other measures, in this study, the eHEALS sum score was converted to an overall score between 0 and 100 based on minimum-maximum normalization. A simplified Chinese eHEALS was used in this study [32].

Well-being

The Investigating Choice Experiments Capability Measure for Adults (ICECAP-A) is a generic and preference-based instrument that evaluates an individual's capability well-being [33]. Each dimension of the ICECAP-A comprises 1 item with 4 response options that range from "not capable" to "fully capable," to measure the different aspects related to capability well-being. The results of the ICECAP-A can convert to a summarized utility score that ranges from 0 to 1 to support the economic evaluation of social care interventions [34]. In this study, the ICECAP-A sum score was calculated using the scoring formula provided by the University of Birmingham. To facilitate comparability with the other measures, in this study, we converted the original ICECAP-A utility score to a range between 0 and 100. The Chinese version of the ICECAP-A was used [35].

Patient SSDM

Patient SSDM was assessed using a self-administered questionnaire. It was developed based on our previous patient engagement framework [36] and index [37], findings from literature review, focus group interviews (including patients, doctors, nurses, and policy makers), and expert discussion. It assessed patient satisfaction with decision-making, along with doctors in clinical practice. The SSDM comprises 5 items to measure different dimensions of satisfaction with the SDM. They are (1) "Did doctors provide several selections for you when making decisions (selection)," (2) "Did doctors carefully listen to your health problems when making decisions (Listen)," (3) "Did doctors respect your willingness when making choices (Respect)," (4) "Did doctors fully discuss your concerns with you when making decisions (Discussion)," and (5) "Did doctors fully understand your preferences and needs when making decisions (preference)?" Each item was rated on a 5-point scale ranging from "strongly disagree" to "strongly agree." The instrument showed good content, construct (Multimedia Appendix 1), convergent validity (Multimedia Appendix 2), and high internal consistency reliability (Cronbach α=.93). The overall score of the SSDM was calculated by adding up the scores of each item, which were then converted to a range of 0 to 100 based on minimum-maximum normalization.

Depressive Disorder

The Patient Health Questionnaire-2 was used to assess whether patients experienced depressed mood over the past 2 weeks. An individual with a score of 3 or above (range: 0-6) was recognized as someone with a depressive disorder [38].

Statistical Analysis

Patients' background characteristics (section 1: demographics; section 2: SES; section 3: lifestyle; and section 4: health status) were presented with the mean and SD of the eHEALS, SSDM, and ICECAP-A sum scores. The Kruskal-Wallis one-way analysis of variance (multiple groups) and Wilcoxon signed-rank test (2 groups) were used to compare the differences in the eHEALS, SSDM, and ICECAP-A sum scores of patients with different background characteristics. The Wilcoxon signed-rank test was also used to assess the relationship between level of eHealth literacy and SSDM and well-being. Patients' level of eHealth literacy was recategorized into high (≥30) and low (<30) on the basis of the median of the original eHEALS sum score. In addition, patients' depressive status was considered in the analysis of the relationship between 3 measures. Three ordinary least square multivariate regression models were developed to assess the relationships between measures adjusted by patients' background characteristics. In the first model, the dependent variable was eHealth literacy, and the independent variables were SSDM, well-being, and patients' background characteristics. In the second model, SSDM was the dependent variable, and the independent variables were eHealth literacy and patients' background characteristics. In the third model, the dependent variable was capability well-being, and the independent variables were eHealth literacy and patients' background characteristics. The objective of the first model was to assess the relationship between patients' eHealth literacy and their socioeconomic determinants, whereas the other two models assessed how patients' eHealth literacy can predict the changes in their SSDM and well-being, after adjusting for background characteristics. The Bland-Altman (B-A) plot was used to assess the agreement between three measures. The mean scores of the eHEALS, SSDM, and ICECAP-A were plotted on the x-axis, and the differences between them was plotted on the y-axis. The observations clustered evenly around a horizontal line representing y=0, reflecting good agreement between the measures [39]. R software (R Foundation for Statistical Computing) was used to perform all statistical analyses. The level of statistical significance was set at $P \leq .05$.

Results

Participants' Characteristics and the Results of Measures

Table 1 shows that more than half of the patients were male (288/569, 50.6%), and approximately 18.6% (106/569) and 19.2% (109/569) were aged <30 years and >60 years, respectively. Nearly 51.1% (291/569) of the patients indicated living with at least one kind of chronic condition, and 27.9% (159/569) reported having a depressive disorder. Patients who were male ($eHEALS_{male}$=68.5 vs $eHEALS_{female}$=64.3), were highly educated ($eHEALS_{tertiary}$=69.6 vs $eHEALS_{no\ or\ primary}$=62.9), were childless ($eHEALS_{no\ child}$=72.7 vs $eHEALS_{with\ child}$=59.4), fully employed ($eHEALS_{fully\ employed}$=68.5 vs $eHEALS_{nonemployed}$=61.7), were without chronic conditions ($eHEALS_{no\ chronic\ conditions}$=68.5 vs $eHEALS_{with\ chronic\ conditions}$=64.4), and indicated no depressive disorder ($eHEALS_{no\ depression}$=67.6 vs $eHEALS_{with\ depression}$=63.3) reported a higher level of eHealth literacy. Younger patients ($SSDM_{\geq 61\ years}$=88.6 vs $SSDM_{16\text{-}30\ years}$=84.2) tended to show higher SSDM. Patients, who resided in rural areas ($ICECAP\text{-}A_{rural}$=79.3 vs $ICECAP\text{-}A_{urban}$=75.7) and were well paid ($ICECAP\text{-}A_{\geq Chinese\ ¥6401\ (US\ \$960.15)}$=81.9 vs $ICECAP\text{-}A_{\leq Chinese\ ¥1800\ (US\ \$270)}$=73.5) were highly likely to report better capability well-being.

Table 1. Patients' characteristics and scores of the eHEALS[a], SSDM[b], and ICECAP-A[c].

	Patients, n (%)	eHEALS		SSDM		ICECAP-A	
		Value, mean (SD)	*P* value[d]	Value, mean (SD)	*P* value	Value, mean (SD)	*P* value
Overall	569 (100)	66.4 (21.2)	—[e]	85.7 (17.0)	—	77.5 (15.8)	—
Sex			.02		.55		.14
Female	281 (49.4)	64.3 (21.4)		85.3 (17.5)		79.8 (16.2)	
Male	288 (50.6)	68.5 (21)		86.1 (16.6)		83.6 (15.3)	
Age (years)			.003		.03		.85
16-30	106 (18.6)	71.7 (19.9)		84.2 (18.8)		77.9 (15)	
31-40	132 (23.2)	69.6 (18.2)		84.1 (16.3)		78.1 (15.7)	
41-50	116 (20.4)	66.0 (19.4)		84.4 (17.5)		78.4 (15.3)	
51-60	106 (18.6)	63.7 (22.4)		87.5 (18.3)		76.8 (15.4)	
≥61	109 (19.2)	60.5 (24.7)		88.6 (13.9)		76.3 (17.6)	
Education			.01		.40		<.001
No or primary	90 (15.8)	62.9 (25.2)		87.8 (15)		70.9 (19.7)	
Secondary	215 (37.8)	64.0 (22.1)		84.8 (18.4)		75.6 (16.1)	
Tertiary or above	264 (46.4)	69.6 (18.5)		85.7 (16.5)		81.3 (12.9)	
Marital status			.03		.70		.92
Single	95 (16.7)	72.5 (17.9)		84.5 (17.5)		78.7 (13.6)	
Married	446 (78.4)	65.1 (21.8)		85.8 (17)		77.3 (15.9)	
Divorced, widow, or widower	28 (4.9)	67.1 (19.9)		87.0 (16.3)		75.8 (20.3)	
Family registry			.86		.35		.03
Rural	279 (49.1)	66.5 (20.3)		86.5 (16.3)		79.3 (14.1)	
urban	290 (50.9)	66.3 (22.2)		84.8 (17.7)		75.7 (17.2)	
Number of children			.02		.52		.05
0	104 (18.3)	72.7 (17.8)		84 (18.5)		78.9 (14.3)	
1	170 (29.9)	66.4 (21.3)		86.7 (15)		79.5 (15.3)	
2	202 (35.5)	66.5 (20.4)		85.4 (16.4)		75.7 (16.3)	
≥3	93 (16.3)	59.4 (24.3)		86.4 (20.0)		76.1 (16.7)	
Caregiver			.42		.31		.48
No	414 (72.8)	65.8 (21.7)		85.4 (17.1)		77.1 (16.2)	
Yes	155 (27.2)	68 (20.0)		86.4 (16.9)		78.5 (14.7)	
Living status			.86		.41		.98
Live with family or others	512 (89.9)	66.4 (21.4)		85.6 (16.9)		77.4 (16)	
Live alone	57 (10.1)	66.6 (20.4)		86.2 (18.6)		78.3 (13.8)	
Employment status			.005		.11		.29
Employed	394 (69.2)	68.5 (20.2)		85.1 (17.1)		78.1 (15.1)	
Nonemployed	175 (30.8)	61.7 (22.8)		87.1 (16.9)		76.1 (17.3)	
Disposable income per month (Chinese ¥ [US $])			.55		.62		<.001
≤1800 (270)	155 (27.2)	64.7 (22.6)		84.9 (18.4)		73.5 (18.1)	
1801-3800 (270.15-570)	146 (25.7)	65.5 (22.9)		87.4 (15.4)		76.6 (15.8)	

	Patients, n (%)	eHEALS		SSDM		ICECAP-A	
		Value, mean (SD)	*P* value[d]	Value, mean (SD)	*P* value	Value, mean (SD)	*P* value
3801-6400 (570.15-960)	127 (22.3)	68.7 (18.2)		85.5 (16.3)		78.6 (14.0)	
≥6401 (960.15)	141 (24.8)	67.3 (20.3)		84.9 (17.7)		81.9 (13.4)	
Health insurance			.97		.98		.93
FHS[f]	30 (5.3)	68.3 (18.9)		87.3 (14.9)		81.1 (11.7)	
UEBMI[g]	258 (45.3)	67 (20.3)		85.3 (17.2)		78.3 (14.3)	
URBMI[h]	132 (23.2)	65.6 (21.8)		85.8 (18.4)		75.8 (18.7)	
NRCMS[i]	131 (23)	65.7 (22.9)		85.7 (16.4)		76.9 (16.0)	
No	18 (3.2)	67.2 (23.4)		87.7 (12.5)		77.2 (16.6)	
BMI[j]			.79		.03		.03
Normal	242 (42.5)	66.5 (21.9)		83.7 (18.4)		75.7 (17.1)	
Abnormal	327 (57.5)	66.3 (20.8)		87.1 (15.8)		78.9 (14.6)	
Smoking			.04		.17		.71
No	408 (71.7)	70.2 (21.0)		84.9 (17.3)		77.7 (15.5)	
Sometimes	83 (14.5)	69 (21.2)		85.6 (12.0)		76.9 (18.3)	
Everyday	78 (13.8)	65.2 (21.9)		89.7 (19.4)		76.9 (14.3)	
Healthy diet per week			.20		.49		.23
Few	87 (15.3)	65.4 (21.1)		84.8 (20.2)		74.4 (18.0)	
Sometimes	310 (54.5)	66.4 (20.8)		85.1 (16.9)		77.8 (15.0)	
Everyday	172 (30.2)	70.2 (21.9)		87.2 (15.5)		78.6 (15.1)	
Exercise per week			.009		.28		.004
Never	151 (26.5)	63.2 (23.7)		84.3 (17.8)		73.2 (19.0)	
Sometimes	321 (56.4)	68.8 (20.2)		86.1 (16.3)		78.6 (14.0)	
Always	97 (17.1)	63.5 (19.6)		86.5 (18.2)		80.7 (14.6)	
Chronic condition			.008		.56		.003
No	278 (48.9)	68.5 (20.9)		85 (17.6)		79.3 (15.5)	
Yes	291 (51.1)	64.4 (21.4)		86.3 (16.4)		75.7 (15.9)	
Depressive disorder			.04		.07		<.001
No	410 (72.1)	67.6 (20.9)		86.3 (16.8)		80.6 (13.8)	
Yes	159 (27.9)	63.3 (21.9)		84 (17.5)		69.4 (17.7)	
Self-reported health condition			.50		.11		.08
Severe threat to life	113 (19.9)	66.6 (22.1)		88 (16.3)		73.9 (17.3)	
Moderate threat to life	113 (19.9)	64 (22.1)		83.6 (18.4)		77.7 (15.4)	
Mild threat to life	136 (23.9)	65.6 (19.4)		85.6 (15.5)		78.1 (14.2)	

	Patients, n (%)	eHEALS		SSDM		ICECAP-A	
		Value, mean (SD)	*P* value[d]	Value, mean (SD)	*P* value	Value, mean (SD)	*P* value
No threat to life	207 (36.3)	68.2 (21.4)		85.6 (17.5)		78.9 (16.0)	

[a]eHEALS: eHealth Literacy Scale.

[b]SSDM: satisfaction with shared decision-making.

[c]ICECAP-A: Investigating Choice Experiments Capability Measure for Adults.

[d]*P* value was calculated based on a Kruskal-Wallis one-way analysis of variance (multiple groups) and Wilcoxon signed-rank test (2 groups).

[e]Not available.

[f]FHS: free health care scheme.

[g]UEBMI: urban employee basic medical insurance.

[h]URBMI: urban resident basic medical insurance.

[i]NRCMS: new rural cooperative medical care system.

[j]BMI: normal: 18.5≤BMI<23; abnormal: BMI<18.5 or BMI≥23.

Relationship Among eHealth Literacy, SSDM, and Well-being

Table 2 presents the outcomes of SSDM and well-being in patients with different levels of eHealth literacy. Patients with a higher level of eHealth literacy reported higher SSDM and better well-being than those with a lower level of eHealth literacy. For patients with depressive disorder, the difference in SSDM in patients with different levels of eHealth literacy was statistically nonsignificant.

Table 2. Satisfaction with SDM[a] and well-being in different groups of eHealth literacy and stratified by patients' depressive disorder and chronic condition status.

	Satisfaction with SDM	Well-being
Overall		
High eHealth literacy, mean (SD)	88.7 (14.7)	81.1 (14.7)
Low eHealth literacy, mean (SD)	82.4 (18.7)	73.7 (16.1)
P value[b]	<.001	<.001
With depressive disorder		
High eHealth literacy, mean (SD)	86.2(16.2)	73.3(18)
Low eHealth literacy, mean (SD)	82.2 (18.4)	66.1 (16.8)
P value	.10	.004
Without depressive disorder		
High eHealth literacy, mean (SD)	89.5 (14.2)	83.6 (12.4)
Low eHealth literacy, mean (SD)	82.5 (18.9)	77.1 (14.6)
P value	<.001	<.001
With chronic conditions		
High eHealth literacy, mean (SD)	88.4 (14.7)	82.6 (14.4)
Low eHealth literacy, mean (SD)	80.5 (20.1)	75 (15.8)
P value	<.001	<.001
Without chronic conditions		
High eHealth literacy, mean (SD)	89 (14.8)	79.1 (14.8)
Low eHealth literacy, mean (SD)	83.9 (17.5)	72.7 (16.4)
P value	<.001	<.001

[a]SDM: shared decision-making.

[b]*P* value was calculated based on Wilcoxon signed-rank test.

Results of the Regression Analysis

The results of multivariate regression models showed that there was a significant and positive relationship between eHealth literacy and SSDM and well-being after adjusting for patients' background characteristics (Table 3). Model 1 demonstrated that patients who were living alone (β=−6.82; *P*=.03) and nonemployed (β=−4.55; *P*=.02) showed a lower level of eHealth literacy. Models 2 and 3 showed that, after adjustment, eHealth literacy was a statistically significant factor predicting the change in SSDM (β=.17; *P*<.001) and well-being (β=.15; *P*<.001), respectively. There was a positive relationship between patients' well-being and their educational level, income, and depressive status.

Table 3. Regression analysis of eHealth literacy and satisfaction with shared decision-making (SSDM) and well-being[a].

Variables	β (95% CI)					
	Model 1 (DV[b]=eHEALS[c])	*P* value	Model 2 (DV=SSDM)	*P* value	Model 3 (DV=ICECAP-A[d])	*P* value
eHealth literacy	__[e]	—	.17 (0.11 to 0.24)	<.001	.15 (0.09 to 0.21)	<.001
Satisfaction in SDM	.22 (0.12 to 0.32)	<.001	—	—	—	—
Well-being	.26 (0.14 to 0.38)	<.001	—	—	—	—
Sex (male)	2.99 (−1.11 to 7.1)	.15	−1.58 (−5.03 to 1.87)	.37	.29 (−2.67 to 3.25)	.85
Age (years)						
31-40	−2.05 (−8.87 to 4.78)	.56	.68 (−5.03 to 6.4)	.81	2.58 (−2.32 to 7.49)	.30
41-50	−4.96 (−12.24 to 2.33)	.18	.58 (−5.52 to 6.67)	.85	4.98 (−0.25 to 10.22)	.06
51-60	−6.37 (−14.15 to 1.41)	.11	4.46 (−2.05 to 10.97)	.18	5.2 (−0.39 to 10.79)	.07
≥61	−7.91 (−16.35 to 0.53)	.07	5.97 (−1.09 to 13.03)	.1	5.88 (−0.18 to 11.94)	.06
Education						
Secondary	−2.78 (−8.35 to 2.78)	.33	−2.67 (−7.31 to 1.97)	.26	4.51 (0.53 to 8.49)	.03
Tertiary or above	−.7 (−7.23 to 5.83)	.83	−.91 (−6.33 to 4.5)	.74	8.34 (3.7 to 12.99)	.004
Marital status						
Married	−2.63 (−12.44 to 7.17)	.60	−2.26 (−10.48 to 5.97)	.59	−.35 (−7.41 to 6.71)	.92
Divorced, widow, or widower	3.57 (−8.77 to 15.91)	.57	−4.92 (−15.26 to 5.42)	.35	.94 (−7.94 to 9.81)	.84
Urban resident	−1.8 (−6.18 to 2.59)	.42	−.3 (−3.98 to 3.38)	.87	−.07 (−3.23 to 3.09)	.97
Children						
1	−2.84 (−13.15 to 7.47)	.59	4.02 (−4.62 to 12.66)	.36	.01 (−7.41 to 7.43)	.99
2	−1.06 (−11.56 to 9.43)	.84	2.77 (−6.03 to 11.56)	.54	−2.25 (−9.79 to 5.3)	.56
≥3	−6.89 (−18.22 to 4.43)	.23	2.87 (−6.63 to 12.38)	.55	1.29 (−6.87 to 9.45)	.76
Caregiver (yes)	.1 (−3.8 to 4.16)	.93	1.33 (−2 to 4.67)	.43	−.04 (−2.9 to 2.82)	.98
Live alone	−6.82 (−13.28 to −0.36)	.04	2.91 (−2.52 to 8.34)	.29	1.35 (−3.31 to 6.02)	.57
Nonemployed	−4.55 (−9.02 to −0.08)	.04	1.43 (−2.33 to 5.19)	.45	.35 (−2.88 to 3.57)	.83
Income (Chinese ¥ [US $])						
1801-3800 (270.15-570)	−1.34 (−6.23 to 3.54)	.59	2.05 (−2.04 to 6.15)	.32	1.94 (−1.58 to 5.45)	.28
3801-6400 (570.15-960)	−1.32 (−6.9 to 4.25)	.64	1.13 (−3.54 to 5.8)	.63	2.43 (−1.58 to 6.44)	.23
≥6401 (960.15)	−4.74 (−10.69 to 1.21)	.12	.47 (−4.51 to 5.45)	.85	4.84 (0.56 to 9.11)	.03
Insurance						
Urban employee basic medical insurance	−3.03 (−10.82 to 4.76)	.45	−2.17 (−8.7 to 4.35)	.51	−2.48 (−8.08 to 3.13)	.39
Urban resident basic medical insurance	−2.21 (−10.47 to 6.06)	.60	−1.53 (−8.47 to 5.4)	.66	−2.07 (−8.02 to 3.88)	.50
New rural cooperative medical care system	−2.12 (−10.91 to 6.67)	.64	−1.91 (−9.28 to 5.46)	.61	−.02 (−6.35 to 6.31)	.99
No	−4.95 (−17.27 to 7.38)	.43	.66 (−9.68 to 11)	.90	2.05 (−6.82 to 10.93)	.65
BMI (abnormal)	−2.29 (−5.78 to 1.19)	.20	3.17 (0.26 to 6.08)	.03	2.38 (−0.12 to 4.87)	.06
Smoking status						
Sometimes	2.56 (−2.83 to 7.96)	.35	6.04 (1.54 to 10.54)	.01	.01 (−3.85 to 3.87)	.99
Everyday	2.84 (−2.8 to 8.48)	.32	2.13 (−2.6 to 6.86)	.38	−1.49 (−5.55 to 2.57)	.47
Healthy diet						
Sometimes	−6.28 (−11.19 to −1.38)	.01	.79 (−3.35 to 4.92)	.71	2.37 (−1.18 to 5.91)	.19

Variables	β (95% CI)					
	Model 1 (DV[b]=eHEALS[c])	*P* value	Model 2 (DV=SSDM)	*P* value	Model 3 (DV=ICECAP-A[d])	*P* value
Everyday	−2.45 (−7.92 to 3.01)	.38	2.15 (−2.42 to 6.73)	.36	2.61 (−1.31 to 6.54)	.19
Exercise						
Sometimes	4.13 (0.02 to 8.24)	.04	1.43 (−2.01 to 4.87)	.41	3.7 (0.74 to 6.65)	.01
Always	−1.94 (−7.36 to 3.47)	.48	1.89 (−2.63 to 6.4)	.41	4.9 (1.03 to 8.77)	.01
Chronic condition (yes)	.07 (−3.6 to 3.75)	.97	1.12 (−1.95 to 4.19)	.47	−2.5 (−5.14 to 0.13)	.06
Depressive disorder (yes)	−1.6 (−5.62 to 2.42)	.43	−1.94 (−5.2 to 1.32)	.24	−8.55 (−11.35 to −5.75)	<.001
Moderate threat to life	−2.84 (−8.24 to 2.55)	.30	−3.41 (−7.92 to 1.1)	.14	2.71 (−1.16 to 6.58)	.17
Mild threat to life	−2.62 (−7.93 to 2.69)	.33	−1.76 (−6.21 to 2.69)	.44	.64 (−3.18 to 4.47)	.74
No threat to life	−.62 (−5.66 to 4.42)	.81	−1.34 (−5.57 to 2.89)	.53	.1 (−3.53 to 3.73)	.96

[a]Reference: female, 16-30 years, no or primary education, single, rural resident, no child, no caregiver, live with family or others, income ≤ Chinese ¥1800 (US $270), free health care scheme insurance, normal BMI, no smoking, few healthy diets, no exercise, no chronic conditions, no depressive disorder, and severe threat to life.

[b]DV: dependent variable.

[c]eHEALS: eHealth Literacy Scale.

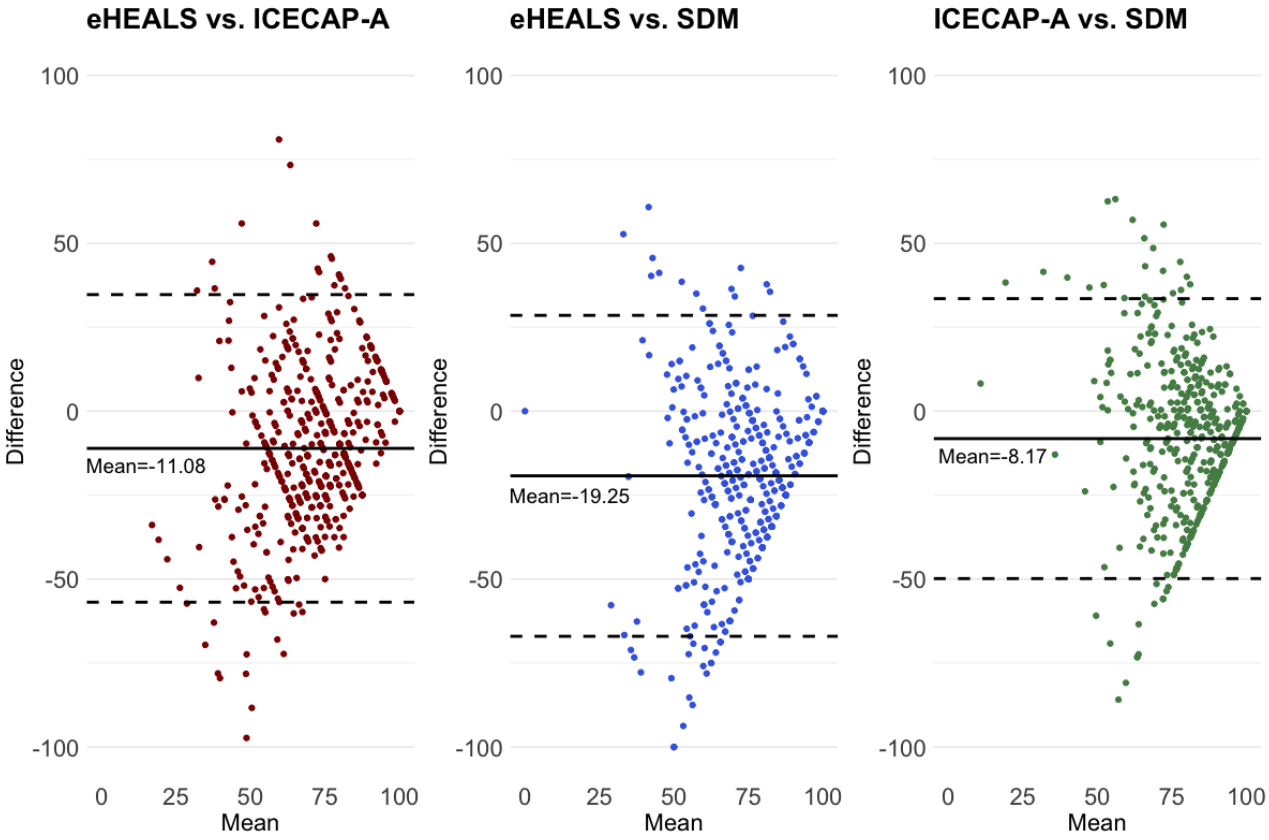
[d]ICECAP-A: Investigating Choice Experiments Capability Measure for Adults.

[e]Not available.

Agreement Between Measures

Although the B-A plot shows a wide limit of agreement interval between the three measures, systematic differences were detected. A good agreement was observed in patients who reported a high level of eHealth literacy, SSDM, and well-being; however, patients who reported a low level of eHealth literacy, SSDM, and well-being were more likely to show less consistent results across the measures, indicating low agreement (Figure 1).

Figure 1. Agreement between scores of the eHEALS, satisfaction with SDM, and ICECAP-A. eHEALS: eHealth Literacy Scale; ICECAP-A: Investigating Choice Experiments Capability Measure for Adults; SDM: shared decision-making.



Discussion

Principal Findings

This study extended the findings of previous studies by demonstrating a statistically significant association between eHealth literacy and SSDM and capability well-being in a sample of Chinese patients. However, when patients reported a low level of eHealth literacy, its association with SSDM and well-being turned to weak and inconsistent. Our findings suggested that providing training to improve patients' eHealth literacy may be a useful way to strengthen their ability to search and use web-based health and health care information to improve their activity in clinical decision-making and well-being. However, although the internet carries a vast range of information resources and services to help people manage their health, we noticed that disparities in using the internet are persistent in people with low SES (unemployed status and unhealthy lifestyle) and, therefore, affect their potential to maintain and improve eHealth literacy and limit their ability to navigate the health care system. In addition, there seemed to be a negative relationship between patients' mental health status and their use of internet-based knowledge and skills to improve SSDM. However, further research is needed to support this finding, as it has not been studied extensively.

Comparisons With Previous Studies

Our results firstly exhibited that there is a positive relationship between eHealth literacy and Chinese patients' SSDM, which is in line with the findings of previous studies. For example, an Iranian study indicated that eHealth literacy is positively associated with SDM and patient communication patterns in patients with multiple myeloma [20]. Another German study noted that the regular use of eHealth services facilitated the decision-making process for patients with cancer and their families [40]. Other studies have reported that the internet and web-based courses are fundamental in improving patients' communication skills with medical personnel [3,41], reducing their overall medical expenses [42], and increasing their confidence and knowledge to be involved in decision-making [43]. In addition, we found that patients who reported being very satisfied with SDM (satisfied with all five dimensions of the SSDM) obtained a similar mean score on the eHEALS (Multimedia Appendix 3). This indicated that improving eHealth literacy might be a multifaceted strategy to promote all the existential dimensions of the SDM [44] and, in turn, positively associated with patients' physical, psychological, social, and spiritual well-being [45]. Considering that the internet is increasingly serving as a major source of health information for both medical professionals and patients, improving eHealth literacy could be a cost-effective way to transfer the current paternalistic pattern of medical services to a more patient-centric model [2]. Moreover, doctors' attitude toward SDM was identified as another important factor that supported SDM in improving patients' well-being in clinical practice [46], which should be considered for future studies.

Studies examining the relationship between patients' well-being and eHealth-related interventions have recently been explored. For example, Villani et al [47] indicated that eHealth interventions can significantly reduce emotional suppression of patients with cancer and increase their cancer-related emotional well-being. Pagliari [48] and Patel et al [49] found that the proliferation of personal health information technology has enhanced people's ability to manage their own health care, communicate with providers through social networks, meet their informational needs, search patient educational resources, and make preferred health decisions. However, research examining the role of eHealth literacy in using these techniques to improve patients' well-being is in its infancy. This study found that patients who reported a high level of eHealth literacy were more likely to show full capability well-being than those who reported a low level of eHealth literacy. This supports the notion that patients with high levels of eHealth literacy are confident and capable of handling internet-based tools to improve their health and well-being [21,22,50]. Further, unlike studies that used nonpreference-based instruments to measure well-being, in this study, the ICECAP-A, a preference-based measure, generated outcomes that not only reflect a patient's current well-being status but also provide information to support the estimation of social care–related quality of life and facilitate a cost-effectiveness analysis of eHealth-related interventions and policies [33]. Furthermore, the results of the B-A plot exhibited a poor agreement between eHEALS, SSDM, and ICECAP-A when patients reported a low score on those measures, unlike those who reported a high score. Methodologically, this may be because, in this study, few patients reported having low well-being and unsatisfactory SDM, and less than 19.8% (113/569) of the respondents reported having poor health. Thus, we could not validate our findings in these populations. Due to the cross-sectional design, no causal relationships can be concluded. Therefore, further studies are required.

The results of bivariate analysis indicated that patients with high SES and healthy lifestyle are more likely to indicate a high level of eHealth literacy; however, the multivariable regression analysis showed a different picture. This is consistent with the mixed findings of the relationship between individuals' socioeconomic determinants and their level of eHealth literacy, as reported in previous studies. For example, Lwin et al [51] found that women and men did not differ in their reported frequencies of evaluating eHealth information; however, no older adult respondents (>55 years) were involved in their study. Conversely, a study in China found that female respondents showed a higher level of eHealth literacy than male respondents; however, they used a revised version of the eHEALS, and all the respondents were older than 45 years [52]. In addition, Stellefson et al [53] indicated that women and older adults living with chronic obstructive pulmonary disease showed a low level of eHealth literacy in the United States. Wong and Cheung [54] also indicated that older primary care service users in Hong Kong are highly likely to report low levels of eHealth literacy. The findings of multivariable regression analysis showed that, when adjusted for patients' background characteristics, there was no significant relationship between patients' eHealth literacy and their educational level. This suggests that eHealth literacy does not comprise only basic literacy but also an accumulation of knowledge and skills to navigate the internet to use health care services. Norman [31] noted that eHealth literacy is not

just a combination of the capability to use computers and traditional health literacy but is a meta-literacy comprising different facets of literacy. For example, an undereducated patient with chronic conditions may show higher eHealth literacy than highly educated patients who have recently been diagnosed with cancer. Our findings highlight the importance of promoting eHealth literacy.

When patients reported having depressive disorders, the difference in SSDM between those with high and low eHealth literacy was statistically insignificant. This is not inconsistent with previous findings in those patients with good skills in searching, assessing, and correctly using web-based health care information may lead to decreased levels of hospitalization-related mental disorders and improve their long-term quality of life and well-being [55-57]. However, no study has directly investigated the relationship between eHealth literacy and SDM considering the potential effect of patients' mental health status. Neter and Brainin [11] confirmed that there is insufficient evidence on the association between eHealth literacy and emotional states of anxiety and depression. A US-based study indicated that social media could benefit patients with chronic obstructive pulmonary disease by helping them cope with mental health issues such as anxiety and depression [53]. Another study found that African Americans who researched depression, anxiety, and stress on the internet showed a significantly higher mean score of the eHEALS than those who did not [58]. Interventions focused on internet-related health literacy, such as mental eHealth literacy, require further investigation. Furthermore, in this study, depressive status was assessed using the Patient Health Questionnaire-2, the sensitivity of which has been disputed by some previous studies [59,60]. Other mental disorders, such as anxiety and stress, should be evaluated in future studies.

Limitations

It is important to address the limitations of this study. First, this was a cross-sectional study; thus, no causal relationships could be concluded. Second, all the respondents were recruited from inpatient departments in hospitals; the issue of a single information source may affect the validity of our findings. In addition, compared with the data from the 2019 Guangdong census, our respondents were slightly older and comprised a higher proportion of rural residents (Multimedia Appendix 4). This implies that there was some degree of selection bias, which may affect the generalizability of our findings. Third, we did not assess the associations between eHealth literacy, well-being, and SDM stratified by patient disease groups, which might also affect the generalizability of our findings. Fourth, the ICECAP-A score was not estimated based on the preference weight of the Chinese population, which is currently unavailable. This may have affected the validity of our findings. Fifth, all the information was self-reported by the patients, which may have generated recall bias. Finally, the information of patients who refused to participate in the survey was not recorded, which might have led to a degree of information bias.

Conclusions

According to the findings of this study, patients with a high level of eHealth literacy were more likely to experience an optimal SDM and improved capability well-being. This suggests that the implementation of interventions to strengthen patients' eHealth literacy could improve their optimal use of health care services and the efficiency of the health and social care system. In addition, univariable analysis demonstrated that patients with low SES showed insufficient eHealth literacy, which may affect their ability to buffer against the negative impacts of an adverse event on their health. It is important for policy makers to understand the facilitators and barriers to improve patients' eHealth literacy and to develop strategies to enhance their health behaviors and health outcomes. Moreover, the effects of patients' mental health status on the relationship between eHealth literacy and SSDM require further investigation.

Acknowledgments

This study was funded by a grant from the Guangdong Basic and Applied Basic Research Foundation (Ref ID: 2021A1515011973).

Authors' Contributions

RHX was involved in the study conceptualization and design, data analysis and interpretation, software use, the writing of the original draft, review, and editing. LMZ was involved in software use, visualization, writing, review, and editing. ELYW was involved in study concept and design, supervision, writing, review, and editing. DW was involved in the study concept and design process, the provision of study materials or patients, the collection and assembly of data, supervision, writing, review, and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Results of the confirmatory factor analysis of the satisfaction with shared decision-making.
[DOCX File , 13 KB - jmir_v23i9e26721_app1.docx]

Multimedia Appendix 2
Correlation between the satisfaction with shared decision-making and 9-item Shared Decision-Making Questionnaire.

[DOCX File , 13 KB - jmir_v23i9e26721_app2.docx]

Multimedia Appendix 3
eHEALS scores stratified by items of the ICECAP-A and satisfaction with SDM. eHEALS: eHealth Literacy Scale; ICECAP-A: Investigating Choice Experiments Capability Measure for Adults; SDM: shared decision-making. *<italic>P</italic><.05; **<italic>P</italic><.01; ***<italic>P</italic><.001
[PNG File , 99 KB - jmir_v23i9e26721_app3.png]

Multimedia Appendix 4
Comparisons between the sample and Guangdong general population.
[DOCX File , 20 KB - jmir_v23i9e26721_app4.docx]

References

1. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
2. Paige SR, Krieger JL, Stellefson M, Alber JM. eHealth literacy in chronic disease patients: an item response theory analysis of the eHealth literacy scale (eHEALS). Patient Educ Couns 2017 Feb:320-326. [doi: 10.1016/j.pec.2016.09.008] [Medline: 27658660]
3. Silver MP. Patient perspectives on online health information and communication with doctors: a qualitative study of patients 50 years old and over. J Med Internet Res 2015 Jan 13;17(1):e19 [FREE Full text] [doi: 10.2196/jmir.3588] [Medline: 25586865]
4. Karnoe A, Furstrand D, Christensen KB, Norgaard O, Kayser L. Assessing competencies needed to engage with digital health services: development of the eHealth literacy assessment toolkit. J Med Internet Res 2018 May 10;20(5):e178 [FREE Full text] [doi: 10.2196/jmir.8347] [Medline: 29748163]
5. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 09;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
6. Parker S, Prince A, Thomas L, Song H, Milosevic D, Harris MF, IMPACT Study Group. Electronic, mobile and telehealth tools for vulnerable patients with chronic disease: a systematic review and realist synthesis. BMJ Open 2018 Aug 29;8(8):e019192 [FREE Full text] [doi: 10.1136/bmjopen-2017-019192] [Medline: 30158214]
7. Quinn S, Bond R, Nugent C. Quantifying health literacy and eHealth literacy using existing instruments and browser-based software for tracking online health information seeking behavior. Comput Hum Behav 2017 Apr;69:256-267. [doi: 10.1016/j.chb.2016.12.032]
8. George C, Whitehouse D, Duquenoy P. eHealth: Legal, Ethical and Governance Challenges. Berlin, Germany: Springer Science & Business Media; 2012.
9. Hansen AH, Bradway M, Broz J, Claudi T, Henriksen Ø, Wangberg SC, et al. Inequalities in the use of eHealth between socioeconomic groups among patients with type 1 and type 2 diabetes: cross-sectional study. J Med Internet Res 2019 May 29;21(5):e13615 [FREE Full text] [doi: 10.2196/13615] [Medline: 31144669]
10. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017 Apr 27;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
11. Neter E, Brainin E. Association between health literacy, eHealth literacy, and health outcomes among patients with long-term conditions: a systematic review. Eur Psychol 2019 Jan;24(1):68-81. [doi: 10.1027/1016-9040/a000350]
12. Xie B. Effects of an eHealth literacy intervention for older adults. J Med Internet Res 2011 Nov 3;13(4):e90 [FREE Full text] [doi: 10.2196/jmir.1880] [Medline: 22052161]
13. Hsu W, Chiang C, Yang S. The effect of individual factors on health behaviors among college students: the mediating effects of eHealth literacy. J Med Internet Res 2014 Dec 12;16(12):e287 [FREE Full text] [doi: 10.2196/jmir.3542] [Medline: 25499086]
14. Lin C, Ganji M, Griffiths MD, Bravell ME, Broström A, Pakpour AH. Mediated effects of insomnia, psychological distress and medication adherence in the association of eHealth literacy and cardiac events among Iranian older patients with heart failure: a longitudinal study. Eur J Cardiovasc Nurs 2020 Feb;19(2):155-164. [doi: 10.1177/1474515119873648] [Medline: 31516036]
15. Nabi RL, Prestin A, So J. Facebook friends with (health) benefits? Exploring social network site use and perceptions of social support, stress, and well-being. Cyberpsychol Behav Soc Netw 2013 Oct;16(10):721-727. [doi: 10.1089/cyber.2012.0521] [Medline: 23790356]
16. Triberti S, Savioni L, Sebri V, Pravettoni G. eHealth for improving quality of life in breast cancer patients: A systematic review. Cancer Treat Rev 2019 Mar;74:1-14. [doi: 10.1016/j.ctrv.2019.01.003] [Medline: 30658289]

17. Koay K, Schofield P, Jefford M. Importance of health literacy in oncology. Asia Pac J Clin Oncol 2012 Mar;8(1):14-23. [doi: 10.1111/j.1743-7563.2012.01522.x] [Medline: 22369440]
18. Brehaut JC, O'Connor AM, Wood TJ, Hack TF, Siminoff L, Gordon E, et al. Validation of a decision regret scale. Med Decis Making 2003;23(4):281-292. [doi: 10.1177/0272989X03256005] [Medline: 12926578]
19. Vahdat S, Hamzehgardeshi L, Hessam S, Hamzehgardeshi Z. Patient involvement in health care decision making: a review. Iran Red Crescent Med J 2014 Jan;16(1):e12454 [FREE Full text] [doi: 10.5812/ircmj.12454] [Medline: 24719703]
20. Nejati B, Lin C, Aaronson NK, Cheng AS, Browall M, Lin C, et al. Determinants of satisfactory patient communication and shared decision making in patients with multiple myeloma. Psychooncology 2019 Jul;28(7):1490-1497. [doi: 10.1002/pon.5105] [Medline: 31087365]
21. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015 Mar 17;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
22. Bodie GD, Dutta MJ. Understanding health literacy for strategic health marketing: eHealth literacy, health disparities, and the digital divide. Health Mark Q 2008;25(1-2):175-203. [doi: 10.1080/07359680802126301] [Medline: 18935884]
23. Car J, Tan WS, Huang Z, Sloot P, Franklin BD. eHealth in the future of medications management: personalisation, monitoring and adherence. BMC Med 2017 Apr 05;15(1):73 [FREE Full text] [doi: 10.1186/s12916-017-0838-0] [Medline: 28376771]
24. Number of internet users in China from 2008 to 2020. Statista. 2021. URL: https://www.statista.com/statistics/265140/number-of-internet-users-in-china/ [accessed 2021-04-23]
25. Guideline on the development and promotion of providing internet-based healthcare services in China. URL: http://www.gov.cn/zhengce/content/2018-04/28/content_5286645.htm [accessed 2020-05-07]
26. Brennan PF, Strombom I. Improving health care by understanding patient preferences: the role of computer technology. J Am Med Inform Assoc 1998;5(3):257-262 [FREE Full text] [doi: 10.1136/jamia.1998.0050257] [Medline: 9609495]
27. Guideline on the development and promotion of providing internet-based healthcare services in China. 2018. URL: http://www.gov.cn/zhengce/content/2018-04/28/content_5286645.htm [accessed 2020-05-07]
28. Brørs G, Norman CD, Norekvål TM. Accelerated importance of eHealth literacy in the COVID-19 outbreak and beyond. Eur J Cardiovasc Nurs 2020 Aug;19(6):458-461 [FREE Full text] [doi: 10.1177/1474515120941307] [Medline: 32667217]
29. Dong D, Xu RH, Wong EL, Hung C, Feng D, Feng Z, et al. Public preference for COVID-19 vaccines in China: a discrete choice experiment. Health Expect 2020 Dec;23(6):1543-1578 [FREE Full text] [doi: 10.1111/hex.13140] [Medline: 33022806]
30. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
31. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
32. Xu RH, Zhou L, Lu SY, Wong EL, Chang J, Wang D. Psychometric validation and cultural adaptation of the simplified Chinese eHealth literacy scale: cross-sectional study. J Med Internet Res 2020 Dec 07;22(12):e18613 [FREE Full text] [doi: 10.2196/18613] [Medline: 33284123]
33. Al-Janabi H, Flynn TN, Coast J. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. Qual Life Res 2012 Feb;21(1):167-176 [FREE Full text] [doi: 10.1007/s11136-011-9927-2] [Medline: 21598064]
34. Flynn TN, Huynh E, Peters TJ, Al-Janabi H, Clemens S, Moody A, et al. Scoring the Icecap-a capability instrument. Estimation of a UK general population tariff. Health Econ 2015 Mar;24(3):258-269 [FREE Full text] [doi: 10.1002/hec.3014] [Medline: 24254584]
35. Tang C, Xiong Y, Wu H, Xu J. Adaptation and assessments of the Chinese version of the ICECAP-A measurement. Health Qual Life Outcomes 2018 Mar 12;16(1):45 [FREE Full text] [doi: 10.1186/s12955-018-0865-3] [Medline: 29530092]
36. Wong EL, Xu RH, Lui S, Cheung AW, Yeoh E. Development of conceptual framework from the view of patients and professionals on patient engagement: a qualitative study in Hong Kong SAR, China. Open J Nurs 2018 May;08(05):303-316. [doi: 10.4236/ojn.2018.85026]
37. Xu RH, Cheung AW, Wong EL. Development and validation of an instrument to measure patient engagement in Hong Kong Special Administrative Region, China. Patient Prefer Adherence 2018 Sep 4;12:1667-1675 [FREE Full text] [doi: 10.2147/PPA.S171026] [Medline: 30233147]
38. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. J Gen Intern Med 2007 Nov;22(11):1596-1602 [FREE Full text] [doi: 10.1007/s11606-007-0333-y] [Medline: 17874169]
39. Giavarina D. Understanding Bland Altman analysis. Biochem Med (Zagreb) 2015 Jun 5;25(2):141-151 [FREE Full text] [doi: 10.11613/BM.2015.015] [Medline: 26110027]
40. Halwas N, Griebel L, Huebner J. eHealth literacy, Internet and eHealth service usage: a survey among cancer patients and their relatives. J Cancer Res Clin Oncol 2017 Nov;143(11):2291-2299. [doi: 10.1007/s00432-017-2475-6] [Medline: 28699035]

41. Lu X, Zhang R. Impact of physician-patient communication in online health communities on patient compliance: cross-sectional questionnaire study. J Med Internet Res 2019 May 13;21(5):e12891 [FREE Full text] [doi: 10.2196/12891] [Medline: 31094342]
42. Veroff D, Marr A, Wennberg DE. Enhanced support for shared decision making reduced costs of care for patients with preference-sensitive conditions. Health Aff (Millwood) 2013 Feb;32(2):285-293. [doi: 10.1377/hlthaff.2011.0941] [Medline: 23381521]
43. Hoffmann TC, Del Mar C, Santhirapala R, Freeman A. Teaching clinicians shared decision making and risk communication online: an evaluation study. BMJ Evid Based Med 2020 Sep 22 (forthcoming) [FREE Full text] [doi: 10.1136/bmjebm-2020-111521] [Medline: 32962972]
44. Gulbrandsen P, Clayman ML, Beach MC, Han PK, Boss EF, Ofstad EH, et al. Shared decision-making as an existential journey: aiming for restored autonomous capacity. Patient Educ Couns 2016 Sep;99(9):1505-1510. [doi: 10.1016/j.pec.2016.07.014] [Medline: 27460801]
45. Netter E, Brainin E, Baron-Epel O. The third digital divide in the health domain: is internet use for health purposes associated with health benefits? In: eHealth: Current Evidence, Promises, Perils and Future Directions. Bingley: Emerald Publishing Limited; 2018:75.
46. Stevenson FA, Barry CA, Britten N, Barber N, Bradley CP. Doctor-patient communication about drugs: the evidence for shared decision making. Soc Sci Med 2000 Mar;50(6):829-840. [doi: 10.1016/s0277-9536(99)00376-7] [Medline: 10695980]
47. Villani D, Cognetta C, Repetto C, Serino S, Toniolo D, Scanzi F, et al. Promoting emotional well-being in older breast cancer patients: results from an eHealth intervention. Front Psychol 2018 Nov 27;9:2279 [FREE Full text] [doi: 10.3389/fpsyg.2018.02279] [Medline: 30538650]
48. Pagliari C. Design and evaluation in eHealth: challenges and implications for an interdisciplinary field. J Med Internet Res 2007 May 27;9(2):e15 [FREE Full text] [doi: 10.2196/jmir.9.2.e15] [Medline: 17537718]
49. Patel V, Arocha J, Ancker J. Cognitive Informatics in Health and Biomedicine: Understanding and Modeling Health Behaviors. Cham: Springer; 2017.
50. Stellefson M, Paige SR, Alber JM, Chaney BH, Chaney D, Apperson A, et al. Association between health literacy, electronic health literacy, disease-specific knowledge, and health-related quality of life among adults with chronic obstructive pulmonary disease: cross-sectional study. J Med Internet Res 2019 Jun 06;21(6):e12165 [FREE Full text] [doi: 10.2196/12165] [Medline: 31172962]
51. Lwin MO, Panchapakesan C, Sheldenkar A, Calvert GA, Lim LK, Lu J. Determinants of eHealth literacy among adults in China. J Health Commun 2020 May 03;25(5):385-393. [doi: 10.1080/10810730.2020.1776422] [Medline: 32552607]
52. Lin Z, Zhang Y, Matteson M, Li X, Tu X, Zhou Y, et al. Older adults' eHealth literacy and the role libraries can play. J Libr Inf Sci 2020 Oct 21;53(3):488-498. [doi: 10.1177/0961000620962847]
53. Stellefson ML, Shuster JJ, Chaney BH, Paige SR, Alber JM, Chaney JD, et al. Web-based health information seeking and eHealth literacy among patients living with Chronic Obstructive Pulmonary Disease (COPD). Health Commun 2018 Dec;33(12):1410-1424 [FREE Full text] [doi: 10.1080/10410236.2017.1353868] [Medline: 28872905]
54. Wong DK, Cheung M. Online health information seeking and eHealth literacy among patients attending a primary care clinic in Hong Kong: a cross-sectional survey. J Med Internet Res 2019 Mar 27;21(3):e10831 [FREE Full text] [doi: 10.2196/10831] [Medline: 30916666]
55. Willis E, Royne MB. Online health communities and chronic disease self-management. Health Commun 2017 Mar;32(3):269-278. [doi: 10.1080/10410236.2016.1138278] [Medline: 27218836]
56. Duff OM, Walsh DM, Furlong BA, O'Connor NE, Moran KA, Woods CB. Behavior change techniques in physical activity eHealth interventions for people with cardiovascular disease: systematic review. J Med Internet Res 2017 Aug 02;19(8):e281 [FREE Full text] [doi: 10.2196/jmir.7782] [Medline: 28768610]
57. Wildevuur SE, Simonse LW. Information and communication technology-enabled person-centered care for the "big five" chronic conditions: scoping review. J Med Internet Res 2015 Mar 27;17(3):e77 [FREE Full text] [doi: 10.2196/jmir.3687] [Medline: 25831199]
58. James DC, Harville C. Ehealth literacy, online help-seeking behavior, and willingness to participate in mHealth chronic disease research among African Americans, Florida, 2014-2015. Prev Chronic Dis 2016 Dec 17;13:E156 [FREE Full text] [doi: 10.5888/pcd13.160210] [Medline: 27854421]
59. Bunevicius A, Deltuva V, Tamasauskas S, Tamasauskas A, Bunevicius R. Screening for psychological distress in neurosurgical brain tumor patients using the Patient Health Questionnaire-2. Psychooncology 2013 Aug;22(8):1895-1900. [doi: 10.1002/pon.3237] [Medline: 23233453]
60. Liu Z, Yu Y, Hu M, Liu H, Zhou L, Xiao S. PHQ-9 and PHQ-2 for screening depression in Chinese rural elderly. PLoS One 2016 Mar 15;11(3):e0151042 [FREE Full text] [doi: 10.1371/journal.pone.0151042] [Medline: 26978264]

Abbreviations

B-A: Bland-Altman
eHEALS: eHealth Literacy Scale

ICECAP-A: Investigating Choice Experiments Capability Measure for Adults
SDM: shared decision-making
SES: socioeconomic status
SSDM: satisfaction with shared decision-making

Edited by R Kukafka; submitted 31.12.20; peer-reviewed by E Neter, Y Yu; comments to author 19.01.21; revised version received 06.03.21; accepted 10.08.21; published 24.09.21.

Please cite as:
Xu RH, Zhou LM, Wong ELY, Wang D
The Association Between Patients' eHealth Literacy and Satisfaction With Shared Decision-making and Well-being: Multicenter Cross-sectional Study
J Med Internet Res 2021;23(9):e26721
URL: https://www.jmir.org/2021/9/e26721
doi:10.2196/26721
PMID:34559062

Original Paper

Versatile GCH Control Software for Correction of Loads Applied to Forearm Crutches During Gait Recovery Through Technological Feedback: Development and Implementation Study

Gema Chamorro-Moriana[1], PhD; Jose Luis Sevillano[2], PhD; V Perez-Cabezas[3], PhD

[1]Department of Physiotherapy, Area of Physiotherapy Research Group CTS-305, University of Seville, Seville, Spain

[2]Department of Architecture and Technology of Computers, Robotics and Technology of Computers Research Group TEP-108, University of Seville, Seville, Spain

[3]Department of Nursing and Physiotherapy, Empowering Health by Physical Activity, Exercise and Nutrition Research Group CTS-1038, University of Cadiz, Cadiz, Spain

Corresponding Author:
Gema Chamorro-Moriana, PhD
Department of Physiotherapy
Area of Physiotherapy Research Group CTS-305
University of Seville
Calle Avicena, S/N
Seville, 41009
Spain
Phone: 34 954486554
Email: gchamorro@us.es

Abstract

Background: Measuring weight bearing is an essential aspect of clinical care for lower limb injuries such as sprains or meniscopathy surgeries. This care often involves the use of forearm crutches for partial loads progressing to full loads. Therefore, feasible methods of load monitoring for daily clinical use are needed.

Objective: The main objective of this study was to design an innovative multifunctional desktop load-measuring software that complements GCH System 2.0–instrumented forearm crutches and monitors the applied loads, displaying real-time graphical and numerical information, and enabling the correction of inaccuracies through feedback technology during assisted gait. The secondary objective was to perform a preliminary implementation trial.

Methods: The software was designed for indoor use (clinics/laboratories). This software translates the crutch sensor signal in millivolts into force units, records and analyzes data (10-80 Hz), and provides real-time effective curves of the loads exerted on crutches. It covers numerous types of extrinsic feedback, including visual, acoustic (verbal/beeps), concurrent, terminal, and descriptive feedback, and includes a clinical and research use database. An observational descriptive pilot study was performed with 10 healthy subjects experienced in bilateral assisted gait. The Wilcoxon matched-pairs signed-rank test was used to evaluate the load accuracy evolution of each subject (ie, changes in the loads exerted on crutches for each support) among various walks, which was interpreted at the 95% confidence level.

Results: GCH Control Software was developed as a multifunctional desktop tool complementing GCH System 2.0–instrumented forearm crutches. The pilot implementation of the feedback mechanism observed 96/100 load errors at baseline (walk 0, no feedback) with 7/10 subjects exhibiting crutch overloading. Errors ranged from 61.09% to 203.98%, demonstrating heterogeneity. The double-bar feedback found 54/100 errors in walk 1, 28/100 in walk 2, and 14/100 in walk 3. The first walk with double-bar feedback (walk 1) began with errors similar to the baseline walk, generally followed by attempts at correction. The Wilcoxon matched-pairs signed-rank test used to evaluate each subject's progress showed that all participants steadily improved the accuracy of the loads applied to the crutches. In particular, Subject 9 required extra feedback with two single-bar walks to focus on the total load. The participants also corrected the load balance between crutches and fluency errors. Three subjects made one error of load balance and one subject made six fluctuation errors during the three double-bar walks. The latter subject performed additional feedback with two balance-bar walks to focus on the load balance.

Conclusions: GCH Control Software proved to be useful for monitoring the loads exerted on forearm crutches, providing a variety of feedback for correcting load accuracy, load balance between crutches, and fluency. The findings of the complementary

implementation were satisfactory, although clinical trials with larger samples are needed to assess the efficacy of the different feedback mechanisms and to select the best alternatives in each case.

(J Med Internet Res 2021;23(9):e27602) doi:10.2196/27602

KEYWORDS

control and monitoring software; feedback technology; motor control; gait; crutches; assisted gait for partial weight-bearing; functional recovery of the gait; unloading of lower limb musculoskeletal injury; rehabilitation; physical therapy; lower limb; injury; injuries; feedback technology; crutches

Introduction

Gait is a basic motor function of humans [1-3]; thus, many health (eg, physiotherapy, orthopedics, biomechanics) and engineering (eg, computing, mechanics, robotics) professionals, among others, have pooled their efforts to analyze and assess gait for recovery or optimization purposes.

New feedback-based technologies are especially useful in the field of rehabilitation [4] to reeducate an altered function or teach a new one, such as in the functional recovery of walking [5,6]. These aspects are fundamental objectives of physiotherapy [7,8], which enable subjects to voluntarily control and modify certain body functions or biological processes if they are given new information about these functions and processes. This is the basic principle underlying feedback mechanisms [7,9,10].

This paper focuses on *extrinsic feedback*, which is provided by external sources [11,12]. The wide range of well-known feedback technologies include visual [5,13-15], acoustic [16-18], and haptic [11,16,19] technologies, which are usually adapted to each individual user in a coherent manner [20]. Other classifications and types of feedback depend on when the information is applied, what type of data it offers, or what objective it pursues, among other factors. Some confusing terms in the literature are clarified below.

First, when feedback is given, it can be *concurrent* (ie, *simultaneously* provided during the intervention) or *terminal* (ie, retarded or postresponse feedback, provided when the action is finished). Concurrent feedback can be *continuous* or *intermittent*. Terminal feedback can be *immediate* [21], *delayed*, or even *summary* (provided after several repetitions of the movements or actions) [22].

Second, according to the type of information received, the feedback can be classified as *knowledge of performance* (KP) [23], which offers performance features (eg, if the subject keeps looking ahead while walking) and *knowledge of results* (KR) [24,25], which involves judgment of a correct or wrong action, or calculation of scores using rating scales. Feedback can also be *descriptive* or *prescriptive*, providing information about how to correct the errors [22]. For example, a physical therapist may describe a gait error by saying that patients are looking at their feet, or will try to correct the error by explaining that they should look at a fixed point ahead of them because the ground is smooth and they do not need to look down.

The extraordinary diversity of available feedback and the creativity in how it is applied have increased the opportunities for adapting and optimizing each intervention based on the user's needs.

Technological advances have prompted the development of new retraining walking techniques based on feedback provided by monitored instrumentation systems and associated software, as confirmed in various studies [14,15,26,27]. In general, these instruments capture real-time information on the gait [2,4,6,13,15-18,20,21,27], allowing the clinician to decide whether the user should receive concurrent feedback or whether it is more convenient to delay the information and verbally provide terminal feedback.

Functional recovery of walking and assisted gait are generally related to different areas of health, including neurology [28,29], gerontology [30-32], and rheumatology [33]; however, one context in which the need to control and optimize the loads applied to the forearm crutches is most clearly evident involves patients with a lower limb musculoskeletal injury [34], who have to relearn how to automate the walking gesture correctly [35]. Traumatological physiotherapy therefore very often involves using forearm crutches to partially unload the injured lower limb, progressing to full weight-bearing. Examples of such applications include sprained knees [36] or ankles [37], surgery for knee meniscopathies [38], and hip arthroplasties [21]. Consequently, the load exerted on the body part in question is an essential assisted gait parameter [39], and therefore this load must be objectified and controlled to optimize it in line with the pathology, recovery process phase, and user's characteristics. Another factor is the current tendency to strongly recommend that the affected lower limb support as much weight as possible without damaging the injury. Underloading could lead to circulatory and muscle tone deficits, resulting in a decrease in osteoblastic action and increase in osteoclastic action, as well as inhibition of the joint and muscle plantar proprioceptive receptors that would imply a functional deficit. In contrast, overloading could lead to compressions or undue stress of structures even without regeneration or in the recovery process [6].

Load-measuring tools are required to both determine the ideal load in each case, which entails specific action protocols based on scientific evidence, and to objectively and progressively increase the loads [4]. At present, there are some instruments available that measure loads, such as insoles [4] or force platforms [2,15]; however, these tools do not allow for monitoring the loads exerted on the forearm crutches in every type of support, the balance between them (in bilateral-assisted gait), the simultaneity of their supports, or even their rhythm, among other relevant factors. These drawbacks are overcome by the use of instrumented forearm crutches such as the GCH

System [6] that can monitor the loads applied to them so that the loads applied to the structure in question are reduced. However, these tools must be used together with a program that allows users and/or health professionals, in clinical or laboratory conditions, to obtain as much information as possible about the assisted gait performance. To the best of our knowledge, there is no scientific evidence related to the control software of instrumented forearm crutches and its implemented feedback mechanisms. As a result, the main objective of this study was to design an innovative multifunctional desktop load-measuring software that complements the GCH System 2.0–instrumented forearm crutches and monitors the loads applied to them, displaying real-time graphical and numerical information, as well as the correction of inaccuracies through feedback technology during assisted gait. The secondary objective was to perform a preliminary implementation study of the newly developed software.

Methods

GCH System 2.0

GCH Control Software implements GCH System to measure the loads applied to forearm crutches. Its application is particularly suitable in physiotherapy and traumatology, fields that frequently work on the functional recovery of gait for patients with lower limb musculoskeletal injury [40,41]. The rehabilitation process starts with total or partial unloading of the injured limb, followed by a gradual increase in loads aided by crutches until achieving complete functional recovery [42].

The instrumented crutches must first be described to understand how the software works. GCH System involves the coupling of a miniature force sensor inside the distal area of the crutch, and the measures are wirelessly transmitted. Each crutch contains an ultralow-power microcontroller with an input voltage of 2.4 V direct current and battery/autonomy of 6000 mAh (8 hours). The outgoing signal is detected by a USB receiver connected to a computer (“fixed system”) using a virtual communication port, or by a receptor built into a watch, mobile phone, pendant, or other portable device. In the latter case, patients have autonomy to practice aided gait outdoors with the usual obstacles encountered in daily life, such as steps and slopes, without any professional supervision. The technical specifications of GCH System 2.0 are reported elsewhere [6]. Each of these fixed and portable systems requires an independent implementation program with different features to allow each patient to achieve their specific therapeutic goals. The fixed system is described below in line with the main objective of the study.

GCH Control Software 1.0

Design

GCH Control Software is an instrumentation program that has to be used on a PC; therefore, it was designed for use during patient care in the clinic or in a laboratory for research purposes.

GCH Control Software was programmed in C# in a Windows operating environment, and another program is needed as a server (GCH Server). Several GCH Control Software sessions can run simultaneously on the same server so that several patients can use the instrumented crutches at the same time. Other specific programs such as Framework 4.0 by Microsoft’s SqLite libraries for Windows and the Codejock graphical software libraries may have to be installed.

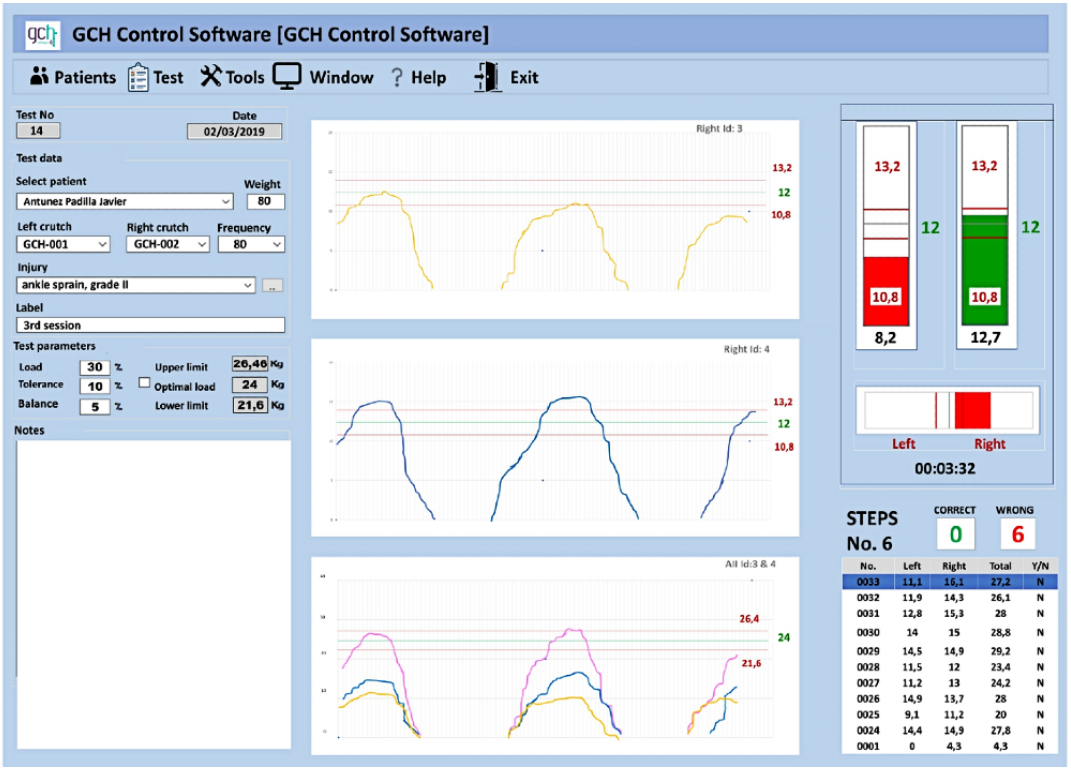
The software was developed by the first author (GC) and is registered in the Andalusia Regional Intellectual Property Registry under file number SE-690-15 and registration number 201599901533465. The source code is not available. Note that the Spanish language is used in all messages as the software was designed for use with Spanish patients.

The software translates the crutch sensor signal in millivolts into units of force (kilogram-force, which is a nonstandard unit but commonly used in our clinical context), records and analyzes the data, and provides a real-time on-screen numerical and graphical display of the loads exerted by the subject on one crutch if the assisted gait is unilateral or on two crutches if it is bilateral. The program records data up to 80 Hz, providing effective curves of weight-bearing at each crutch support.

Figure 1 shows the general interface of the program, displaying data about patients, pathologies, crutch loads recommended by the physiotherapist, permitted plus or minus tolerance margins of error in the maximum loads of each crutch step and in the load balance between crutches, and a notepad. The line graphs display the independent loads of each crutch, unified load of both crutches, and overlapping of the three previous loads. Bar graphs display the following loads: independent loads for each crutch (double bar), with the option to view the unified load (single bar), and a horizontal bar that represents the load balance between both crutches (balance bar). All of these data are displayed numerically, with a correct and wrong crutch step count (according to the load accuracy).

A detailed step-by-step description and operation of the software is provided after identifying the crutches to be used in the GCH Server.

Figure 1. Program interface screenshot. From top to bottom: the left column shows user's data, clinical test data, and observations; the middle shows line graphs of right crutch load, left crutch load, right and left crutch overlapping graphs, and the sum of both loads; and the right column shows double bar (right and left crutch vertical bars), balance bar, and correct and wrong crutch step counter (according to the load accuracy).

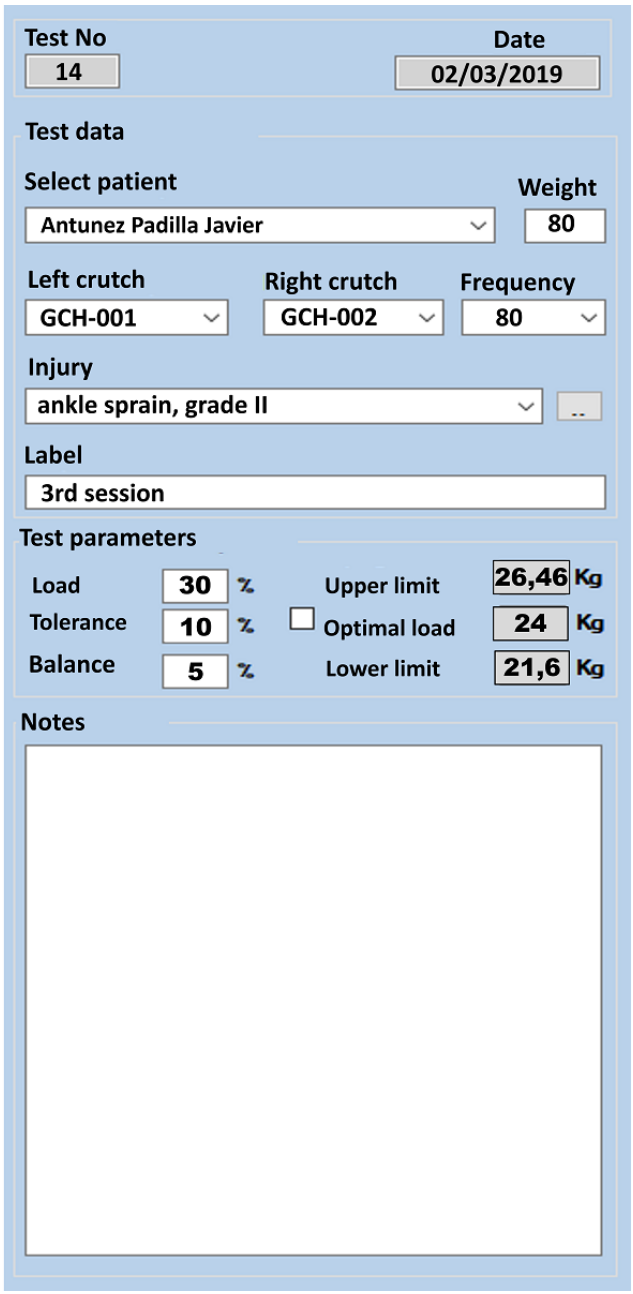


Data Logging Database

The database includes user registration information such as the patient's affiliation data and clinically relevant personal background. Each crutch is identified as "right" or "left," and the gait is classified as unilateral or bilateral. Preoperative clinical and technical data (Figure 2) are required for an autonumber test, such as the subject's current weight (because body weight percentages will be used), the frequency at which the data will be collected (20, 40, or 80 Hz), the current pathology, and a possible test identification label.

The next step is to record the ideal load to be exerted on the crutch, as recommended by the physiotherapist (ie, the load reduction applied to the injured limb, in percentage body weight or kilograms, which changes automatically); the permitted margins of error for overloading or underloading; and the permitted tolerance in the load balance between the two crutches, only in the case of bilateral assisted walking.

Figure 2. Left-hand column of the general interface shown in Figure 1, including (from top to bottom): autonumber test, date, patient's name, current weight, crutch identification, data frequency, lesion, test label, test parameters (ideal load requested, tolerance for load accuracy errors, tolerance for imbalance between both crutches), and notepad.



Graphical and Numerical Data Provided by the Software

The program uses several graphics to display the user's walking performance in real time: line graphs (see Figure 1, middle column) and bar graphs (see Figure 1, right column).

Line graphs are only for use by the professional, in which the right crutch load, the left crutch load, and their sum are shown in different colors. There are three possible versions of line graphs: "L+R" (left + right), which shows the crutches separately; "L and R" (left and right), where a graphic overlay of the crutches is added to display both of them simultaneously; and "ALL," which displays the superimposed graph of the two crutches plus the total load graph.

Both the researcher and clinician can view bar graphs on the computer (see Figure 1, right column, upper), and the user can see these graphs displayed on an extendable screen by a projector. In this case, there are three main display options. The first is a double-bar display, which involves two vertical bars (one for each crutch), showing the different loads. This method is useful for monitoring the load amount and how it is distributed during bilateral walking to ensure it is balanced and simultaneous, without fluctuations. This version offers the most complete information. The second option is a single bar that adds up the total loads exerted by both crutches. In this case, patients focus their attention on the total load, provided that they do not have any load asymmetry problems between crutches. The information is interchangeable with the double-bar feedback display. The third display option is the balance bar. With this horizontal bar, the user's attention focuses on

balancing loads between both crutches; therefore, this display is not used in unilateral assisted gait.

All of the line and bar graphics have three horizontal lines (see Figure 1): the middle green line shows the ideal load, and the red lines above and below indicate the permitted plus or minus tolerance margin (numerical data already included in the software).

The program interface displays other interesting data such as the number of correct and wrong crutch steps in each test, along with the respective load of each crutch (see Multimedia Appendix 1).

Each test is recorded and automatically dated in the database, making it possible to access, display, and analyze graphs and numerical data regarding load accuracy, mistakes made, patient progression, and other factors, thereby enabling comparisons and statistical analyses (Figures 3 and 4).

Figure 3. Screenshot of an example database with load results (Hz), displaying crutch identifiers, left crutch load, right crutch load, total load, time (ms), and time stamp. In this case, the right crutch starts the load earlier than the left crutch. The maximum peak of strength of the left crutch is greater than that of the right, but both coincide in time. Finally, the support time of the crutches is higher than 1 second.

ID Crutches	Left Loading	Right Loading	Total Loading(L+R)	Test Time(ms)	TimeStamp
1 - 2	0	0	0	40176	17:01:07.838
1 - 2	0	0	0	40239	17:01:07.900
1 - 2	0	0	0	40303	17:01:07.963
1 - 2	0	168	168	40365	17:01:08.025
1 - 2	0	1895	1895	40429	17:01:08.087
1 - 2	0	2559	2559	40487	17:01:08.150
1 - 2	331	4307	4638	40550	17:01:08.212
1 - 2	3282	6417	9699	40620	17:01:08.275
1 - 2	6361	8892	15253	40681	17:01:08.337
1 - 2	10545	10950	21495	40747	17:01:08.399
1 - 2	14920	11416	26336	40799	17:01:08.462
1 - 2	16685	12330	29015	40862	17:01:08.524
1 - 2	18041	14396	32437	40928	17:01:08.587
1 - 2	21677	16503	38180	40994	17:01:08.649
1 - 2	24451	17577	42028	41050	17:01:08.711
1 - 2	25850	17998	43848	41112	17:01:08.774
1 - 2	26246	19046	45292	41176	17:01:08.836
1 - 2	26401	19478	45879	41237	17:01:08.899
1 - 2	25420	20118	45538	41298	17:01:08.961
1 - 2	24542	21282	45824	41361	17:01:09.023
1 - 2	23626	21725	45351	41424	17:01:09.086
1 - 2	19158	17497	36655	41488	17:01:09.148
1 - 2	5219	4130	9349	41548	17:01:09.211
1 - 2	0	0	0	41614	17:01:09.273
1 - 2	91	97	188	41673	17:01:09.335
1 - 2	11	0	11	41736	17:01:09.398
1 - 2	0	5	5	41797	17:01:09.460
1 - 2	6	0	6	41860	17:01:09.523
1 - 2	0	0	0	41922	17:01:09.585

Figure 4. Screenshot of an example database showing maximum loads for each crutch support and feedback results, displaying crutch identifiers, maximum left crutch load, maximum right crutch load, total load, crutch support number, correct crutch supports, wrong crutch supports, time (ms), and time stamp. Note that this information enables easy assessment of loading errors, as well as of the homogeneity and symmetry between the maximum loads on the crutches.

ID Crutches	Left Loading (KG)	Right Loading (Kg)	Total Loading (L+R) (Kg)	Step Number	Correct Steps	Wrong Steps	Test Time (ms)	TimeStamp
1 - 2	0	2	2	1		1	737	17:00:30.086
1 - 2	5,3	10,8	16,1	2		2	4668	17:00:53.486
1 - 2	7,2	6,8	14	3		3	38551	17:01:06.777
1 - 2	26,4	21,7	48,1	4		4	41421	17:01:09.273
1 - 2	11,6	15,5	27,1	5		5	43044	17:01:11.020
1 - 2	16,6	20,1	36,7	6		6	44854	17:01:12.830
1 - 2	17,6	21,5	39,1	7		7	46538	17:01:14.577
1 - 2	28,8	21,8	50,6	8		8	52591	17:01:20.505
1 - 2	15,7	21,6	37,3	9		9	54276	17:01:22.190
1 - 2	23,2	17	40,2	10		10	56148	17:01:24.062
1 - 2	17,2	23,3	40,5	11		11	58020	17:01:25.934
1 - 2	19	22,2	41,2	12		12	63948	17:01:31.924
1 - 2	17,1	19,1	36,2	13		13	65820	17:01:33.734
1 - 2	12,9	19,4	32,3	14		14	67692	17:01:35.668
1 - 2	18,5	18,5	37	15	1		69439	17:01:37.415
1 - 2	22,7	17,2	39,9	16		15	75804	17:01:43.718
1 - 2	16,6	18,8	35,4	17		16	77614	17:01:45.465
1 - 2	22,3	17,5	39,5	18		17	79361	17:01:47.399
1 - 2	18,1	18,2	36,3	19	2		81357	17:01:49.209
1 - 2	4,8	9,4	14,2	20		18	107066	17:02:22.655
1 - 2	10,3	10,9	21,2	21		19	129909	17:02:41.563
1 - 2	2,8	2	4,8	22		20	135583	17:02:43.996
1 - 2	20,3	18,4	38,7	23		21	137955	17:02:45.993
1 - 2	13,9	21,7	35,6	24		22	139702	17:02:47.865
1 - 2	15,6	15,1	30,7	25		23	141574	17:02:49.612
1 - 2	17,1	16,3	33,4	26	3		143446	17:02:51.422
1 - 2	4	3,4	7,4	27		24	144943	17:02:52.982

Feedback Mechanism

GCH Control Software gives professionals numerical and graphic information about the loads exerted on the crutches, the precision and homogeneity of these loads, and the simultaneity or balance between crutches. All of these data can be used to assess the subjects' assisted gait and give them oral feedback for correcting their mistakes. The software also includes a useful and feasible feedback mechanism that tells patients directly about their performance and warns them if they make mistakes while walking, which they can then self-correct. For design of the software, the most frequent and efficient feedback methods described in the literature were identified (ie, immediate visual feedback followed by terminal and immediate acoustic feedback [43]), and the most appropriate approaches were selected and implemented in the system's user interface.

We implemented a simple beep as auditory feedback. A high-pitched beep is used when the subject overloads the affected limb, and the most annoying sound is used in this case owing to the clinical risk involved. The beep length must be short enough (ie, 0.3 seconds) to start just after the stand phase (crutch support and foot) ends and before the crutches and the foot next touch the ground. Different alternatives such as line charts, gauge charts, and bar charts were tested. Among them, vertical bar charts were found to be easier to interpret, especially if there are two bars (one for each crutch), which also makes it possible to compare the load amount and simultaneity between the two crutches.

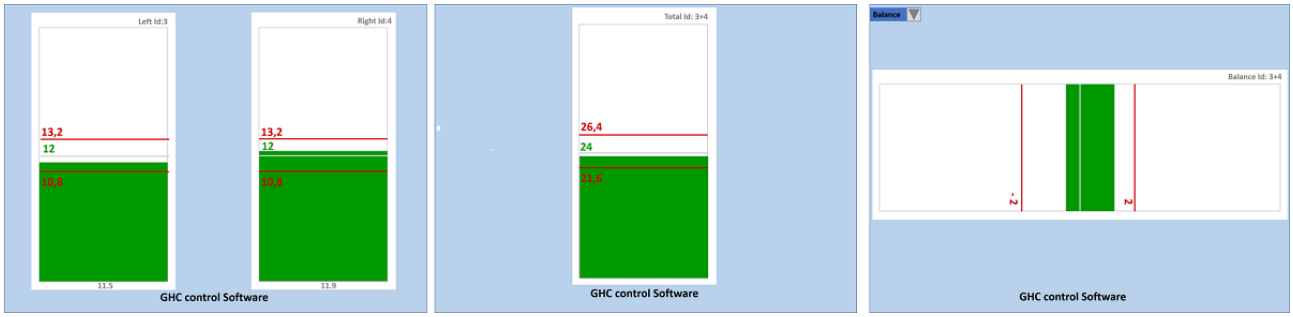
The bar graphs mentioned above are preferably displayed to patients on an extendable screen (Figure 5), using a projector that is large enough to show straight line paths at least 16 meters long (see Clinical Implementation section). The bar graphs are applied one at a time, as each graph is designed to focus the user's attention on the parameter that the physiotherapist considers to be the most necessary. Thus, priority is given to correcting the most serious mistakes first and the least significant mistakes last, until the assisted gait is improved, enabling the subject's functional recovery.

Similar to the graphics that professionals see on the computer, all of the bars feature a green line to show the ideal load recommended by the physiotherapist and two red lines to indicate the minimum required or maximum allowed load (Figure 5). In the first two double-bar and single-bar versions, the bar content turns red when it is outside the tolerated load range or turns green if it is in the correct range. The error beep is therefore only heard when the red color appears to complement the visual feedback.

If a balance bar is displayed, as long as the right and left crutch load balance is maintained, the bar content stays green in the

middle or slightly shifts to the side bearing the most weight (Figure 5), but without exceeding the red lines that show the error tolerance margins. If the margins are exceeded, the bar content turns red and the subject hears a beep. It does not make any sense to use this screen if the assisted gait is one-sided (ie, with only one crutch).

Figure 5. Screenshots of the three possible extendable screens projected in front of patients to show them visual feedback of the loads exerted on the forearm crutches during bilateral assisted walking. From left to right: double bar (one for each crutch), single bar (sum of both crutches), and balance bar (balance between the two crutches). The bars are green in this case because the loads are within the tolerated margins of error. The balance bar shows that the subject is applying more force to the right crutch than to the left crutch, but within the permitted margin of error.



Clinical Implementation

Study Design

Once the desktop software had been designed and developed, and the numerical and graphic measurements had been confirmed in a validation and reliability study [6], a pilot observational descriptive study was performed to obtain preliminary information on the clinical performance of the feedback mechanism implemented by GCH Control Software.

Participants

The sample consisted of 10 participants, including 5 women (50%) and 5 men (50%), with an age range from 21 to 55 years (mean 37.40 years, SD 11.86) and weight range between 52.6 and 81.7 kilograms (mean 64.21 kilograms, SD 9.90), all selected by nonprobability and convenience sampling. Details are given in Multimedia Appendix 2. All subjects had a normal BMI according to the US National Heart, Lung, and Blood Institute.

Inclusion criteria were healthy subjects aged between 18 and 60 years; had previous experience in partially unloading an affected lower limb by bilateral assisted gait with forearm crutches; scoring 3-4 in each of the Chamorro Assisted Gait Scale (CHAGS) items [1] (10 items, 0-4 points each); with normal gait, and asymptomatic when walking at a free cadence; and passing a simple static equilibrium test, consisting of keeping one's balance on each foot for 30 seconds without moving the body [44].

Exclusion criteria were suffering from an evident disorder of overall coordination and physical skill that could alter the aided gait with crutches, or suffering from visual or acoustic disorders that would prevent the individual from receiving biofeedback during the intervention.

This research protocol was approved by the Ethics Committee of Virgen Macarena University Hospital (Seville, Spain). All subjects provided written informed consent before participating in the experimental study.

Measurements and Data Collection

Measurements were taken with GCH System 2.0 [6], and numerical and graphic data were monitored by GCH Control Software. The maximum vertical ground reaction force on the crutches was measured immediately after each support (ie, after each stride). The monitored data were immediately logged in the software database for subsequent use. The pilot implementation was carried out under laboratory conditions.

The subjects walked at a free cadence, taking steps as frequently as they considered most comfortable, along 15.5 linear meters (for a minimum of 10 strides or 20 steps) on an even ground and facing a projector that displayed the image projected on the outward journey (Figure 6). They were asked to walk bilaterally with forearm crutches in two stages [1] to partially unload their previously injured lower limb. In other words, the lower limb to be unloaded is supported by the crutch (if unilateral) or by both crutches (if bilateral) at the same time. Thus, the vertical ground reaction force is applied simultaneously to the sole of the foot and to the crutches so that the subject's body weight is distributed evenly (Figure 7).

When the subjects leaned on the injured lower limb, 50% of the body weight had to be unloaded. In other words, 50% of the ground's vertical reaction force was applied to the crutches. A ±10% load tolerance margin was permitted (±5% of the ideal load). The crutch balance tolerance margin was 5%. For example, the two crutches should ideally bear a weight of 50 kilograms for a patient weighing 100 kilograms, but a 47.5-52.5 kilogram distribution would be considered correct. This is easy to observe in the single-bar graph (which unifies the loads). In the double-bar graph, each crutch should ideally bear 25 kilograms and a 23.75-26.25 kilogram distribution would be considered correct. Furthermore, a difference of more than 2.5 kilograms between the load of one crutch and the other is not allowed in the balance bar.

Participants were allowed to practice for 2 minutes and were reminded of the homogeneous nature of the maximum loads. Since they had mastered this type of walking, no more time was needed to confirm their automation and check that they met the correct CHAGS scale rating [1] (as described above for the inclusion criteria).

Figure 6. User walking toward the screen with double-bar feedback. The aided gait is bilateral (two forearm crutches) with a two-stage partial load (simultaneous heel and crutches support).

Figure 7. Bilateral aided gait with a partial load in two-stages, with simultaneous heel and crutches support.

Users received visual and beep-based acoustic feedback at the same time. The physiotherapist gave no verbal feedback. The 10 participants were assessed to see if they applied the correct or wrong maximum loads requested by the physiotherapist (ie, ideal load) to the forearm crutches during bilateral assisted walking. Each participant took a minimum of four walks. They began each walk with both feet on the same line and then took their first step with the foot to be unloaded together with both crutches simultaneously. Participants performed the first walk (walk 0) without any feedback to obtain capacity baseline data, and then walked three more times with the double-bar feedback (Figure 5). That is, they were warned both visually and acoustically if they applied the wrong load to the crutches immediately at the end of each crutch support. The two independent bars also showed them when the loads applied to both crutches were simultaneous and balanced. Since this can provide too much feedback in some cases, extra walks with other simpler types of feedback (eg, three walks with a single bar or three walks with a balance bar) were applied in line with participants' needs (up to a maximum of 10 walks). The physiotherapist decided who had to continue based on observational assessment and computer screen feedback to

evaluate the tests. The system recorded 10 strides for each walk, or 10 crutch weight-bearing per stride or unloaded step. In other words, for 10 participants×10 loads (crutches)×a minimum/maximum of 4 of 10 walks, there were at least 400 measurements and a maximum of 1000 measurements available for analysis.

Statistical Analysis

The descriptive analysis included the means and percentages of maximum loads applied to the crutches at each support. The Wilcoxon matched-pairs signed-rank test was used to evaluate the load accuracy evolution. The analysis provided the *P* value to evaluate each subject's progress, comparing each of their walks. The practice walks of different subjects were not compared with each other. A confidence level of 95% was considered, and the experimental *P* value was evaluated at a significance level of 5%. The data obtained were organized and analyzed with IBM SPSS statistical software (version 22.0).

Results

Table 1 describes the crutch loads of the first 4 walks (1 without feedback, 2-4 with feedback). Loads are displayed as a percentage of the ideal load requested (50% of each subject's body weight), with the optimum load subsequently set to 100%. The ideal load permissible tolerance was ±10% (ie, ±5%, or a load of 95% to 105%). See Multimedia Appendix 3 for a more general visual presentation of these data.

A total of 96/100 errors were observed in the first walk (walk 0, no feedback), most of which involved crutch overloading by 7/10 subjects (subjects 1, 2, 3, 4, 5, 7, and 9). Errors ranged from 61.09% (subject 8) to 203.98% (subject 1). The loads of subjects 2, 6, and 9 were particularly nonhomogeneous.

The first walk with double-bar feedback (walk 1) began with errors similar to the no-feedback walk errors, generally followed by attempts at correction. Therefore, progress toward the ideal load and fluctuations above and below the optimal load were observed until the optimal load was achieved and maintained as much as possible in the successive walks.

Table 1. Percentage of the loads applied on the crutches during the first 4 walks, with an optimum load of 100% and permitted error range of 95%-105%.

Feedback with crutch supports	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10
Walk 0: Computer feedback[a]										
1	201.08-O[b]	133.98-O	146.39-O	121.07-O	107.98-O	90.41-U[c]	149.87-O	70.70-U	143.20-O	81.67-U
2	198.19-O	130.93-O	144.68-O	125.52-O	109.89-O	86.64-U	153.81-O	72.87-U	157.48-O	78.40-U
3	199.28-O	137.59-O	143.45-O	126.71-O	109.13-O	88.70-U	153.28-O	72.56-U	169.39-O	77.31-U
4	202.17-O	108.46-O	142.47-O	123.44-O	106.84-O	91.44-U	153.28-O	68.53-U	167.01-O	81.31-U
5	203.98-O	121.78-O	143.70-O	122.55-O	107.60-O	92.81-U	147.77-O	71.01-U	148.64-O	82.76-U
6	201.45-O	120.39-O	145.41-O	126.11-O	*103.04*[d]	*95.21*	154.59-O	66.36-U	158.16-O	78.40-U
7	198.55-O	122.05-O	145.17-O	123.74-O	*104.56*	94.52-U	150.66-O	67.91-U	166.67-O	79.49-U
8	201.81-O	131.21-O	143.70-O	122.55-O	109.89-O	86.99-U	147.51-O	62.33-U	171.43-O	76.23-U
9	203.25-O	128.99-O	142.23-O	126.41-O	*104.94*	87.67-U	155.12-O	61.09-U	172.11-O	81.31-U
10	195.66-O	110.68-O	145.90-O	120.18-O	107.22-O	82.53-U	148.03-O	80.62-U	164.97-O	86.03-U
Walk 1 (double bar): Screen feedback[e]										
1	150.09-O	144.80-O	158.14-O	123.44-O	117.11-O	83.22-U	162.47-O	66.98-U	148.30-O	78.40-U
2	138.88-O	154.79-O	154.47-O	122.26-O	112.93-O	88.70-U	144.88-O	88.06-U	191.84-O	91.83-U
3	118.63-O	128.16-O	80.29-U	*97.63*	*96.20*	91.78-U	122.57-O	109.77-O	143.20-O	*101.27*
4	*101.99*	134.26-O	131.95-O	108.01-O	*98.48*	*96.23*	109.71-O	*102.02*	94.90-U	*102.00*
5	109.95-O	82.39-U	124.11-O	72.11-U	*97.72*	94.52-U	*101.05*	*104.19*	72.45-U	*101.63*
6	92.22-U	85.71-U	85.92-U	76.85-U	*97.34*	*97.95*	*100.52*	*102.95*	87.76-U	*99.46*
7	*98.01*	93.20-U	*95.96*	82.79-U	*100.38*	*100.00*	*100.79*	*100.47*	*104.42*	*101.27*
8	*103.44*	*97.36*	*99.14*	93.77-U	*99.24*	90.75-U	*97.38*	107.60-O	153.74-O	*95.83*
9	*99.46*	*101.25*	*104.53*	*96.44*	*98.86*	94.52-U	*102.10*	106.05-O	*104.08*	94.74-U
10	*98.73*	110.40-O	*99.39*	*96.14*	*98.10*	*97.95*	*101.57*	*100.47*	120.41-O	*100.91*
Walk 2 (double bar): screen feedback[e]										
1	129.48-O	117.61-O	106.49-O	84.87-U	*100.00*	106.85-O	109.45-O	110.70-O	142.18-O	*103.09*
2	111.75-O	106.52-O	*104.53*	93.47-U	*101.90*	*103.77*	106.56-O	88.99-U	131.29-O	*104.17*
3	*100.18*	*97.64*	*99.39*	108.90-O	*100.38*	*101.37*	*102.89*	91.78-U	*96.60*	107.80-O
4	*95.84*	*100.69*	*99.88*	105.64-O	*101.90*	*100.68*	*104.46*	93.64-U	87.07-U	*102.00*
5	*101.99*	107.91-O	*95.96*	108.90-O	*103.04*	108.56-O	*103.94*	*97.98*	*97.62*	*101.27*
6	*103.80*	*102.08*	*99.14*	*102.67*	*100.38*	*103.42*	*99.21*	111.01-O	119.05-O	*103.81*
7	*101.63*	*97.36*	*96.45*	*98.22*	*99.24*	*102.40*	*100.00*	*103.88*	116.33-O	*99.46*
8	*100.90*	*98.75*	*95.23*	83.98-U	*98.48*	*101.71*	*101.31*	*101.71*	*104.42*	*98.37*
9	*101.27*	*100.14*	92.78-U	*99.70*	*100.00*	*103.08*	*102.10*	*99.53*	*100.68*	*100.54*
10	*100.54*	*101.80*	*98.16*	*102.97*	*101.52*	*101.71*	*99.74*	*97.67*	109.86-O	*103.45*
Walk 3 (double bar): screen feedback[e]										
1	*99.82*	107.35-O	*97.43*	91.69-U	*98.48*	*101.37*	*99.21*	109.15-O	144.90-O	*99.09*
2	91.86-U	*104.85*	*97.43*	*96.74*	*100.00*	*103.42*	*98.69*	*104.81*	130.61-O	*97.28*
3	*98.01*	*102.91*	*99.39*	*99.11*	*98.86*	105.14-O	*100.79*	*101.09*	*97.28*	*99.82*
4	*102.35*	*104.02*	*101.10*	*102.67*	*98.10*	*99.66*	*102.10*	*100.47*	86.05-U	*103.81*
5	*104.52*	*103.19*	*102.57*	*101.19*	*98.48*	*99.32*	*98.16*	*102.64*	*101.36*	*102.72*

Feedback with crutch supports	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10
6	105.24-O	*102.36*	*97.43*	*99.41*	*100.76*	*97.95*	*103.15*	*101.71*	*96.60*	*98.00*
7	*102.71*	*104.30*	*99.39*	*98.52*	*103.04*	*98.63*	*104.20*	*99.53*	*100.68*	*97.28*
8	*103.44*	105.41-O	*97.18*	*103.26*	*100.38*	*101.71*	*102.62*	*103.88*	111.22-O	*98.37*
9	*103.44*	107.91-O	*99.63*	*100.59*	*98.86*	*100.00*	*101.31*	*101.09*	115.31-O	*100.18*
10	*100.18*	105.13-O	*99.39*	*101.48*	*98.10*	*104.79*	*102.89*	*99.22*	*103.74*	*103.45*

[a]Only for physiotherapist/researcher.

[b]O: overloaded.

[c]U: underloaded.

[d]Correctly applied loads are in italics.

[e]For patients.

Figure 8 presents the number of errors in the loads applied on the crutches in the first 4 walks, distinguishing between underload errors, overload errors, and the total for each subject and walk. The numerical data are shown in Multimedia Appendix 4.

Table 2 and Figure 9 show the mean percentage loads exerted on the crutches in each walk. The optimal load is considered to be 100%, and the error tolerance range (10%) was between 95% and 105%. The mean error table (Table 2) would be meaningless without Table 1 and Figure 8. For example, if a subject makes extreme overload and underload errors, the mean might seem to be within the tolerated range and therefore correct, but it is not. However, if all errors are overload or underload errors, or the loads are homogeneous, the clinical interpretation of the results in Table 2 would be different.

The Wilcoxon matched-pairs signed-rank test, which was used to evaluate each subject's progress, showed significant differences in the reduction of overload errors when comparing the baseline values to each of their feedback walks: walk 0 vs walk 1 (P=.02), walk 0 vs walk 2 (P=.03), walk 0 vs walk (P=.02). However, there were no significant differences in the values of the underload errors. The practice walks of different subjects were not compared with each other. A confidence level of 95% was considered, and the experimental P value was evaluated at a significance level of 5%. The table with all P values is provided in Multimedia Appendix 5.

Figure 8. Graphical representation of the number of errors in the weight supported by the crutches on each walk and per subject.

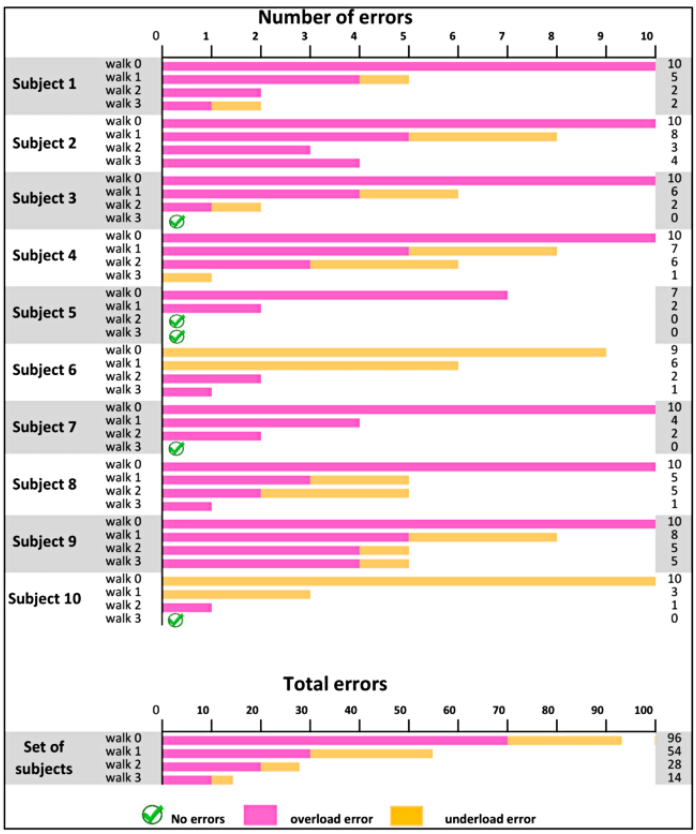


Table 2. Mean percentage (range) loads applied on the crutches during the first 4 walks, with an optimal load of 100% and a permitted error range of 95%-105%.

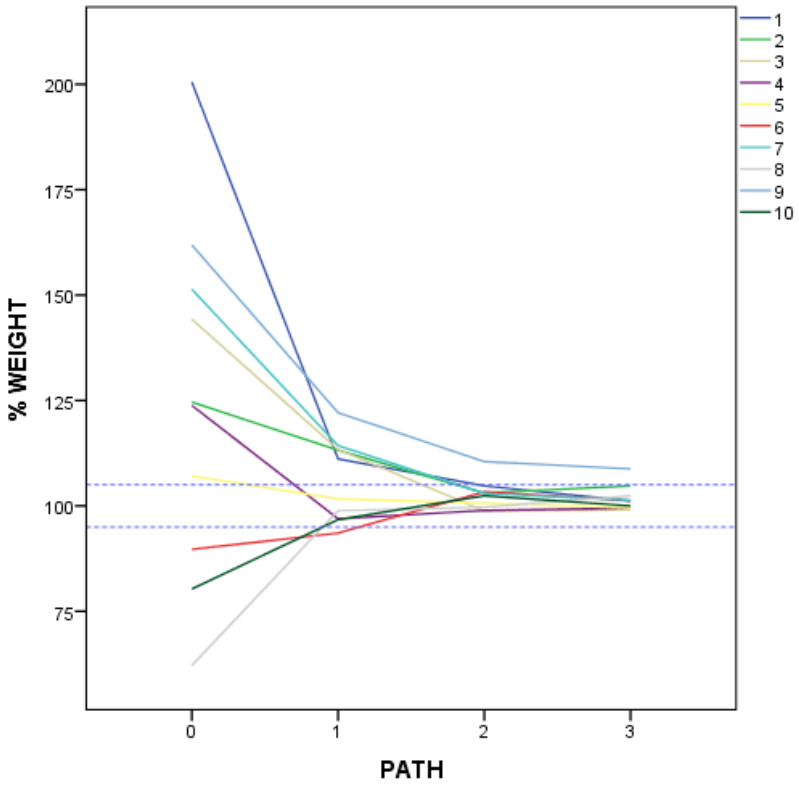
Subject	Walk 0	Walk 1	Walk 2	Walk 3
1	200.54 (195.66-203.98)-U[a]	111.14 (92.22-150.09)-U	*104.74 (95.84-129.48)*[b]	*101.16 (91.86-105.24)*
2	124.61 (108.46-137.59)-U	113.23 (82.39-154.79)-U	*103.05 (97.36-117.61)*	*104.74 (102.36-107.91)*
3	144.31 (142.23-146.39)-U	113.39 (80.29-158.14)-U	*98.80 (92.78-106.49)*	*99.09 (97.18-102.57)*
4	123.83 (120.18-126.71)-O[c]	*96.94 (72.11-123.44)*	*98.93 (83.98-108.90)*	*99.47 (91.69-103.26)*
5	107.11 (103.04-109.89)-O	*101.64 (96.20-117.11)*	*100.68 (98.48-103.04)*	*99.51 (98.10-103.04)*
6	89.69 (82.53-95.21)-O	93.56 (83.22-100.00)-O	*103.35 (100.68-108.56)*	*101.20 (97.95-105.14)*
7	151.39 (147.51-155.12)-U	114.30 (97.38-162.47)-U	*102.97 (99.21-109.45)*	*101.31 (98.16-104.20)*
8	62.14 (61.09-72.87)-O	*98.86 (66.98-109.77)*	*99.69 (88.99-111.01)*	*102.36 (99.22-109.15)*
9	161.91 (143.20-172.11)-U	122.11 (72.45-191.84)-U	110.51 (87.07-142.18)-U	108.78 (86.05-144.90)-U
10	80.29 (76.23-86.03)-O	*96.73 (78.40-102.00)*	*102.40 (98.37-107.80)*	*100.00 (97.28-103.81)*

[a]U: underload mean error.

[b]Correctly applied mean loads are in italics.

[c]O: overload mean error.

Figure 9. Line graph representing the mean percentage loads (y-axis) for each subject (different colors) in the 4 walks (x-axis).



The results regarding the right and left crutch load imbalance errors made during the three double-bar walks were as follows: 0 errors for subjects 1, 3, 5, 6, 7, and 10; 1 error for subjects 4 (walk 1), 8 (walk 2), and 9 (walk 1); and 6 fluctuation errors for subject 2 (3 errors in walk 1, 2 errors in walk 2, and 1 error in walk 3).

Subjects 2 and 9 performed extra walks with other types of feedback, namely balance bar and single bar. The research physiotherapist made this decision after noting that these subjects were finding it difficult to achieve their objectives and needed to focus attention on a specific aspect.

Subject 9 performed 2 more walks with the single-bar feedback to focus on the total load, and progressed from 8, 5, and 5 errors in the three double-bar walks (Figure 8 and Multimedia Appendix 4), respectively, to 3 (first single-bar walk) and 1 (second single-bar walk) errors, thereby completing the test.

Subject 2 performed 2 more walks with the balance-bar feedback to focus on the left and right crutch load balance, and progressed from 3, 2, and 1 imbalances with load fluctuations in the three

double-bar walks, respectively, to 2 imbalances with fluctuations (first balance-bar walk) and 0 without fluctuations (second balance-bar walk), thereby completing the test.

Discussion

Principal Results

The innovative multifunctional desktop program GCH Control Software 1.0 monitors crutch loads, which can be controlled by real-time graphic and numerical displays during assisted gait practice. The software contains a versatile feedback mechanism correcting any wrongly applied loads, ensuring greater clinical accuracy. The integrated database can be used not only to objectively evaluate the progress in loads of the subjects but also to comparatively analyze the system's efficiency and draw up clinical action protocols. Moreover, we performed a pilot implementation trial to assess the functionality of the feedback mechanism.

The application environment is basically associated with orthopedics and traumatology [6,38,41,45], and particularly to temporary processes of functional gait recovery of patients with lower limb injuries or surgeries [45] such as ligamentoplasty of the anterior cruciate ligament of the knee [36], a knee or hip prosthesis [21], or a knee osteotomy [40]. In all of these cases, treatment of the lesion starts with no load on the injured limb, which is followed as recovery progresses by partial loads that gradually increase up to a total body weight load. The GCH System is also useful in the fields of gerontology [31,32], neurology [46], pediatrics [46,47], or rheumatology [33], as mentioned earlier. If the applications are extended to these fields where the lesions tend to be chronic, orthotic devices must also be maintained over time [48], increasing the opportunities of using the method. New research prospects providing innovative clinical action protocols for users with disabilities or even normal age deterioration are therefore possible.

This study, like many other research projects [4,15,23,43,49-51], advocates technology and information technology effectiveness in functional recovery processes in general, and specifically in gait. However, effectiveness does not only refer to the intervention itself, in this case based on technological feedback, but also extends to the (global or analytical) initial, continuous, and final assessments of each essential gait parameter [52]. GCH Control Software evaluates the maximum force applied to the crutches in each support and therefore indirectly the unloading of the affected lower limb, which is one of the essential parameters [1,39] in assisted gait with two-stage partial unloading. This peak clinical moment matches the midstance phase of the affected limb (ie, where the limb remains upright and with single stance or without the help of the other limb in the air; midswing phase of the contralateral limb) [53,54]. Unloading relies exclusively on the crutches. The software also monitors if the assistive devices touch the ground at the same time and if their loads are balanced and increase progressively without fluctuations. These data are related to other essential parameters of two-stage gait with forearm crutches, as shown by the CHAGS [1] scale of observational aided gait assessment. Three of its ten load-independent items are directly related to this information: "simultaneous support of both crutches and foot," "step rhythm" (all steps have to be carried out in the same time), and "fluency" (assisted gait is performed automatically, decisively, and without hesitation).

The consequence drawn by the authors is that mutually complementary and objective technological and observational measurement tools need to be applied [55]. Similarly, despite a major goal of technology system-based clinical interventions to offer patients autonomy, they must also rely on the human eye, hands, intuition, and professional decisions [20]. In the specific case of the GCH Control Software feedback mechanism, the physiotherapist's presence is shown to optimize the technology features.

Versatile Feedback

GCH Control Software can be adapted very easily to each patient's specific characteristics, regardless of their main disease, such as their level of coordination skills [56], previous experience [56] in handling forearm crutches, age, and visual or acoustic disorders. This is because the software provides a wide range of extrinsic feedback, ensuring individualized interventions for achieving specific or general therapeutic goals when recovering gait functions. This feedback can be provided to the clinical professional, the user, or both at the same time, although the clinical professional, usually a physiotherapist, receives more detailed and extensive information than the user who is performing a dual task [57], and should only be given very specific feedback in small amounts for the intervention to be effective.

As for the feedback modalities considered (ie, acoustic, visual, and acoustic-visual), patients receive two kinds of acoustic feedback: computer-based verbal feedback from the physiotherapist and the beeps generated by the software. The advantage of the verbal feedback [49] is that the physiotherapist can decide what information to give in line with the patient's needs. This is particularly useful if the subject has poor coordination skills, as noted in the dual tasks [49]. Subjects can be given only information about their results (KR feedback) or also about their performance (KP feedback), and can receive information that only describes a mistake (descriptive feedback) or that suggests the strategy for correcting it (prescriptive feedback). The clinician can also choose between giving the information while the patient is walking (concurrent feedback) or afterward (terminal feedback), and among immediate terminal feedback, delayed terminal feedback, or summary terminal feedback (after several walks). The beeps are immediate terminal feedback, because subjects receive the signal immediately when they begin to take a step using support for the affected limb with crutches. The acoustic signal tells subjects if they have reached the correct maximum load at the end of each support, which is why it is given immediately. However, during free cadence walking [58], steps are taken very quickly, and the period of time between the maximum load and end of the support is too imperceptible so that the signal can be considered a real-time signal, and thus it can be interpreted clinically as concurrent feedback. This is particularly true if the gait is regarded as a cyclical process [59] and each crutch support or step is not assessed independently.

With regard to visual information, the main advantage is its continuity (continuous concurrent feedback) by showing how the loads applied on the crutches steadily increase and decrease in double-bar, single-bar, and balance-bar feedback. Moreover, the double-bar and single-bar feedback show if each crutch support is correct or incorrect (immediate terminal feedback).

The software's multiple independent biofeedback options can be extended by using various combinations of these options. Numerous studies have assessed the effectiveness of various types of feedback [26,27,49,60,61] and their combinations. A recent systematic review [43] of technology-based feedback and its efficacy in improving gait parameters concluded that immediate visual feedback was the most widely used, followed by terminal and immediate acoustic feedback.

GCH Pilot Implementation

In line with several previous studies [11,23,27,51], the results obtained in this study suggest that feedback mechanism–based interventions are satisfactory. The pilot experiment showed that subjects applied significantly more accurate loads on the forearm crutches after only a few minutes, or even a few seconds in some cases. These preliminary data should be confirmed in subsequent longer studies with larger samples. Nevertheless, these findings suggest that interventions are more effective, and even more efficient, owing to the feedback immediacy. Although the participant inclusion criteria were favorable to achieve positive results (ie, all participants had to have experience in two-stage bilateral assisted walking with partial unloading, and their technical gesture had to be acceptable based on a score of 3-4 points for each CHAGS item [1]), only one participant (Subject 5) had previous experience of load measurements, and with the same software.

The crutch load measurements without feedback (ie, walk 0) showed that all subjects made significant overload and underload mistakes (even with twice the requested load). Most of the subjects (7/10) overloaded the crutches, thereby unloading their body weight more than necessary, with associated clinical disadvantages. For example, in older patients with osteoporosis [32,42], unnecessary unloading is not appropriate because it inhibits osteoblastic action and increases osteoclastic action. Even if there is no underlying disease, the tendency is always to use the largest possible weight-bearing load [42], without damaging the injury, to stimulate circulation and muscle tone, and therefore optimize the patient's overall functional recovery. This is why frequent two-stage assisted walking without the injured lower limb touching the ground (foot in the air) is not recommended, unless necessary. The negative clinical consequences of underloading the crutches can be even more severe with certain injuries or types of surgery such as noncemented knee or hip prostheses, osteotomy [40,41], or autologous chondrocyte implantation [62]. In such cases, overloading the injured area can lead to further surgery, longer functional recovery times, or other side effects or sequelae, among other drawbacks [42]. Several subjects (3/10), despite being coordinated and having used crutches previously, made these mistakes constantly without feedback. The case of Subject 10 was particularly interesting, whose technical gesture with homogeneous loads were perfect, and remained well below the requested load on the crutches without feedback. This is a clear example of the usefulness of GCH Control Software, even with skilled and experienced patients.

In the three double-bar feedback walks, maximum crutch load accuracy errors decreased in all participants from the first to the third walk. At the start of the tests, and especially the first test, participants made more mistakes and differences between the requested and applied force were larger. Shortly afterward, the visual and acoustic information provided by the software helped them to adjust the loads. The biggest errors were made during the first walk; however, 3 subjects did not make any mistake in the third walk, and another 3 subjects only made one mistake, representing a minimum difference. One subject did not make any mistakes in the second or third walk. However, two of the subjects did not achieve the expected results in the double-bar walk.

Subject 2 began with very high loads and made 8/10 mistakes in the first walk, which decreased, but to a lesser extent than observed for the majority of participants (ie, 3 errors made in the second walk and 4 errors made in the third walk). During the 3 walks, it was observed that the crutches did not always load evenly, and the right and left bars fluctuated, even though the final weight of each crutch met the requirements (6/30 crutch balance errors). The physiotherapist decided to continue the intervention and apply the third type of feedback (balance bar) to allow the subject to focus their attention on leaning on both crutches at the same time and ensure even crutch loads, regardless of the total amount. As expected, the subject's performance improved remarkably with only one walk, after scoring 3/4 points in the CHAGS scale item on simultaneously using both crutches and the affected limb [1]. Therefore, occasionally, the subject did not meet this goal. The GCH Control Software feedback seems to not only facilitate load monitoring and accuracy but can also detect and correct other erroneous parameters that are equally essential in crutch-assisted gait (eg, leaning on both crutches simultaneously as in the case of Subject 2). This subject is a clear example that technological assessment and feedback (GCH System linked with GCH Control Software), combined with observational assessment and direct feedback from the physiotherapist (CHAGS score), can be used together and in a complementary manner, thereby individualizing each intervention [1].

The number of errors made by Subject 9 dropped in the double-bar feedback walks (from 7 to 5); however, there were still numerous disparate inaccuracies, in line with the prefeedback walk. Nevertheless, the subject maintained the crutch load balance. After analyzing these data, the physiotherapist asked the subject to continue the intervention with the single-bar display, and the subject made only 3 mistakes in the first walk and 1 mistake in the second walk. This improvement was explained by having focused the subject's attention [56] on the total load without displaying independent data from each crutch, which were not necessary in this case. The subject also walked more times and therefore spent more time practicing, which must also be taken into account for interpreting this improvement [56].

At the end of the pilot experiment, the subjects commented that they had been surprised by the difference between how they actually performed and how they thought they had performed. They found the test to be easy, comfortable, and effective. Even those who needed a few extra screens spoke positively about GCH Control Software and GCH System in general. They also stated that despite hearing the acoustic signals, they were generally more aware of the visual signals.

Limitations and Prospects

This study was limited to an extent by the sample size and short duration of the intervention. Longer prospective clinical trials should be performed, with large samples and control groups to ascertain the effectiveness and efficiency of the different types of graphical biofeedback (double bar, single bar, and balance), acoustic feedback (verbal and/or beep), or combined feedback available with GCH Control Software. For example, acoustic feedback using beeps and visual feedback could be applied separately, followed by evaluation of whether (and in which cases) acoustic feedback alone suffices in the long term. As explained earlier, the more options there are, the easier it is to customize or adapt the system to each patient during the treatment, both in line with individual characteristics (eg, coordination) and specific temporary musculoskeletal injuries or chronic situations due to illness or age. Clinical trials are therefore required to address different pathologies and contexts in the various branches of medicine to develop effective clinical action protocols.

Given the rapid progress of information technology, future adaptations of GCH Control Software to new communication systems will be necessary. However, these adaptations would be implemented at the lower levels of the software stack. The way feedback information is presented to patients and physiotherapists would not be affected by these adaptations.

Conclusions

This study describes GCH Control Software, the multifunctional desktop software associated with GCH System 2.0, which measures loads exerted on forearm crutches by patients who need to partially unload a pathological lower limb.

The program graphically and numerically monitors loads in real time during assisted gait interventions, and uses a versatile, efficient feedback mechanism to correct the accuracy of wrongly applied loads. It also shows if subjects use both crutches at the same time in bilateral walking and how evenly they distribute the load between them. The program features a patient and test database that can be used to objectify applied load progress and comparatively analyze the system's effectiveness in different clinical contexts, enabling drawing up clinical action protocols based on scientific evidence.

Satisfactory findings were obtained in the complementary pilot implementation tests of the feedback mechanism functionality. Although an ideal sample was selected, all subjects made errors exerting loads on the forearm crutches during bilateral gait in the baseline test, indicating the necessity or appropriateness of the method presented. They achieved greater accuracy with only one brief session. Although one subject found it difficult to balance the loads between the two crutches and fluidly apply the loads, these issues were ultimately resolved with feedback. However, any clinical data provided herein should be interpreted with caution. Future research should focus on how participants' accuracy changes over several weeks to confirm the findings.

Acknowledgments

The authors would like to thank the Telefonica Chair "Intelligence in Networks" and the "Area of Physiotherapy CTS-305" Research Group of the University of Seville, Spain, for their collaboration.

Authors' Contributions

GC conceptualized the idea for the study and software. GC and JS wrote the Introduction and Methods (GCH Control Software 1.0) sections. GC and VP wrote the Methods (Clinical Implementation) and Results sections. GC and JS drafted the Discussion. GC, JS, and VP critically reviewed the manuscript for important intellectual content. All authors contributed to the final version and approved the final paper for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Crutch step count, displaying the total number of supports, number of correct supports, and number of wrong supports. Each crutch support is also represented numerically, showing the load applied to the left, right, and both crutches, and whether or not the support was correct.
[PNG File , 152 KB - jmir_v23i9e27602_app1.png]

Multimedia Appendix 2
Subject demographics.
[DOCX File , 13 KB - jmir_v23i9e27602_app2.docx]

Multimedia Appendix 3
Visual representation of correct (green), overloaded (pink), and underloaded (gold) loads applied on the crutches during the first 4 walks (related to Table 1).
[DOCX File , 24 KB - jmir_v23i9e27602_app3.docx]

Multimedia Appendix 4
Number of errors in the weight supported by the crutches on each walk and per subject.
[DOCX File , 16 KB - jmir_v23i9e27602_app4.docx]

Multimedia Appendix 5
<italic>P</italic> values from the Wilcoxon matched-pairs signed-rank test.
[DOCX File , 13 KB - jmir_v23i9e27602_app5.docx]

References

1. Chamorro-Moriana G, Ridao-Fernández C, Ojeda J, Benítez-Lugo M, Sevillano JL. Reliability and validity study of the Chamorro Assisted Gait Scale for people with sprained ankles, walking with forearm crutches. PLoS One 2016;11(5):e0155225 [FREE Full text] [doi: 10.1371/journal.pone.0155225] [Medline: 27168236]
2. Van Kammen K, Boonstra A, Reinders-Messelink H, den Otter R. The combined effects of body weight support and gait speed on gait related muscle activity: a comparison between walking in the Lokomat exoskeleton and regular treadmill walking. PLoS One 2014;9(9):e107323 [FREE Full text] [doi: 10.1371/journal.pone.0107323] [Medline: 25226302]
3. Fischer J, Nüesch C, Göpfert B, Mündermann A, Valderrabano V, Hügle T. Forearm pressure distribution during ambulation with elbow crutches: a cross-sectional study. J Neuroeng Rehabil 2014 May 15;11:61 [FREE Full text] [doi: 10.1186/1743-0003-11-61] [Medline: 24731773]
4. van Lieshout R, Stukstette MJ, de Bie RA, Vanwanseele B, Pisters MF. Biofeedback in partial weight bearing: validity of 3 different devices. J Orthop Sports Phys Ther 2016 Dec;46(11):993-1001. [doi: 10.2519/jospt.2016.6625] [Medline: 27733088]
5. Yoshikawa K, Mutsuzaki H, Sano A, Koseki K, Fukaya T, Mizukami M, et al. Training with Hybrid Assistive Limb for walking function after total knee arthroplasty. J Orthop Surg Res 2018 Jul 03;13(1):163 [FREE Full text] [doi: 10.1186/s13018-018-0875-1] [Medline: 29970139]
6. Chamorro-Moriana G, Sevillano J, Ridao-Fernández C. A compact forearm crutch based on force sensors for aided gait: reliability and validity. Sensors (Basel) 2016 Jul 21;16(6):925 [FREE Full text] [doi: 10.3390/s16060925] [Medline: 27338396]
7. Fernández R, Rodríguez B, Barcia B, Souto S, Chouza M, Martínez S. Generalidades sobre feedback (o retroalimentación). Fisioterapia 1998;20(1):3-11 [FREE Full text]
8. Ryan J, Levac DE, Wright FV. Reliability of the motor learning strategies rating instrument in physiotherapy intervention for children with cerebral palsy. Dev Med Child Neurol 2019 Sep;61(9):1061-1066. [doi: 10.1111/dmcn.14177] [Medline: 30740648]
9. Yeom HG, Kim JS, Chung CK. Brain mechanisms in motor control during reaching movements: transition of functional connectivity according to movement states. Sci Rep 2020 Jan 17;10(1):567. [doi: 10.1038/s41598-020-57489-7] [Medline: 31953515]
10. Sultan AS, Mateen Khan MA. Feedback in a clinical setting: A way forward to enhance student's learning through constructive feedback. J Pak Med Assoc 2017 Jul;67(7):1078-1084 [FREE Full text] [Medline: 28770891]
11. High C, McHugh HF, Mills SC, Amano S, Freund JE, Vallabhajosula S. Vibrotactile feedback alters dynamics of static postural control in persons with Parkinson's disease but not older adults at high fall risk. Gait Posture 2018 Jun;63:202-207 [FREE Full text] [doi: 10.1016/j.gaitpost.2018.05.010] [Medline: 29772496]
12. Mortensen D, Bech S, Begault DR, Adelstein BD. The relative importance of visual, auditory, and haptic information for the user's experience of mechanical switches. Perception 2009;38(10):1560-1571. [doi: 10.1068/p5929] [Medline: 19950486]
13. Hunt M, Takacs J, Hart K, Massong E, Fuchko K, Biegler J. Comparison of mirror, raw video, and real-time visual biofeedback for training toe-out gait in individuals with knee osteoarthritis. Arch Phys Med Rehabil 2014 Oct;95(10):1912-1917. [doi: 10.1016/j.apmr.2014.05.016] [Medline: 24910929]
14. Segal N, Glass NA, Teran-Yengle P, Singh B, Wallace RB, Yack HJ. Intensive gait training for older adults with symptomatic knee osteoarthritis. Am J Phys Med Rehabil 2015 Oct;94(10 Suppl 1):848-858 [FREE Full text] [doi: 10.1097/PHM.0000000000000264] [Medline: 25768068]
15. Drużbicki M, Guzik A, Przysada G, Kwolek A, Brzozowska-Magoń A. Efficacy of gait training using a treadmill with and without visual biofeedback in patients after stroke: A randomized study. J Rehabil Med 2015 May;47(5):419-425 [FREE Full text] [doi: 10.2340/16501977-1949] [Medline: 25757954]

16. Lipsitz L, Lough M, Niemi J, Travison T, Howlett H, Manor B. A shoe insole delivering subsensory vibratory noise improves balance and gait in healthy elderly people. Arch Phys Med Rehabil 2015 Mar;96(3):432-439 [FREE Full text] [doi: 10.1016/j.apmr.2014.10.004] [Medline: 25450133]
17. Ki KI, Kim MS, Moon Y, Choi JD. Effects of auditory feedback during gait training on hemiplegic patients' weight bearing and dynamic balance ability. J Phys Ther Sci 2015 May;27(4):1267-1269 [FREE Full text] [doi: 10.1589/jpts.27.1267] [Medline: 25995603]
18. Quinzaños Fresnedo J, Sahagún Olmos R, León Hernández S, Pérez Zavala R, Quiñones Uriostegui I, Solano Salazar C, et al. Efectos a corto plazo del entrenamiento de la marcha en una órtesis robótica (Lokomat®) con retroalimentación auditiva en pacientes con lesión medular incompleta crónica. Rehabilitación 2015 Jan;49(1):30-37. [doi: 10.1016/j.rh.2014.10.002]
19. El-Tamawy M, Darwish MH, Khallaf ME. Effects of augmented proprioceptive cues on the parameters of gait of individuals with Parkinson's disease. Ann Indian Acad Neurol 2012 Oct;15(4):267-272 [FREE Full text] [doi: 10.4103/0972-2327.104334] [Medline: 23349591]
20. Parker J, Mountain G, Hammerton J. A review of the evidence underpinning the use of visual and auditory feedback for computer technology in post-stroke upper-limb rehabilitation. Disabil Rehabil Assist Technol 2011;6(6):465-472. [doi: 10.3109/17483107.2011.556209] [Medline: 21314295]
21. Scheidig A, Jaeschke B, Schuetz B, Trinh TQ, Vorndran A, Mayfarth A, et al. May I keep an eye on your training? Gait assessment assisted by a mobile robot. IEEE Int Conf Rehabil Robot 2019 Jun;2019:701-708. [doi: 10.1109/ICORR.2019.8779369] [Medline: 31374713]
22. Bacon-Bastable S. Nurse as Educator: Principles of Teaching and Learning for Nursing Practice. 4th edition. Burlington, MA: Jones & Bartlett Learning; 2014:99-101.
23. Kearney E, Shellikeri S, Martino R, Yunusova Y. Augmented visual feedback-aided interventions for motor rehabilitation in Parkinson's disease: a systematic review. Disabil Rehabil 2019 May;41(9):995-1011. [doi: 10.1080/09638288.2017.1419292] [Medline: 29316816]
24. Sharma D, Chevidikunnan MF, Khan FR, Gaowgzeh RA. Effectiveness of knowledge of result and knowledge of performance in the learning of a skilled motor activity by healthy young adults. J Phys Ther Sci 2016 May;28(5):1482-1486 [FREE Full text] [doi: 10.1589/jpts.28.1482] [Medline: 27313355]
25. Anderson D, Magill RA, Sekiya H. Motor learning as a function of KR schedule and characteristics of task-intrinsic feedback. J Mot Behav 2001 Mar;33(1):59-66. [doi: 10.1080/00222890109601903] [Medline: 11265058]
26. Basta D, Rossi-Izquierdo M, Soto-Varela A, Greters ME, Bittar RS, Steinhagen-Thiessen E, et al. Efficacy of a vibrotactile neurofeedback training in stance and gait conditions for the treatment of balance deficits: a double-blind, placebo-controlled multicenter study. Otol Neurotol 2011 Dec;32(9):1492-1499. [doi: 10.1097/MAO.0b013e31823827ec] [Medline: 22089958]
27. Zanotto D, Rosati G, Spagnol S, Stegall P, Agrawal S. Effects of complementary auditory feedback in robot-assisted lower extremity motor adaptation. IEEE Trans Neural Syst Rehabil Eng 2013 Sep;21(5):775-786. [doi: 10.1109/tnsre.2013.2242902]
28. Berger A, Horst F, Müller S, Steinberg F, Doppelmayr M. Current state and future prospects of EEG and fNIRS in robot-assisted gait rehabilitation: a brief review. Front Hum Neurosci 2019;13:172. [doi: 10.3389/fnhum.2019.00172] [Medline: 31231200]
29. Berger A, Horst F, Steinberg F, Thomas F, Müller-Eising C, Schöllhorn WI, et al. Increased gait variability during robot-assisted walking is accompanied by increased sensorimotor brain activity in healthy people. J Neuroeng Rehabil 2019 Dec 27;16(1):161 [FREE Full text] [doi: 10.1186/s12984-019-0636-3] [Medline: 31882008]
30. Wade J, Boyles R, Flemming P, Sarkar A, de Riesthal M, Withrow T, et al. Feasibility of automated mobility assessment of older adults via an instrumented cane. IEEE J Biomed Health Inform 2019 Jul;23(4):1631-1638. [doi: 10.1109/jbhi.2018.2873991]
31. Mundt M, Batista JP, Markert B, Bollheimer C, Laurentius T. Walking with rollator: a systematic review of gait parameters in older persons. Eur Rev Aging Phys Act 2019;16:15 [FREE Full text] [doi: 10.1186/s11556-019-0222-5] [Medline: 31528238]
32. Jacobs E, McCrum C, Senden R, van Rhijn LW, Meijer K, Willems PC. Gait in patients with symptomatic osteoporotic vertebral compression fractures over 6 months of recovery. Aging Clin Exp Res 2020 Mar;32(2):239-246 [FREE Full text] [doi: 10.1007/s40520-019-01203-9] [Medline: 31030421]
33. Carroll M, Boocock M, Dalbeth N, Stewart S, Frampton C, Rome K. Ankle joint function during walking in tophaceous gout: A biomechanical gait analysis study. Gait Posture 2018 Jun;63:150-153. [doi: 10.1016/j.gaitpost.2018.04.020] [Medline: 29734124]
34. Rubin G, Monder O, Zohar R, Oster A, Konra O, Rozen N. Toe-touch weight bearing: myth or reality? Orthopedics 2010 Oct 11;33(10):729. [doi: 10.3928/01477447-20100826-02] [Medline: 20954673]
35. Ridao-Fernández C, Ojeda J, Chamorro-Moriana G. Analysis of spatial and temporal step parameters during crutch-assisted gait as a dual-task: a pilot study. J Biomech Eng 2018 Oct 01;140(10):101006. [doi: 10.1115/1.4040020] [Medline: 30029235]
36. LaPrade R, DePhillipo NN, Cram TR, Cinque ME, Kennedy MI, Dornan GJ, et al. Partial controlled early postoperative weightbearing versus nonweightbearing after reconstruction of the fibular (lateral) collateral ligament: a randomized

controlled trial and equivalence analysis. Am J Sports Med 2018 Aug;46(10):2355-2365. [doi: 10.1177/0363546518784301] [Medline: 30015510]
37. Feger M, Glaviano NR, Donovan L, Hart JM, Saliba SA, Park JS, et al. Current trends in the management of lateral ankle sprain in the United States. Clin J Sport Med 2017 Mar;27(2):145-152. [doi: 10.1097/JSM.0000000000000321] [Medline: 27347860]
38. Krych A, Reardon PJ, Johnson NR, Mohan R, Peter L, Levy BA, et al. Non-operative management of medial meniscus posterior horn root tears is associated with worsening arthritis and poor clinical outcome at 5-year follow-up. Knee Surg Sports Traumatol Arthrosc 2017 Mar;25(2):383-389. [doi: 10.1007/s00167-016-4359-8] [Medline: 27761625]
39. Ridao-Fernández C, Chamorro-Moriana G, Ojeda J. Influence of the load exerted over a forearm crutch in spatiotemporal step parameters during assisted gait: pilot study. Biomed Eng Online 2018 Jul 18;17(1):98 [FREE Full text] [doi: 10.1186/s12938-018-0527-z] [Medline: 30021612]
40. Schröter S, Nakayama H, Yoshiya S, Stöckle U, Ateschrang A, Gruhn J. Development of the double level osteotomy in severe varus osteoarthritis showed good outcome by preventing oblique joint line. Arch Orthop Trauma Surg 2019 May;139(4):519-527. [doi: 10.1007/s00402-018-3068-9] [Medline: 30413943]
41. Leibold C, Schmaranzer F, Siebenrock KA, Steppacher SD. Femoral osteotomies for the treatment of avascular necrosis of the femoral head. Oper Orthop Traumatol 2020 May;32(2):116-126. [doi: 10.1007/s00064-019-00642-x] [Medline: 31784775]
42. Chamorro Moriana G, Roldán JR, Rejano JJJ, Martínez RC, Serrano CS. Design and validation of GCH System 1.0 which measures the weight-bearing exerted on forearm crutches during aided gait. Gait Posture 2013 May;37(4):564-569. [doi: 10.1016/j.gaitpost.2012.09.018] [Medline: 23218725]
43. Chamorro-Moriana G, Moreno AJ, Sevillano JL. Technology-based feedback and its efficacy in improving gait parameters in patients with abnormal gait: a systematic review. Sensors (Basel) 2018 Jan 06;18(1):142 [FREE Full text] [doi: 10.3390/s18010142] [Medline: 29316645]
44. Osaki Y, Kunin M, Cohen B, Raphan T. Relative contribution of walking velocity and stepping frequency to the neural control of locomotion. Exp Brain Res 2008 Mar;185(1):121-135. [doi: 10.1007/s00221-007-1139-1] [Medline: 17952421]
45. Hernigou P, Giber D, Dubory A, Auregan JC. Safety of simultaneous versus staged bilateral opening-wedge high tibial osteotomy with locked plate and immediate weight bearing. Int Orthop 2020 Jan;44(1):109-117. [doi: 10.1007/s00264-019-04385-z] [Medline: 31385014]
46. Gama G, Celestino ML, Barela JA, Barela AMF. Gait initiation and partial body weight unloading for functional improvement in post-stroke individuals. Gait Posture 2019 Feb;68:305-310. [doi: 10.1016/j.gaitpost.2018.12.008] [Medline: 30553152]
47. Slavens B, Bhagchandani N, Wang M, Smith PA, Harris GF. Motion analysis of the upper extremities during Lofstrand crutch-assisted gait in children with orthopaedic disabilities. J Exp Clin Med 2011 Oct;3(5):218-227. [doi: 10.1016/j.jecm.2011.09.010]
48. Yeung E, Chow DHK, Su IYW. Kinematic and electromyographic studies on unaided, unilateral and bilateral crutch walking in adolescents with spastic diplegia. Prosthet Orthot Int 2012 Mar;36(1):63-70. [doi: 10.1177/0309364611429722] [Medline: 22130909]
49. Zalech M, Bujak Z. Precision of verbal feedback affects accuracy of motor skill performance. J Sports Med Phys Fitness 2018 Apr;58(4):435. [doi: 10.23736/s0022-4707.17.06749-4]
50. Boehm W, Gruben KG. Development of KIINCE: A kinetic feedback-based robotic environment for study of neuromuscular coordination and rehabilitation of human standing and walking. J Rehabil Assist Technol Eng 2018;5:2055668318793585 [FREE Full text] [doi: 10.1177/2055668318793585] [Medline: 31191950]
51. Zhang Y, Nolan KJ, Zanotto D. Immediate effects of force feedback and plantar somatosensory stimuli on inter-limb coordination during perturbed walking. IEEE Int Conf Rehabil Robot 2019 Jun;2019:252-257. [doi: 10.1109/ICORR.2019.8779565] [Medline: 31374638]
52. Martin K, Hoover D, Wagoner E, Wingler T, Evans T, O'Brien J, et al. Development and reliability of an observational gait analysis tool for children with Down syndrome. Pediatr Phys Ther 2009;21(3):261-268. [doi: 10.1097/PEP.0b013e3181b13bca] [Medline: 19680068]
53. Anukoolkarn K, Vongsirinavarat M, Bovonsunthonchai S, Vachalathiti R. Plantar pressure distribution pattern during mid-stance phase of the gait in patients with chronic non-specific low back pain. J Med Assoc Thai 2015 Oct;98(9):896-901. [Medline: 26591401]
54. Eguchi R, Yorozu A, Fukumoto T, Takahashi M. Estimation of vertical ground reaction force using low-cost insole with force plate-free learning from single leg stance and walking. IEEE J Biomed Health Inform 2020 May;24(5):1276-1283. [doi: 10.1109/jbhi.2019.2937279]
55. Bella GP, Rodrigues NBB, Valenciano PJ, Silva LMAE, Souza RCT. Correlation among the visual gait assessment scale, Edinburgh Visual Gait Scale and Observational Gait Scale in children with spastic diplegic cerebral palsy. Rev Bras Fisioter 2012 May;16(2):134-140. [Medline: 22584771]
56. Wellmon R, Pezzillo K, Eichhorn G, Lockhart W, Morris J. Changes in dual-task voice reaction time among elders who use assistive devices. J Geriatr Phys Ther 2006;29(2):74-80. [doi: 10.1519/00139143-200608000-00006] [Medline: 16914064]

57. Muir-Hunter S, Montero-Odasso M. The attentional demands of ambulating with an assistive device in older adults with Alzheimer's disease. Gait Posture 2017 May;54:202-208. [doi: 10.1016/j.gaitpost.2017.03.011] [Medline: 28359928]
58. Ramakrishnan T, Kim SH, Reed KB. Human gait analysis metric for gait retraining. Appl Bionics Biomech 2019;2019:1286864. [doi: 10.1155/2019/1286864] [Medline: 31814843]
59. Fasano A, Bloem B. Gait disorders. Continuum (Minneap Minn) 2013 Oct;19(5):1344-1382. [doi: 10.1212/01.CON.0000436159.33447.69] [Medline: 24092293]
60. van Lieshout R, Pisters MF, Vanwanseele B, de Bie RA, Wouters EJ, Stukstette MJ. Biofeedback in partial weight bearing: usability of two different devices from a patient's and physical therapist's perspective. PLoS One 2016;11(10):e0165199 [FREE Full text] [doi: 10.1371/journal.pone.0165199] [Medline: 27798674]
61. Pizzolato C, Reggiani M, Saxby DJ, Ceseracciu E, Modenese L, Lloyd DG. Biofeedback for gait retraining based on real-time estimation of tibiofemoral joint contact forces. IEEE Trans Neural Syst Rehabil Eng 2017 Sep;25(9):1612-1621. [doi: 10.1109/tnsre.2017.2683488]
62. Ebert J, Edwards PK, Fallon M, Ackland TR, Janes GC, Wood DJ. Two-year outcomes of a randomized trial investigating a 6-week return to full weightbearing after matrix-induced autologous chondrocyte implantation. Am J Sports Med 2017 Mar;45(4):838-848. [doi: 10.1177/0363546516673837] [Medline: 27881381]

Abbreviations

CHAGS: Chamorro Assisted Gait Scale
KP: knowledge of performance
KR: knowledge of results

Edited by R Kukafka; submitted 02.02.21; peer-reviewed by R Martín-Valero, B Smith, V Shah; comments to author 10.05.21; revised version received 24.06.21; accepted 27.07.21; published 22.09.21.

Please cite as:
Chamorro-Moriana G, Sevillano JL, Perez-Cabezas V
Versatile GCH Control Software for Correction of Loads Applied to Forearm Crutches During Gait Recovery Through Technological Feedback: Development and Implementation Study
J Med Internet Res 2021;23(9):e27602
URL: https://www.jmir.org/2021/9/e27602
doi:10.2196/27602
PMID:34550073

Original Paper

Exploring the Use of Genomic and Routinely Collected Data: Narrative Literature Review and Interview Study

Helen Daniels[1], BSc, PhD; Kerina Helen Jones[1], BSc, PhD; Sharon Heys[1], BSc; David Vincent Ford[1], MBA

Population Data Science, Swansea University, Swansea, United Kingdom

Corresponding Author:
Helen Daniels, BSc, PhD
Population Data Science, Swansea University
Singleton Park
Swansea, SA2 8PP
United Kingdom
Phone: 44 01792606572
Email: h.daniels@swansea.ac.uk

Abstract

Background: Advancing the use of genomic data with routinely collected health data holds great promise for health care and research. Increasing the use of these data is a high priority to understand and address the causes of disease.

Objective: This study aims to provide an outline of the use of genomic data alongside routinely collected data in health research to date. As this field prepares to move forward, it is important to take stock of the current state of play in order to highlight new avenues for development, identify challenges, and ensure that adequate data governance models are in place for safe and socially acceptable progress.

Methods: We conducted a literature review to draw information from past studies that have used genomic and routinely collected data and conducted interviews with individuals who use these data for health research. We collected data on the following: the rationale of using genomic data in conjunction with routinely collected data, types of genomic and routinely collected data used, data sources, project approvals, governance and access models, and challenges encountered.

Results: The main purpose of using genomic and routinely collected data was to conduct genome-wide and phenome-wide association studies. Routine data sources included electronic health records, disease and death registries, health insurance systems, and deprivation indices. The types of genomic data included polygenic risk scores, single nucleotide polymorphisms, and measures of genetic activity, and biobanks generally provided these data. Although the literature search showed that biobanks released data to researchers, the case studies revealed a growing tendency for use within a data safe haven. Challenges of working with these data revolved around data collection, data storage, technical, and data privacy issues.

Conclusions: Using genomic and routinely collected data holds great promise for progressing health research. Several challenges are involved, particularly in terms of privacy. Overcoming these barriers will ensure that the use of these data to progress health research can be exploited to its full potential.

(J Med Internet Res 2021;23(9):e15739) doi:10.2196/15739

KEYWORDS

genomic data; routine data; electronic health records; health data science; genome; data regulation; case study; eHealth

Introduction

Background

The progression of genomics in the last few decades has been remarkable. Since 2001, when the Human Genome Project mapped and sequenced virtually every gene in the human genome, genetic sequencing technology has advanced rapidly in both the public and private domains. Next-generation sequencing costs have plummeted by almost 100%, and research opportunities have grown exponentially as a result [1]. For example, a simple search in the medical database, PubMed, shows that research on genomics has more than quadrupled since 2000, from around 340,000 published articles on this topic growing to 1.5 million by 2020. This increase has translated into quicker diagnoses, better outcomes, and more effective health care for patients [2-4]. Great strides have been made in cancer research, for example, where patients are now being treated according to their own or the tumor's genomic data [5].

Being able to use genomic data in conjunction with routinely collected data holds even greater potential to advance knowledge by including factors wider in scope. Precision medicine requires that novel correlations of genotype, phenotype, and the environment be identified to inform new methods for diagnosing, treating, and preventing disease in a way that is responsive to the individual [5]. Knowledge of gene-environment interactions can also contribute on a population level by informing health and public health services in areas such as service planning, population genetic testing, disease prevention programs, and policy development [6]. Routinely collected data, electronic health records (EHRs) in particular, already hold vast amounts of clinical and environmental information on large numbers of people and preclude the need for lengthy and expensive data collection. Adding these phenotypic data to knowledge about a person's genome can elucidate new knowledge, such as that about gene-environment and gene-drug interactions, and can thus provide a richer understanding of health and disease [7].

Increasing the use of genomic data for health research is a high government priority to understand and address the causes of disease. The potential of integrating genomic and routine data sets has been recognized in the United Kingdom by the Welsh Government [8] via their *Genomics for Precision Medicine Strategy* and by Genomics England [9] with the inception of the 100,000 genome project in 2018, both of whom are investigating ways to link genomic data and EHRs. From a more international perspective, the former president of the United States, Barack Obama, launched the Precision Medicine Initiative to improve individualized care by combining genomic data and EHRs with diet and lifestyle information from US citizens [10]. The UK Chief Medical Officer, Dame Sally Davies's "genomic dream" of mainstreaming genomic medicine into National Health Service (NHS) standard care is becoming ever closer, which means that data from a person's genome will likely be directly recorded into EHRs, making this type of research far more accessible [11].

Objective

As this field prepares to move forward, it is important to take stock of the current state of play in order to highlight new avenues for development, identify challenges, and ensure that adequate data governance models are in place for safe and socially acceptable progress. Previous work has examined the benefits and logistical challenges of integrating genomic and routinely collected data in health care practice, but less is known of this specifically in a research setting [6,12,13]. Therefore, our objective is to add to this literature by conducting a narrative literature review and a series of interviews that would provide an outline of the use of genomic data alongside routinely collected data in health research to date. It focuses on the types of data that have been used, the role of routinely collected data in these studies, the data sources, how researchers access the data, and the challenges surrounding their use. This will inform further work in developing a framework for working with genomic and routinely collected data [14].

Methods

Literature Review

First, we conducted a literature search of research that used genomic data in conjunction with routinely collected data. We define routinely collected data as data collected as a matter of course and not specifically for research [8]. Genomic data refer to the data generated after processing a person's genome, in full or in part, for example, by sequencing [7]. Studies were eligible for inclusion if they had used both types of data in combination to answer a health research question. We included studies of any design published in either peer-reviewed journals or gray literature in the English language.

We searched the following databases to identify these studies: PubMed, Ovid, CINAHL, OpenGrey, CENTRAL, LILACS, and Web of Knowledge from inception until January 31, 2019. We also searched for books, gray literature, and websites. We used a piloted search strategy that included keywords representing genetic and routinely collected data, and the following is the search strategy used for PubMed and modified for use with the remaining databases:

1. (Gene OR Genetic* OR Genome* OR Genomic*)
2. (*Administration record** OR Anonymised OR Anonymized OR Anonymisation OR Anonymization OR *Big data* OR *Clinical record** OR *Data linkage* OR *Data mining* OR *Data science* OR *Education record** OR *Ehealth* OR EHR OR *Electronic data* OR *Health record** OR *Housing record** OR Encrypt* OR *Insurance* OR *Linked data* OR *Medical record** OR *Patient record** OR *Prison record** OR *Publically available* OR *Publicly available* OR Register OR Registry OR Registries OR *Routine data* OR *Routinely collected* OR *Safe haven*)
3. #1 AND #2

This search in PubMed resulted in more than 50,000 hits. After initial pilot screenings, for example, by restricting to only publications within the last 10 years, it was clear that, given the number and heterogeneity of potentially relevant articles, neither a systematic review nor a meta-analysis would be practical. We took a pragmatic approach by scanning these articles to retrieve information on the following items until we reached data saturation, when no new information appeared in the text: types of genomic data; types and roles of routinely collected data; data sources; and data access models, that is, how researchers access the data. We chose examples from each criterion to ensure that we included a range of health conditions and presented them in a narrative format, which followed the format of the criteria given above.

Interviews

To understand the use of genomic and routinely collected data in context, we recruited a purposive sample of individuals who have been involved in leading research projects using a combination of genomic and routinely collected data. We identified potential participants from the literature search outlined above and sent 19 interview invitations via email. Reminders were sent after 2 weeks, and if there was no response, we made no further contact. In total, 11 individuals agreed to

participate in either an in-person or teleconferencing interview, depending on their geographical location. Participants were involved with the following projects: the Swansea Neurology Biobank (Swansea, Wales); Dementia Platform UK (Oxford, England); PsyCymru (Swansea, Wales); UK Biobank (Stockport, England); BC Generations Project (British Columbia, Canada); the Province of Ontario Neurodevelopmental Disorders (POND) Network, IC/ES (Ontario, Canada); Electronic and Medical Records and Genomics (eMERGE) Network (Vanderbilt, United States); and the Sax Institute's 45 and Up study (New South Wales, Australia). We omitted any further information on the participants to maintain their anonymity. We developed interview questions with our advisory board: a group of UK geneticists and data scientists who were interested in using genomic data and our discussions centered around these:

1. What is the purpose of integrating the genetic data with health data?
2. What types of genetic data are being included?
3. Were there particular approvals you had to obtain? And if yes, what were these?
4. What were the main challenges encountered?
5. How did you address the challenges?
6. What is your main model for storing these data?
7. What access model(s) do you use? For example, safe room only, remote access, data released externally to researchers;
8. What are the conditions for access to data?

We followed up interviews by email if any answers needed clarification.

Results

Textbox 1 provides a summary of the results of the literature review and interviews. Multimedia Appendix 1 Table S1 [15-32] provides a detailed summary of the results.

We included 19 studies in this literature review that provided broad examples of the different types of genomic and routinely collected data that can be used together to answer a health-related research question. The countries where this research was based were the United Kingdom [15-20], China [21], United States [22-29], Canada [19,30,31], and Australia [32].

Textbox 1. Summary of results.

Type of genomic data

- Single nucleotide polymorphisms
- Polygenic risk scores
- Gene activity scores
- DNA methylation status

Purpose of combining data

- Genome-wide association studies
- Phenome-wide association studies
- Longitudinal studies
- Candidate gene studies
- Gene profiling studies
- Exploratory studies

Types of routinely collected data

- Electronic health records
- Disease registry data
- Disease registries
- Mortality registers
- Deprivation indices
- Health insurance

Role of routinely collected data

- Identifying cases and controls
- Baseline data
- Deep phenotyping
- Long-term follow-up
- Sociodemographic information

Sources of data

- Databanks
- Biobanks

Governance models for data access

- Publicly available on the web
- Released to researchers
- Data safe havens

Challenges

- Data collection
- Data storage and costs
- Technical and/or software issues
- Data privacy and data protection laws

Types of Genomic Data

Examples found of the types of genomic data used in these studies included single nucleotide polymorphisms (SNPs), gene activity scores, and DNA methylation status. The most frequently used were SNPs [16,17,19,23-29], which represent a single base-pair change in the DNA sequence and are highly

granular; hence, they are popular in health research [33]. Research studies tend to refer to SNPs by their reference SNP number, a unique identifier given to a SNP, or a cluster of SNPs by the National Center for Biotechnology (NCBI) [34].

Polygenic risk scores are closely related to SNPs and were used in 2 of the studies included in this review [15,16]. These predict a person's likelihood of developing a particular disease based on the cumulative effect of a number of genetic variants [15]. Our literature search found other instances of quantitative measures of gene activity combined with routine data, including the 21-gene Recurrence Score, which measures the activity of 21 genes (16 cancer-related and 5 reference) for patients with breast cancer [33] and the enzyme activity score for the *CYP2D6* (cytochrome P450 family 2 subfamily D member 6) gene, which codes for an important drug-metabolizing enzyme, and is highly polymorphic in humans [34]. Rusiecki et al [29] measured changes in DNA methylation status in their participants to investigate whether this predicted posttraumatic stress disorder in US military service members.

Purpose of Combining Genomic and Routine Data

Conducting genome-wide association studies (GWAS) [15,17-19,21], phenome-wide association studies (PheWAS) [24,25], and a combination of both [26,27] were the main purposes of combining genomic and routine data. Both these methods use powerful statistical techniques to find associations between genetic variants (SNPs) and phenotypes, which can then be used to predict the genetic risk factors of disease, levels of gene expression, and even social and behavioral characteristics such as educational attainment, impulsivity, and recreational drug experimentation [35,36]. Given the number of tested associations, these studies require large sample sizes to yield enough statistical power to detect differences in genetic variation between cases and controls [37]. Other study designs included longitudinal studies [16,18,30-32], a candidate gene study [23], case control studies [19,29], a gene profiling study [22], and an exploratory study [28]. Each of these studies was designed either to identify genetic risk factors of the disease or to investigate drug safety or efficacy (Multimedia Appendix 1 Table S1).

Types and Role of Routinely Collected Data

EHRs appear to be, by far, the most common form of routinely collected data used in the studies identified in this review. For the type of research discussed here, data in EHRs have been used to identify eligible participants, for phenotyping, and to provide long-term follow-up on specific health outcomes. These data are collected as a matter of course in health care systems, and depending on their country of origin, EHR content can vary, although usually this digital record will include the patient's name, address, demographics, medical history, care preferences, lifestyle information (such as diet, exercise, and smoking status), and free-text notes [38]. An example of EHR data used in the identified studies was the International Classification of Disease (ICD) coding. This allows clinicians to record the status of a patient in a standardized way, whether it is for disease, disorder, injury, infection, or symptoms [39]. Hebbring et al [26] used ICD codes in EHRs to identify appropriate cases and controls for their PheWAS study of the *HLA-DRB1*150* gene involved in immune regulation. Rusiecki et al [29] used ICD codes to identify cases with a postdeployment diagnosis of posttraumatic stress disorder in their study on the link between gene expression and posttraumatic stress disorder in US military service members.

EHRs were also used to describe and validate a phenotype of interest, which is particularly important for PheWAS studies. The large number of phenotypes used in these analyses need to be clearly defined to ensure that any associations made with genetic variants are precise and replicable [40]. For example, Breitenstein et al [23] used EHRs to define the type 2 diabetes phenotype needed for their candidate gene study based on diagnoses, medications, and laboratory tests. The algorithm used to achieve this was developed by the eMERGE Network (see below for more details on this organization) and has been used successfully many times since [41].

In addition to providing a baseline snapshot of the patient, EHRs are also longitudinal in nature, and this makes them ideally placed to provide long-term follow-up to study participants. An example of this is the Genetics and Psychosis (GAP) study, which looked at a purported association between a variant of the *ZNF804A* gene and poor outcomes after first-episode psychosis [20]. Using individual-level linkage between EHRs and genotype data, this study followed the clinical outcomes for 291 patients over a period of 2 years, and subsequently found strong evidence for their hypothesis.

Other examples of routine data used in genomic research include disease registry morbidity and mortality records [20,22]. Disease registries collect information on clinical outcomes and care for a specific patient population over time. EHRs often feed data into these registries, but registry data can also include patient-reported outcomes and other biometric data, and therefore provide a more holistic view of the patient than an EHR would in isolation, for example, the UK MS Register [42]. Routine data sets need not be individual-level or person-based to prove useful in genetic research. A case in point is the Scottish Index of Multiple Deprivation, which gives an indication of a geographical area's socioeconomic deprivation based on data about employment, income, health, education, housing, crime, and access to services [43]. Clarke et al [15] were able to link participants' postcodes to the Scottish Index of Multiple Deprivation and generate socioeconomic deprivation variables to investigate their association with polygenic risk scores for alcohol dependence.

Sources of Genomic and Routinely Collected Data

Overview

Although individual research projects often collect biological samples and generate their own genomic data for combination with routinely collected data [19,29], studies can also make use of the many sources of genomic and routinely collected data already available. Our interview participants spoke to us about the different data sources they used in their research and provided details on the use of these data and the participants of their projects. From this information, there seem to be two general categories of these sources: databanks and biobanks, the former where only the data are stored, and the latter, which

store both biological samples and genomic data. Below are illustrative examples; therefore, this is not an exhaustive list.

Dementias Platform UK

Dementias Platform UK (DPUK) [16] is a data portal funded by the Medical Research Council that hosts the data of 2 million people from over 40 cohorts relevant to dementia research. Combining these data enhances the individual research power of each study and brings together knowledge from a number of stakeholders to facilitate and accelerate new discoveries. Examples of these cohorts are the GENetic Frontotemporal Dementia Initiative (GENFI; genotype data for *GRN*, *MAPT,* and *C9ORF72* genes) and the Genetic and Environmental Risk in Alzheimer's Disease (GERAD) Consortium (whole exome sequences) [16]. Presently, DPUK provides EHR linkage for Welsh participants via the Secure Anonymized Information Linkage (SAIL) databank [44].

UK Biobank

UK Biobank [45] is a national resource with data from 500,000 participants aged between 40 and 69 years who have donated blood, urine, and saliva samples, have undergone a number of baseline measures, and have provided detailed health and behavioral information about themselves. Genomic data are available for 488,000 participants and comprise SNPs, genotypes, and haplotypes. The UK Biobank holds a number of routine data sets, including hospital inpatient episodes, cancer registrations, and deaths. Studies using the UK Biobank genomic and routinely collected data include genome-wide meta-analyses of depression [17] and identifying candidate gene and disease associations that could help predict adverse drug reactions [26].

Personal Genome Project

The Personal Genomes Project (PGP) is a databank founded in 2005 at Harvard University and now extends worldwide. It provides a web-based platform for individuals (over 100,000 people to date) to share their genomic data publicly, along with their EHRs, and other trait information to progress science without many of the governance restrictions of traditional research. Most of the genomic data on the PGP database are in SNP format, although files of raw sequence data are available for some participants [43].

eMERGE Network

The eMERGE Network is a consortium of American medical institutions whose goal is to use EHR data in combination with a variety of genomic data types to advance translational research. The network also releases its genomic data, including GWAS, whole genome, and whole exome sequence data, along with a subset of phenotypic elements to the broader community of researchers via the dbGaP—an NCBI [46] database of genotypes and phenotypes. Through this mechanism, any research project can be applied to the uploaded data. A wealth of publications have resulted from eMERGE's work; these are available to view on the web [24,25,47].

POND Network

The POND network is an IC/ES [31] initiative based in Ontario, Canada, and involves a cohort of children and young people (the total number of the sample is approximately 3000) with a neurological development disorder, with a particular focus on autism. This is a highly phenotyped cohort with all participants having undergone multiple clinical tests, including those for attention-deficit/hyperactivity disorder, obsessive-compulsive disorder, and family history and demographic data. A subset (n=667) provided consent for linkage with administrative data held in IC/ES via health card numbers. The aim is to identify subgroups of autism based on the co-occurrence of other developmental conditions, other comorbidities, and health service use, and to characterize these groups based on clinical attributes and genomics. The network data are project-level only, although it is anticipated that linkage of genomic data to administrative and health data will become more routine in the future.

BC Generations Project

The BC Generations Project [48] is British Columbia's (BC) largest ever health study and is part of a national initiative—the Canadian Partnership for Tomorrow Project—to aid researchers in answering questions about how environment, lifestyle, and genes contribute to cancer and other chronic diseases. Almost 30,000 participants were involved in the project, and they provided baseline information about their health, diet, lifestyle, and medical and family history. Many have also donated blood and urine samples, and the type of genomic data generated from these samples is based on the needs of the researchers (subject to approval). The BC cohort is one of several provincial cohorts that can be combined for national studies or can be linked to provincial administrative data via PopDataBC [49].

Sax Institute and the 45 and Up Study

Based in Australia, the main business of the Sax Institute is to manage the *45 and up* study, which has been following 260,000 people for over 12 years who provide both routinely collected and self-reported health data [32]. The aim of this study was to collect samples from 50,000 individuals for full genome sequencing. The Sax Institute has a partnership with the Garvan Institute of Medical Research [50], which acts as a genome sequencing facility and makes data available for research subject to approval. The Garvan Institute retains all the genomic data, but with data linkage to the Sax Institute.

Data Governance Access Models

Overview

We surveyed the data access models and information governance systems of genomic and routine data sources identified by our literature search and interviews. From the additional information given to us by our interview participants, we were able to categorize data access models as follows: (1) publicly available on the web, (2) released to researchers, or accessed via a (3) data safe haven.

Publicly Available on the Web

There are a wealth of free, genomic data sources on the web, made available because of individual projects or from the pooled results of a variety of different projects. These data are either downloadable or viewable on the web. For this type of resource, data tend to be at the gene or variant level (eg, GIANT Consortium [51]; GWAS Catalog [52]), but some do hold

individual-level data, eg, Database of Genomic Variants [53]. The Personal Genome Project, described above, is the only data source that provides identifiable, individual-level genomic data [43]. All other data were deidentified.

Released to Researchers

Currently, the most common way for genomic data to be accessed for research is through secure electronic file transfer to the researcher. This generally occurs after successful application to an internal review board (IRB) and signed data use agreement (DUA). All of the published research studies identified by our literature search used this model, unless the genomic data were generated specifically for that particular project. For instance, Cronin et al [24] received genomic data regarding 54 SNPs, and also demographic, vital sign, and billing data derived linked EHRs by the eMERGE network. Other biobanks using this model include Mayo Genome Consortia [23], China Kadoorie Biobank [21], UK Biobank [17,26], BioVu [24], and Generation Scotland [17], although access to BioVu data is only granted to Vanderbilt faculty members [54]. In addition to IRB and DUA procedures, the use of individual-level EHR data supplied by Generation Scotland also requires an application to the NHS Research Ethics Committee [55].

Data Safe Havens

The use of large-scale population data in health research has become increasingly popular in the last few decades, and this has seen the evolution of data safe havens as a way to ensure its safe and secure use. Lea et al [56] defined data safe havens as a system that invokes procedural, technical, and physical controls, including access to data within a secure environment (rather than data release), in order to safeguard the identities of people providing the data. Despite there being a greater tendency in the literature for releasing data to researchers, we have seen from our work with interview participants, a trend toward using a data safe haven system for both younger and more well-established organizations.

The PsyCymru study [18] and the Swansea Neurology Biobank [19] have deposited polygenic risk scores and SNPs, respectively, into the SAIL Databank [44], a data safe haven based at Swansea University. The SAIL Databank provides remote access to many linkable anonymized data sets, and both of these studies have used SAIL to link their genetic data and other phenotypic data to EHRs and other routinely collected data. These data are available only for project access, and currently cannot be shared. However, their intention is to make linked genetic and routinely collected data available for research in the near future [57,58].

The Sax Institute operates in a similar way to the SAIL Databank in that it provides access to genomic and health data via a virtual lab by remote access anywhere in the world. Data access requires approval by two data access committees: one at the Garvan Institute and one at the Sax Institute. The BC Generations Project said, “should the researcher access the project’s data via PopDataBC then they would only be allowed to use the data on within the secured research environment” (Participant 5, BC Generations Project).

For the UK Biobank, the current *modus operandi* releases anonymized data externally to researchers. However, they stated that this is unlikely to be sustainable because genomic data files are too large. The UK Biobank will likely be changing to a remote access model in the near future. The eMERGE Network also confirmed that they are experimenting with remote access into a data safe haven. One of our participants who used IC/ES data explained that “as the data are considered highly sensitive, access is only by an ICES analyst, with results provided to the project lead” (Participant 4, IC/ES).

Challenges

Combining genomic and routine data does not come without its challenges. Our participants, who are currently working with these data, spoke to us about these different challenges during their interviews, and their experiences are summarized below.

Data Collection

A challenge described by several participants concerned data collection. Conducting long-term follow-up over long periods and at regular intervals means that participants need to be invited and reconsented to provide more blood and other health information. In addition, poor quality sequence alignment could render the samples useless, and with repeated use by researchers, blood samples will eventually become exhausted. Each of these issues necessitates a lengthy and expensive process of resampling thousands of individuals and imposes a burden on participants. A possible solution to this, as one participant suggested, is to “join up with others biobanks in order to work towards epidemiologically valid sample sizes” (Participant 1, 45 and Up Study)

Data Storage and Costs

Genomic data are huge, approximately 90 GB for the raw data of 1 whole genome, and this is often the source of technical issues surrounding its use [59]. One participant told us that their main challenge was data storage capacity and that “it may no longer be cost or space efficient for organisations to hold multiple datasets” (Participant 2, UK Biobank). They go on to explain that there is certainly a case to be made for storing genomic data as VCF files only, which keep record only of gene sequence variations and are much smaller and easier to work with. However, this restricts the type of analyses possible, particularly in the advent of new discoveries about the anatomy of the genome. There may be a need to find solutions regarding storage space for raw genomic data and for specialist platforms required to conduct analyses.

Technical and/or Software Issues

Many of the challenges faced by participants working with these data are technical in nature. Participant 3 (eMERGE) described creating a sequencing platform from the ground up, which was much longer than initially projected. Furthermore, they described that multiple sequencing centers needed harmonization, as well as the needs across sites and projects for network-wide data collection. Another participant found difficulties surrounding analysis software and analyst capacity, since this is a very specialized skill (Participant 4, IC/ES). Researchers often conduct genetic data analysis using publicly available, downloadable software applications, which are subject to

frequent updates for improvements. This is a challenge to incorporate this software and keep it up to date. Several other participants spoke of similar software issues. Potential solutions discussed during our interviews were to upscale and have dedicated servers for genomic data, as well as to install specialist toolsets within the secure research environment.

Data Privacy and Data Protection Laws

Given the possibility of identifying individuals from genomic data [60], privacy is of primary concern. We were told that, for one research project using IC/ES data (Participant 4), the privacy approval group was concerned about identifiability due to the genomic data being unique, particularly where there are rare variants. This means that the genomic data contained in variant call files have been brought into the databank to show the feasibility of transfer but has to be integrated into the analytical platform.

The General Data Protection Regulation [61] in the United Kingdom states that if researchers are to rely on the lawful basis of consent to process medical genomic data, the reasons for doing so must be described in an explicit and transparent way. The rapid decrease in the cost of whole genome sequencing in the last few decades has opened many new avenues for genetic research, but this means that it is impossible to predict what this research will actually look like in the future [55]. One participant felt that this means they would need to seek consent from the participant for each new research proposal [62], rather than just once when their sample was taken. They expressed concerns that "continued and lengthy re-contact with participants was not only costly and difficult, it may also be invasive and burdensome" (Participant 11, Swansea Neurology Biobank). Also mentioned was Canada's antidiscrimination laws (Bill S-201: Genetic Nondiscrimination Act) to avoid prejudice on the grounds of genetics (including insurance and employability) [63] (Participant 5, BC Generations Project). Several participants believed that the public acceptability of their work was important, but they "weren't sure how to go about ensuring it" (Participant 7, Dementia Platform UK).

Discussion

Principal Findings

Using genomic and routinely collected data holds great potential for health research. The genomic data we identified in our literature review included SNPs, polygenic risk scores, and gene activity scores. Routine data primarily consisted of EHRs, but we did find other routine data types including registry data and deprivation indices that had been combined with genomic data. This paper shows how genomic research has progressed in recent years—from basic GWA and PheWA studies—to more complex methodologies in which health records are linked to genomic data at the individual level. Associations between genetic variants and phenotypes, identification of drug targets, knowledge about drug toxicity, and effectiveness can all be studied from the combination of genomic and routine data and leveraged for public benefit.

The EHRs created during routinely collected care provide the large sample sizes needed for GWAS analysis without the need for costly and time-consuming prospective data collection. These larger samples allow for greater validity, generalizability, and yield adequate statistical power, while also minimizing participant burden and reducing attrition during follow-up [64]. In addition, these data sets are not limited to a circumscribed number of phenotypes in the way that a traditional research study might be [40]. This richness and diversity of EHR data means that the large number of phenotypes used in PheWAS can be very clearly defined (referred to as deep phenotyping), which enhances the accuracy of phenotype-genotype associations [40].

Most importantly, with regard to EHRs, perhaps, is the additional information that routinely collected data provide about an individual's environment. This can include lifestyle factors, education, work history, pollution, and even traumatic events. Genetic determinism has long been rejected, and we now widely accept the powerful influence that lifestyle and the environment have on the way that our genes are expressed [65-67]. Precision medicine requires novel correlations not only between genotype and phenotype to be made but also with an individual's environment to inform new methods for diagnosing, treating, and preventing disease in a way that is tailored to that individual. Linking genomic data to routine data promises to elucidate important findings for precision medicine research, which will enable researchers to understand the relationship between an individual's genome and their complete life course [68]. More of these types of studies are needed for precision medicine to reach its full potential, to make the intricate genotype-phenotype associations needed to advance the understanding and treatment of disease.

From our interviews, we identified 3 main ways that researchers can access genomic and routine data: publicly available on the web, released to researchers, and via a data safe haven. We also observed that biobanks and databanks seem to be moving toward renouncing a data release model and instead favoring a data safe haven approach. Aside from potentially solving data storage issues, this will also help assuage privacy and governance concerns. Reidentification and disclosure from genomic data is possible [69,70] and can lead to many undesirable consequences: discrimination by health or life insurance companies, societal stigma, and the discovery of a genetic predisposition to a condition when one does not want to be told. This is complicated further because of the familial nature of genes, and disclosure could cause some devastating effects for biological relatives as well [71-73].

Despite this, we must keep in mind that simply because genomic data are unique, this in itself does not render it identifiable. Rather, the reidentification risk of genomic data depends on the way they are accessed, the type of analyses that are conducted, and the format in which the results are finally published [70]. This means that decisions to use certain data access models based on practicality and decreased costs are not sufficient. Genomic research relies heavily on human participation, and the public should be consulted to inform the way in which their data are accessed [74]. There is a plethora of activity taking place to consult the public about health research in general, and the success and acceptability of large-scale data research is owed, in part, to the extensive public engagement activities that

have been taking place [75,76]. Organizations such as the Global Alliance for Genomics and Health [77] are engaging with the public on the use of genomic data in research, but there is still more work to be done regarding the use of genomic and routinely collected data [78].

Limitations

This review was not intended to be exhaustive or systematic; therefore, it does not include all health research studies that have used genomic and routinely collected data. We also excluded biobanks and databanks that do not house routine data; therefore, we only include examples here. This may have resulted in inadvertently excluding some countries and institutions where this research takes place and the types of genomic and routinely collected data that have been used. Not implementing systematic methods in study selection may have introduced bias in this review's conclusions; however, the pragmatic approach used here was deemed sufficient to meet our objectives.

We included 19 studies in this review, only one of these involved data and research from a non-Western country (China) [21]. We did not come across any research that had taken place in low- to middle-income countries, although as this review was not conducted systematically, we may have unintentionally overlooked these. However, the absence of any studies from such countries may be, as Tekola-Ayele and Rotimi [79] explain, a result of having a limited number of well-trained genomic scientists and poor research infrastructure, and due to a less well-established routine data collection infrastructure and procedures such as EHRs [80].

Qualitative interviews may be subject to recruitment bias, which means that some viewpoints and experiences were excluded. In addition, there is limited information about participants; however, as this is a relatively small field of research, it was deemed necessary to maintain participants' privacy.

Conclusions

Given the projected increase in the availability of genomic data, the potential to be obtained from its combination with routine health data is vast. This review has shown examples of what has been done in this field so far with, for example, GWAS and PheWAS plus other study designs. For fields such as pharmacogenomics, these methodologies need to be used further, where using routinely collected data will simplify the process of tracking longer-term outcomes of personalized medical treatments, and elucidate new findings on the effects of the environment on drug-gene interactions. Our take away from this study is that there are several challenges involved in using these data, particularly surrounding privacy. Therefore, it is imperative that appropriate data governance be documented and that public engagement activities take place to ensure socially acceptable practices. Overcoming these barriers will ensure that the use of these data to progress health research can be exploited to its full potential.

Acknowledgments

This study was funded by the UK Medical Research Council: MC_PC_16035.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of findings.
[DOCX File , 15 KB - jmir_v23i9e15739_app1.docx]

References

1. Hayden E. Is the $1,000 genome for real? Nature. 2014. URL: https://www.nature.com/news/is-the-1-000-genome-for-real-14530 [accessed 2021-08-19]
2. England TB rates fall by third since 2012. BBC News. 2018. URL: https://www.bbc.co.uk/news/health-43517246 [accessed 2021-08-19]
3. Farnaes L, Hildreth A, Sweeney NM, Clark MM, Chowdhury S, Nahas S, et al. Rapid whole-genome sequencing decreases infant morbidity and cost of hospitalization. NPJ Genom Med 2018;3:10 [FREE Full text] [doi: 10.1038/s41525-018-0049-4] [Medline: 29644095]
4. No authors listed. Genome sequencing reveals how MRSA spreads. Pharm J 2017 Oct 26:- [FREE Full text] [doi: 10.1211/PJ.2017.20203805]
5. What is precision medicine? Genetics Home Reference. 2020. URL: https://ghr.nlm.nih.gov/primer/precisionmedicine/definition [accessed 2021-08-19]
6. Martin-Sanchez F, Maojo V, Lopez-Campos G. Integrating genomics into health information systems. Methods Inf Med 2018 Feb 07;41(01):25-30. [doi: 10.1055/s-0038-1634309]
7. Wolford BN, Willer C, Surakka I. Electronic health records: the next wave of complex disease genetics. Hum Mol Genet 2018 May 01;27(R1):R14-R21 [FREE Full text] [doi: 10.1093/hmg/ddy081] [Medline: 29547983]
8. Genomics for precision medicine strategy. Welsh Government. 2017 Jul. URL: https://gov.wales/sites/default/files/publications/2019-04/genomics-for-precision-medicine-strategy.pdf [accessed 2021-08-19]

9. The 100,000 genomes project. Genomics England. 2019. URL: https://www.genomicsengland.co.uk/about-genomics-england/the-100000-genomes-project/ [accessed 2021-08-19]
10. Davis J. Obama's next move: precision medicine and genomics venture capitalist? Healthcare IT News. 2016. URL: https://www.healthcareitnews.com/news/obamas-next-move-precision-medicine-and-genomics-venture-capitalist [accessed 2021-08-19]
11. Chief Medical Officer annual report 2016: generation genome. Department of Health and Social Care, Gov.UK. 2016. URL: https://www.gov.uk/government/publications/chief-medical-officer-annual-report-2016-generation-genome [accessed 2021-08-19]
12. Warner JL, Jain SK, Levy MA. Integrating cancer genomic data into electronic health records. Genome Med 2016 Oct 26;8(1):113 [FREE Full text] [doi: 10.1186/s13073-016-0371-3] [Medline: 27784327]
13. Kho AN, Rasmussen LV, Connolly JJ, Peissig PL, Starren J, Hakonarson H, et al. Practical challenges in integrating genomic data into the electronic health record. Genet Med 2013 Oct;15(10):772-778 [FREE Full text] [doi: 10.1038/gim.2013.131] [Medline: 24071798]
14. Jones K, Daniels H, Heys S, Lacey A, Ford DV. Toward a risk-utility data governance framework for research using genomic and phenotypic data in safe havens: multifaceted review. J Med Internet Res 2020 May 15;22(5):e16346 [FREE Full text] [doi: 10.2196/16346] [Medline: 32412420]
15. Clarke T, Smith AH, Gelernter J, Kranzler HR, Farrer LA, Hall LS, et al. Polygenic risk for alcohol dependence associates with alcohol consumption, cognitive function and social deprivation in a population-based cohort. Addict Biol 2016 Mar 10;21(2):469-480 [FREE Full text] [doi: 10.1111/adb.12245] [Medline: 25865819]
16. Welcome to the data portal. Dementias Platform UK. URL: https://portal.dementiasplatform.uk/ [accessed 2021-08-19]
17. Hall LS, Adams MJ, Arnau-Soler A, Clarke TK, Howard DM, Zeng Y, Scotland Generation, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, et al. Genome-wide meta-analyses of stratified depression in Generation Scotland and UK Biobank. Transl Psychiatry 2018 Jan 10;8(1):9 [FREE Full text] [doi: 10.1038/s41398-017-0034-1] [Medline: 29317602]
18. Lloyd K, McGregor J, John A, Craddock N, Walters JT, Linden D, et al. A national population-based e-cohort of people with psychosis (PsyCymru) linking prospectively ascertained phenotypically rich and genetic data to routinely collected records: overview, recruitment and linkage. Schizophr Res 2015 Aug;166(1-3):131-136. [doi: 10.1016/j.schres.2015.05.036] [Medline: 26139629]
19. Pickwick J. Swansea University scientists take part in largest study into the genetics of common epilepsy. Swansea University. 2017. URL: https://www.swansea.ac.uk/press-office/news-archive/2017/swanseauniversityscientiststakepartinlargeststudyintothegeneticsofcommonepilepsy.php [accessed 2021-08-19]
20. Wickramasinghe A, Tulloch AD, Hayes RD, Chang CK, Broadbent M, Di Forti M, et al. Associations between the schizophrenia susceptibility gene ZNF804A and clinical outcomes in psychosis. Transl Psychiatry 2015 Dec 15;5(12):e698 [FREE Full text] [doi: 10.1038/tp.2015.198] [Medline: 26670283]
21. Millwood IY, Bennett DA, Holmes MV, Boxall R, Guo Y, Bian Z, China Kadoorie Biobank Collaborative Group. Association of CETP gene variants with risk for vascular and nonvascular diseases among chinese adults. JAMA Cardiol 2018 Jan 01;3(1):34-43 [FREE Full text] [doi: 10.1001/jamacardio.2017.4177] [Medline: 29141072]
22. Afghahi A, Mathur M, Thompson C, Mitani A, Rigdon J, Desai M, et al. Use of gene expression profiling and chemotherapy in early-stage breast cancer: a study of linked electronic medical records, cancer registry data, and genomic data across two health care systems. J Oncol Pract 2016 Jun;12(6):e697-e709 [FREE Full text] [doi: 10.1200/jop.2015.009803]
23. Breitenstein MK, Simon G, Ryu E, Armasu SM, Weinshilboum RM, Wang LM, et al. Using EHR-linked biobank data to study metformin pharmacogenomics. Stud Health Technol Inform 2015;210:914-918 [FREE Full text] [Medline: 25991289]
24. Cronin RM, Field JR, Bradford Y, Shaffer CM, Carroll RJ, Mosley JD, et al. Phenome-wide association studies demonstrating pleiotropy of genetic variants within FTO with and without adjustment for body mass index. Front Genet 2014 Aug 5;5:250 [FREE Full text] [doi: 10.3389/fgene.2014.00250] [Medline: 25177340]
25. Crosslin DR, Carrell DS, Burt A, Kim DS, Underwood JG, Hanna DS, et al. Genetic variation in the HLA region is associated with susceptibility to herpes zoster. Genes Immun 2015;16(1):1-7 [FREE Full text] [doi: 10.1038/gene.2014.51] [Medline: 25297839]
26. Diogo D, Tian C, Franklin CS, Alanne-Kinnunen M, March M, Spencer CC, et al. Phenome-wide association studies across large population cohorts support drug target validation. Nat Commun 2018 Oct 16;9(1):4285 [FREE Full text] [doi: 10.1038/s41467-018-06540-3] [Medline: 30327483]
27. Hebbring SJ. The challenges, advantages and future of phenome-wide association studies. Immunology 2014 Feb;141(2):157-165 [FREE Full text] [doi: 10.1111/imm.12195] [Medline: 24147732]
28. Knisely MR, Carpenter JS, Draucker CB, Skaar T, Broome ME, Holmes AM, et al. CYP2D6 drug-gene and drug-drug-gene interactions among patients prescribed pharmacogenetically actionable opioids. Appl Nurs Res 2017 Dec;38:107-110 [FREE Full text] [doi: 10.1016/j.apnr.2017.10.001] [Medline: 29241501]
29. Rusiecki JA, Byrne C, Galdzicki Z, Srikantan V, Chen L, Poulin M, et al. PTSD and DNA Methylation in select immune function gene promoter regions: a repeated measures case-control study of U.S. military service members. Front Psychiatry 2013 Jun 24;4:56 [FREE Full text] [doi: 10.3389/fpsyt.2013.00056] [Medline: 23805108]

30. Dhalla A, McDonald TE, Gallagher RP, Spinelli JJ, Brooks-Wilson AR, Lee TK, et al. Cohort Profile: the British Columbia Generations Project (BCGP). Int J Epidemiol 2019 Apr 01;48(2):377-38k. [doi: 10.1093/ije/dyy160] [Medline: 30169793]
31. IC/ES - Data Discovery Better Health. 2019. URL: https://www.ices.on.ca/ [accessed 2021-08-19]
32. About the 45 and up study. Sax Institute. 2019. URL: https://www.saxinstitute.org.au/our-work/45-up-study/ [accessed 2021-08-19]
33. What are Single Nucleotide Polymorphisms (SNPs)? MedlinePlus. 2020. URL: https://medlineplus.gov/genetics/understanding/genomicresearch/snp/ [accessed 2021-08-25]
34. dbSNP. National Center for Biotechnology Information. 2021. URL: https://www.ncbi.nlm.nih.gov/snp/ [accessed 2021-08-25]
35. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet 2017 Jul 06;101(1):5-22 [FREE Full text] [doi: 10.1016/j.ajhg.2017.06.005] [Medline: 28686856]
36. Robinson JR, Denny JC, Roden DM, Van Driest SL. Genome-wide and phenome-wide approaches to understand variable drug actions in electronic health records. Clin Transl Sci 2018 Mar;11(2):112-122 [FREE Full text] [doi: 10.1111/cts.12522] [Medline: 29148204]
37. Bush WS, Moore JH. Chapter 11: genome-wide association studies. PLoS Comput Biol 2012;8(12):e1002822 [FREE Full text] [doi: 10.1371/journal.pcbi.1002822] [Medline: 23300413]
38. The electronic health records system In the UK. Centre for Public Impact. 2017. URL: https://www.centreforpublicimpact.org/case-study/electronic-health-records-system-uk/ [accessed 2021-08-25]
39. International classification of diseases 11th revision. World Health Organization. URL: https://icd.who.int/en [accessed 2021-08-25]
40. Pendergrass SA, Crawford DC. Using electronic health records to generate phenotypes for research. Curr Protoc Hum Genet 2019 Jan;100(1):e80 [FREE Full text] [doi: 10.1002/cphg.80] [Medline: 30516347]
41. Kho AN, Hayes MG, Rasmussen-Torvik L, Pacheco JA, Thompson WK, Armstrong LL, et al. Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study. J Am Med Inform Assoc 2012;19(2):212-218 [FREE Full text] [doi: 10.1136/amiajnl-2011-000439] [Medline: 22101970]
42. Middleton RM, Rodgers WJ, Chataway J, Schmierer K, Rog D, Galea I, et al. Validating the portal population of the United Kingdom Multiple Sclerosis Register. Mult Scler Relat Disord 2018 Aug;24:3-10. [doi: 10.1016/j.msard.2018.05.015] [Medline: 29860199]
43. The personal genome project. Personal Genome Project: Global Network. 2021. URL: https://www.personalgenomes.org/gb [accessed 2021-08-25]
44. SAIL Databank. 2021. URL: https://saildatabank.com/ [accessed 2001-08-25]
45. Biobank homepage. Biobank UK. 2021. URL: https://www.ukbiobank.ac.uk/ [accessed 2021-08-25]
46. dbGap database. National Center for Biotechnology Information. 2021. URL: https://www.ncbi.nlm.nih.gov/gap/ [accessed 2021-08-25]
47. Publications. eMERGE Networks. 2021. URL: https://emerge-network.org/publications/ [accessed 2021-08-25]
48. BC generations homepage. BC Generations Project. 2021. URL: https://www.bcgenerationsproject.ca/ [accessed 2021-08-25]
49. Population data BC homepage. Population Data BC. 2021. URL: https://www.popdata.bc.ca/ [accessed 2021-08-25]
50. Genomes of healthy older Australians to accelerate medical research in NSW and beyond. Garvan Institute of Medical Research. 2016. URL: https://www.garvan.org.au/news-events/news/genomes-of-healthy-older-australians-to-accelerate-medical-research-in-nsw-and-beyond [accessed 2021-08-25]
51. GIANT: Genetic Investigation of ANthropometric Traits. GIANT consortium. 2019. URL: https://portals.broadinstitute.org/collaboration/giant/index.php/Main_Page [accessed 2021-08-25]
52. GWAS Catalog: The NHGRI-EBI Catalog of human genome-wide association studies. GWAS Catalog. 2021. URL: https://www.ebi.ac.uk/gwas/ [accessed 2021-08-25]
53. Database of genomic variants. 2021. URL: http://dgv.tcag.ca/dgv/app/home [accessed 2021-08-25]
54. BioVu. Department of Biomedical Informatics. 2021. URL: https://www.vumc.org/dbmi/biovu [accessed 2021-08-25]
55. Smith BH, Campbell H, Blackwood D, Connell JA, Connor MJ, Deary IJ, et al. Generation Scotland: the Scottish family health study; a new resource for researching genes and heritability. BMC Med Genet 2006 Oct 02;7:74 [FREE Full text] [doi: 10.1186/1471-2350-7-74] [Medline: 17014726]
56. Lea NC, Nicholls J, Dobbs C, Sethi N, Cunningham J, Ainsworth J, et al. Data safe havens and trust: toward a common understanding of trusted research platforms for governing secure and ethical health research. JMIR Med Inform 2016 Jul 21;4(2):e22 [FREE Full text] [doi: 10.2196/medinform.5571] [Medline: 27329087]
57. Health and Care Research Wales Annual Report 2017-18. Health and Care Research Wales. 2018. URL: https://healthandcareresearchwales.org/sites/default/files/2020-10/Health_and_Care_Research_Wales_Annual_Report_17-18_eng.pdf [accessed 2021-08-25]
58. Genomics for precision medicine strategy. Genomics Partnership Wales. 2017. URL: https://genomicspartnership.wales/genomics-precision-medicine-strategy/ [accessed 2021-08-25]

59. Brace yourself for large-scale whole genome sequencing. Mass Genomics. 2014. URL: http://massgenomics.org/2014/11/brace-yourself-for-large-scale-whole-genome-sequencing.html [accessed 2021-08-25]
60. Rocher L, Hendrickx JM, de Montjoye Y. Estimating the success of re-identifications in incomplete datasets using generative models. Nat Commun 2019 Jul 23;10(1):3069 [FREE Full text] [doi: 10.1038/s41467-019-10933-3] [Medline: 31337762]
61. Guide to the UK General Data Protection Regulation (UK GDPR). Information Commissioner's Office. 2020. URL: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/ [accessed 2021-08-25]
62. Steinsbekk K, Solberg B. Biobanks--when is re-consent necessary? Public Health Ethics 2011 Nov 03;4(3):236-250. [doi: 10.1093/phe/phr031]
63. The genetic non-discrimination act: an overview. Canadian Civil Liberties Association. 2021. URL: https://ccla.org/genetic-non-discrimination-act-overview/ [accessed 2021-08-25]
64. Qu HQ, Tien M, Polychronakos C. Statistical significance in genetic association studies. Clin Invest Med 2010 Oct 01;33(5):E266-E270 [FREE Full text] [doi: 10.25011/cim.v33i5.14351] [Medline: 20926032]
65. Allen GE. The social and economic origins of genetic determinism: a case history of the American Eugenics Movement, 1900-1940 and its lessons for today. Genetica 1997;99(2-3):77-88. [doi: 10.1007/BF02259511] [Medline: 9463076]
66. Thomas D. Gene--environment-wide association studies: emerging approaches. Nat Rev Genet 2010 Apr;11(4):259-272 [FREE Full text] [doi: 10.1038/nrg2764] [Medline: 20212493]
67. Silventoinen K, Rokholm B, Kaprio J, Sørensen TI. The genetic and environmental influences on childhood obesity: a systematic review of twin and adoption studies. Int J Obes (Lond) 2010 Jan;34(1):29-40. [doi: 10.1038/ijo.2009.177] [Medline: 19752881]
68. Jones K, Lacey A, Perkins B, Rees M. Data safe havens to combine health and genomic data: benefits and challenges. Int J Popul Data Sci 2017 Apr 19;1(1):- [FREE Full text] [doi: 10.23889/ijpds.v1i1.348]
69. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. Science 2013 Jan 18;339(6117):321-324 [FREE Full text] [doi: 10.1126/science.1229566] [Medline: 23329047]
70. Erlich Y, Narayanan A. Routes for breaching and protecting genetic privacy. Nat Rev Genet 2014 Jun;15(6):409-421 [FREE Full text] [doi: 10.1038/nrg3723] [Medline: 24805122]
71. Dyer C. Huntington's case raises questions about passing gene test results to relatives. BMJ 2017 May 18;357:j2433. [doi: 10.1136/bmj.j2433] [Medline: 28522429]
72. Laurie G. Genetic Privacy: A Challenge to Medico-Legal Norms. Cambridge: Cambridge University Press; 2002.
73. Mullen C. Decisions, consent and expectations of the individual. In: Widdows H, Mullen C, editors. The Governance of Genetic Information: Who Decides?. Cambridge: Cambridge University Press; 2009:51-72.
74. Shabani M, Bezuidenhout L, Borry P. Attitudes of research participants and the general public towards genomic data sharing: a systematic literature review. Expert Rev Mol Diagn 2014 Nov;14(8):1053-1065. [doi: 10.1586/14737159.2014.961917] [Medline: 25260013]
75. Cameron D, Pope S, Clemence M. Dialogue on dataxploring the public's views on using administrative data for research purposes. Economic and Social Research Council (ESRC). 2014. URL: https://esrc.ukri.org/files/public-engagement/public-dialogues/dialogue-on-data-exploring-the-public-s-views-on-using-linked-administrative-data-for-research-purposes/ [accessed 2021-08-25]
76. The one-way mirror: public attitudes to commercial access to health data. Wellcome Trust. 2016 Mar. URL: https://wellcome.org/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf [accessed 2021-08-25]
77. Enabling responsible genomic data sharing for the benefit of human health. Global Alliance for Genomics and Health. URL: https://www.ga4gh.org/ [accessed 2021-08-21]
78. Jones K, Daniels H, Squires E, Ford DV. Public views on models for accessing genomic and health data for research: mixed methods study. J Med Internet Res 2019 Aug 21;21(8):e14384 [FREE Full text] [doi: 10.2196/14384] [Medline: 31436163]
79. Tekola-Ayele F, Rotimi CN. Translational genomics in low- and middle-income countries: opportunities and challenges. Public Health Genomics 2015;18(4):242-247 [FREE Full text] [doi: 10.1159/000433518] [Medline: 26138992]
80. Akhlaq A, Sheikh A, Pagliari C. Barriers and facilitators to health information exchange in low- and middle-income country settings: a systematic review protocol. J Innov Health Inform 2015 Mar 27;22(2):284-292 [FREE Full text] [doi: 10.14236/jhi.v22i2.98] [Medline: 26245242]

Abbreviations

BC: British Columbia
EHR: electronic health record
eMERGE: Electronic and Medical Records and Genomics
GAP: Genetics and Psychosis
GWAS: Genome-Wide Association Studies
ICD: International Classification of Disease
NHS: National Health Service

PheWAS: phenome-wide association studies
POND: Province of Ontario Neurodevelopmental Disorders
SAIL: Secure Anonymized Information Linkage

Edited by R Kukafka; submitted 02.08.19; peer-reviewed by S Musy, C Hockham, S Thiebes, X Pastor; comments to author 18.05.20; revised version received 01.10.20; accepted 15.07.21; published 24.09.21.

Please cite as:
Daniels H, Jones KH, Heys S, Ford DV
Exploring the Use of Genomic and Routinely Collected Data: Narrative Literature Review and Interview Study
J Med Internet Res 2021;23(9):e15739
URL: https://www.jmir.org/2021/9/e15739
doi:10.2196/15739
PMID:34559060

Original Paper

Presentation, Treatment, and Natural Course of Severe Symptoms of Urinary Tract Infections Measured by a Smartphone App: Observational and Feasibility Study

Akke Vellinga[1], BSc, MSc, PhD; Karen Farrell[2], MA; Roisin Fallon[2], MA; Daniel Hare[3], BMed; Una Sutton-Fitzpatrick[4], BMed; Martin Cormican[5,6], MD

[1]School of Medicine, National University of Ireland Galway, Galway, Ireland

[2]School of Medicine, Health Research Board Primary Care Clinical Trials Network Ireland, National University of Ireland Galway, Galway, Ireland

[3]Department of Clinical Microbiology, University Hospital Galway, Galway, Ireland

[4]Tallaght University Hospital, Dublin, Ireland

[5]Health Service Executive, Antimicrobial Resistance and Infection Control Team, Dublin, Ireland

[6]Discipline of Bacteriology, National University of Ireland Galway, Galway, Ireland

Corresponding Author:
Akke Vellinga, BSc, MSc, PhD
School of Medicine
National University of Ireland Galway
1 Distillery Road
Galway, H91TK33
Ireland
Phone: 353 91495194
Email: akke.vellinga@nuigalway.ie

Abstract

Background: Urinary tract infections (UTIs) are one of the most common conditions in women. Current information on the presentation, management, and natural course of the infection is based on paper diaries filled out and subsequently posted by patients.

Objective: The aim of this study is to explore the feasibility of a smartphone app to assess the natural course and management of UTIs.

Methods: A smartphone app was developed to collect data from study participants presenting with symptoms of UTI in general practice. After initial demographic and treatment information, symptom severity was recorded by the patient after a reminder on their smartphone, which occurred twice daily for a period of 7 days or until symptom resolution.

Results: A total of 181 women aged 18-76 years downloaded the smartphone app. The duration of symptoms was determined from the results of 178 participants. All patients submitted a urine sample, most patients were prescribed an antibiotic (163/181, 90.1%), and 38.7% (70/181) of the patients had a positive culture. Moderately bad or worse symptoms lasted a mean of 3.8 (SD 3.2; median 4) days, and 70.2% (125/178) of the patients indicated that they were cured on day 4 after consultation. This compares with other research assessing symptom duration and management of UTIs using paper diaries. Patients were very positive about the usability of the smartphone app and often found the reminders supportive. On the basis of the feedback and the analysis of the data, some suggestions for improvement were made.

Conclusions: Smartphone diaries for symptom scores over the course of infections are an efficient and acceptable means of collecting data in research.

(J Med Internet Res 2021;23(9):e25364) doi:10.2196/25364

KEYWORDS

urinary tract infections; general practice; smartphone application; mobile phone

Introduction

Background

Urinary tract infections (UTIs) are one of the most common conditions for which women consult their general practitioner (GP) [1]. More than half of the women consult their GP for a UTI at least once in their lives [2,3].

UTIs are considered self-limiting infections [4,5], but antibiotics are generally prescribed empirically [6-8]. The prescription of antimicrobials for UTI accounts for a substantial amount of prescriptions in primary care and contributes to an increased risk of adverse drug effects, the burden of resistant infections [9,10], and increased costs [11]. In women presenting in primary care with suspected UTI, which turned out to be susceptible to the antibiotic prescribed, the duration of more severe symptoms was 3.3 days on average [2]. However, the duration increased to 4.7 days for women with resistant infections and 4.3 days for women with symptoms but no significant bacterial growth (urethral syndrome) [2]. An international comparison of the presentation and management of UTI in four European countries found wide differences in management, with antibiotics prescribed for 59%-95% of the presenting women; however, there were no differences in the time taken to resolve moderately bad to worse symptoms, which took a median of 4 days [12].

All studies reporting on the management of UTI base their findings on the use of paper diaries, which are provided to women at the time of consultation and are requested to be returned by post after symptoms are resolved [5,13-15]. Such diaries are also used in UTI trials where antibiotic treatment is compared with symptomatic treatment for UTI in general practice [16,17].

In one observational study that recorded the natural course of UTIs, 64% of the symptom diaries were recovered [2], and in another observational cohort of four countries, 70% of the diaries were recovered [18]. A Danish study observing the effect of point-of-care testing for UTI recovered 85% of the patient diaries [19]. All of these studies have used additional resources to contact patients and urge them to post their diaries.

Compliance with paper diaries has been compared with electronic diary keeping for patients recording pain three times daily for 21 days. The difference between electronic and paper recording was analyzed, and the results showed a high level of feigned compliance in the use of paper diaries [20]. Different versions of electronic diaries have been piloted and tested in various environments to record symptoms, such as in inflammatory bowel disease [21], overactive bladder [22], and asthma [23].

Smartphone apps are increasingly used in care innovation research and provide new opportunities to develop interventions [24]. In pain settings, apps have been most widely used to record diary entries both in observational [25] and clinical trial settings [26]; however, apps have also been implemented for other conditions [27-29]. In these studies, smartphone diaries have been shown to be highly efficient and generally perform better than paper-based diaries.

In our own SIMPle (Supporting the Improvement and Management of Prescribing for UTI) study, we used an observational study to follow patients with UTI using a smartphone app [30]. The results of this feasibility study showed that a smartphone app was well received and that once a patient committed to reporting their symptoms, they would continue doing so for the duration of the study [31].

Objective

The objective of this study is to develop and evaluate the feasibility of a smartphone app to assess the natural course and management of UTIs.

Methods

Aim

This study explored the feasibility of a smartphone app to assess the natural course and management of UTIs in women presenting in primary care and to document the outcomes of urine culture, antibiotic resistance of bacterial growth, and antibiotic prescription using a smartphone app with reminders for data collection.

Design

This study is an observational feasibility study of women presenting with symptoms of UTI.

Setting and Participants

Between November 2018 and December 2019, 6 GPs in the west of Ireland enrolled adult female patients with symptoms of UTI. After obtaining consent, the patient received a patient number, which was texted to a central number to record the patient's phone number, and a link to download the app was returned. Both patients and practitioners received a payment for each download of the smartphone app.

Follow-up

A telephone call was made to all patients on days 4 and 28 (or as soon as possible after this day). The aim of the day 4 phone call was to confirm the symptom score. In addition, information on adverse events, experience using the app, and return consultation with a GP were recorded.

Smartphone App

A smartphone app was developed to collect data from the participants (Figure 1). At enrollment, patients were requested to record demographic information (age, recent UTI, marital status, children, and medical card) as well as treatment details (antibiotic treatment: yes or no, name of antibiotic, and duration).

Figure 1. Smartphone app.

To record the severity of symptoms, patients received a reminder twice a day, once in the morning at 9 AM (or 10 AM and 11 AM if no entry was made) and once in the afternoon at 6 PM (or 7 PM and 8 PM if no entry was made). Patients were reminded to provide their symptom score for 7 days and, if no recovery occurred, up to 14 days or until symptom recovery. Symptom recovery was defined as a score of <2 (moderate) for each symptom.

Symptom severity included four symptoms: dysuria (painful and/or burning urination), frequency of urination, urgency of urination, and lower abdominal pain. Symptoms were scored from 0 (not at all) to 4 (as bad as it gets). An additional question—*Do you feel cured?*—was included after the symptom score to allow comparison with a Scandinavian trial in which this was the main outcome score [32].

On days 0 and 5, patients were also asked to fill out a UTI-related questionnaire on impairment to daily life activities on the Activity Impairment Assessment (AIA) [33] as well as self-rated health in general and on the particular day (on a visual analog scale from 0 to 100) [34]. The AIA is a five-item questionnaire that assesses the amount of times an individual's work or regular activities has been affected because of their UTI. Patients respond to the AIA items on a five-point Likert scale, with the response options *none of the time*, *a little of the time*, *some of the time*, *most of the time*, and *all of the time*. The questions were as follows: did you have to cut down on time at work, could you accomplish less, were you limited in the kind of work you did, did you have difficulty performing work, and did your UTI interfere with social activity?

Sample Size

Considering a mean difference of half a day clinically significant, a sample of 92 patients (α=.05 and β=.2) should detect such a difference in the mean duration of symptoms (score moderate or worse on the four items: dysuria, frequency of urination, the urgency of urination, and lower abdominal pain [5]) of 3.8 (SD 3.0) on day 4.

Data Analysis

The duration of symptoms included the day the patient consulted their GP. The mean duration of symptoms was based on the day their mean total symptom score was moderately bad or worse and similarly for the duration of each symptom separately. The mean duration to cure is based on a yes answer to the question *Do you feel cured?* Symptom resolution was based on the day that all symptoms were 0. As the app was filled out for a duration of 7 days, the return of symptoms within the 7 days was included in the duration of symptoms. The end of symptoms was also indicated on the first day when a patient ceased making any more entries.

Antibiotic prescriptions were recorded on the day of the consultation, and patients were asked every day if they were still taking the antibiotic. The antibiotic prescribed was assessed retrospectively in the context of the microbiological culture results to assess if the organism tested was susceptible to the antibiotic chosen empirically.

Urethral syndrome was defined as having symptoms of suspected UTI, with no significant bacterial growth detected on culture [2].

Analysis was performed using STATA 13.0 (StataCorp LLC) and IBM SPSS, version 26. Tableau was used to visualize the data.

Laboratory Analysis

Urine samples were collected from each patient. Samples were sent to the Microbiological Lab of the University Hospital Galway for analysis. A positive sample was based on a colony count of >10,000 cfu/ml pure growth observed after overnight incubation on chromogenic agar. Mixed growth was also recorded and included for further analysis. Organisms detected in pure culture were identified by matrix-assisted laser desorption ionization–time of flight (Bruker). Susceptibility testing was performed using the EUCAST (European Committee on Antimicrobial Susceptibility Testing) disk diffusion methodology and interpretive criteria.

Ethical Approval

Ethical approval was obtained from the Irish College of General Practitioners. Individual informed consent was obtained from all participants before enrollment in the study. Consent had to be confirmed at registration when downloading the smartphone app.

Availability of Data and Materials

Access to anonymized patient data can be obtained from the corresponding author upon request. The code for the smartphone app can be obtained from the corresponding author and used freely if referenced appropriately.

Results

Overview

A total of 181 patients downloaded the smartphone app. Their mean age was 29.7 (SD 14; median 22, range 18-76) years. Of the women, 28.1% (51/181) were married or in a relationship, 69.1% (125/181) did not have children, 18.2% (33/181) had one or two children, and 12.7% (23/181) had three or more children. More than one-third (67/181, 37%) of the patients were entitled to free medical and GP care, whereas 62.9% (114/181) had to pay for their GP visit as a private patient. Of the participants, 65.7% (119/181) had or were pursuing a university degree and 34.2% (62/181) had secondary education. A total of 39.8% (54/181) were working, 6.1% (11/181) were homemakers, and 54.1% (98/181) were students.

Of the 181 patients, 3 never filled out any symptom score, whereas all others provided at least one symptom score. On day 4, 39 did not provide a symptom score, but 13 of them provided one on the previous day.

Of the 181 patients, 163 (90%) were prescribed an antibiotic. Nitrofurantoin was most frequently prescribed (66.3%, 108/163), followed by trimethoprim (20/163, 12.3%), co-amoxicillin (11/163, 6.7%), amoxicillin (8/163, 4.9%), fosfomycin (5/163, 3.1%), and quinolone (2/163, 1.2%).

The question that was asked at every entry, that is, if they took their antibiotic or if they used any pain medication, was filled out irregularly and could not be analyzed.

Symptom Duration

Symptoms started on average 5.6 days (SD 6.3) before they consulted their GP, with a median of 4 days (IQR 5). The mean duration of symptoms after they consulted their GP was 4.2 days (SD 3.0) and the mean time to cure (indicated by the patient) was 3.8 days (SD 3.2). Table 1 and Figure 2 show the mean duration of moderately bad or worse symptoms in more detail. The most common symptom rated moderately bad or worse on the first day was the urgency of urination (114/178, 64%), followed by frequency of urination (110/178, 61.8%), lower abdominal pain (98/178, 55.1%), and dysuria (71/178, 39.9%); the duration of each symptom is shown in Multimedia Appendix 1. There were no significant differences between the types of UTI and if an antibiotic was prescribed or not. An overview of the percentage of patients who indicated that they felt cured each day since the consultation is presented in supplementary Figure 2. Overall, 70.2% (125/178) of patients indicated that they felt cured on the fourth day after consultation (day 5), and this was 71.8% (115/160) among those who had an antibiotic prescribed.

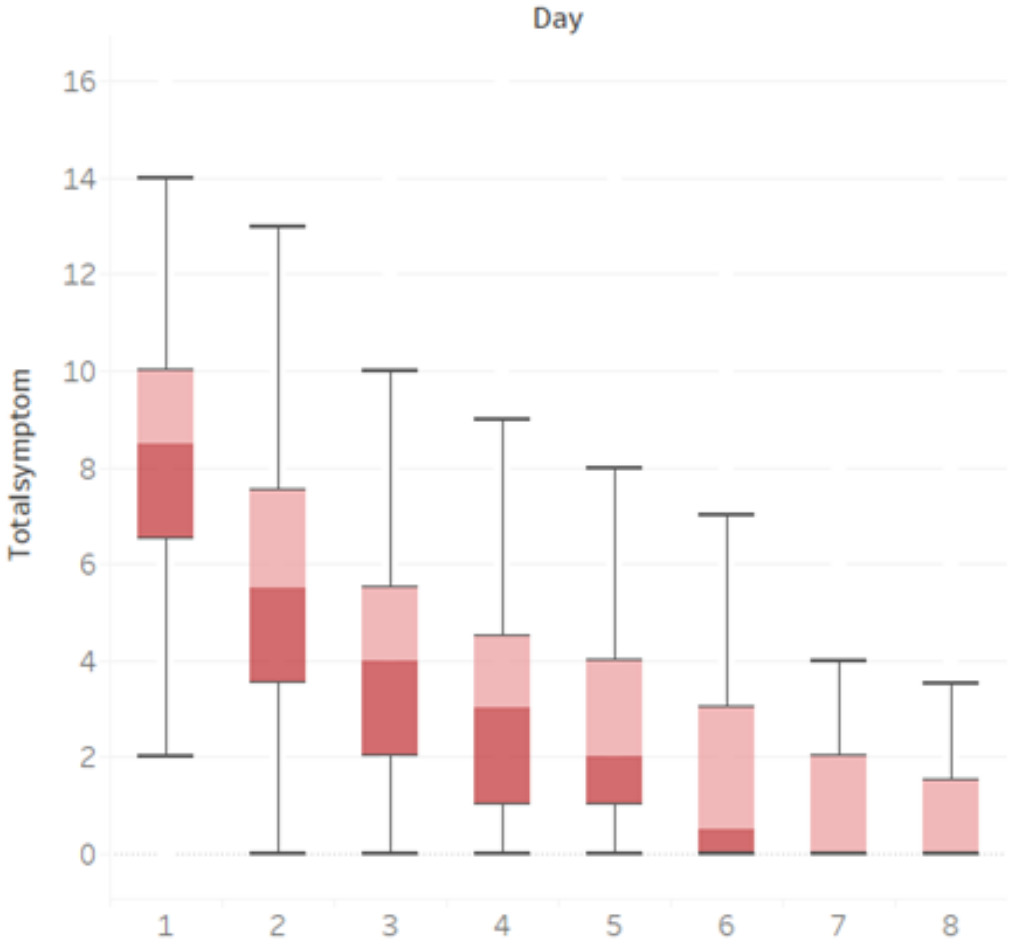
Table 1. Mean (SD) and median of the duration of symptoms according to urinary tract infection type and antibiotic prescription (N=178).

Variables	Symptoms	Dysuria	Frequency of urination	Urgency of urination	Lower abdominal pain	Resolution	Cured
Duration (days)							
Value, mean (SD)	4.2 (3.0)	2.0 (2.6)	2.6 (2.6)	3.0 (2.9)	2.9 (3.3)	10.7 (20.9)	3.8 (3.2)
Value, median	4	1	3	2	2	6	3
UTI [a] type, mean (SD)							
UTI sensitive to AB[b] (n=36)	4.7 (3.5)	3.0 (0.6)	3.1 (3.3)	3.5 (3.5)	3.7 (3.6)	13.4 (26.3)	3.5 (3.1)
UTI resistant to AB or no AB (n=7)	5.0 (2.7)	3.6 (3.6)	3.4 (2.1)	3.4 (2.5)	3.0 (2.9)	6.3 (4.0)	4.3 (4.1)
UTI unknown sensitivity (n=26)	3.4 (1.8)	1.7 (1.6)	1.7 (1.2)	2.4 (1.7)	2.0 (1.9)	5.4 (2.2)	3.2 (2.0)
No significant bacterial growth (urethral syndrome; n=92)	4.3 (3.2)	1.6 (2.1)	2.8 (2.7)	2.9 (2.9)	3.1 (3.5)	12.0 (23.2)	4.1 (3.5)
Antibiotic, mean (SD)							
Yes (n=160)	4.3 (3.0)	2.0 (2.5)	2.7 (2.6)	3.0 (2.9)	3.0 (3.2)	10.3 (20.4)	3.8 (3.2)
No (n=18)	3.4 (3.5)	2.3 (3.6)	2.2 (2.6)	2.9 (3.5)	2.2 (3.6)	10.4 (22.3)	4.2 (3.6)

[a]UTI: urinary tract infection.

[b]AB: antibiotic.

Figure 2. Total symptom score by day. Each box represents the interval of 50% of the symptom scores, with the difference in color at the mean score. Whiskers indicate the range of the symptom score on each day.



Overall, 90% (163/181) of patients had moderate to worse dysuria for 4 days or less, 85.1% (154/181) had a frequency of urination, 81.2% (147/181) had the urgency of urination, and 66.9% (121/181) had lower abdominal pain. Overall, symptoms lasted 4 days or less for 56.9% (103/181) of the patients after seeing their doctor (Multimedia Appendix 2).

Microbiology

Overall, 38.7% (70/181) patients had a positive culture. Enterobacteriaceae (mainly *E. coli*) were cultured from 27.6% (50/181) of patients, followed by *S. saprophyticus* from 5.5% (10/181) of patients. A total of 17.7% (32/181) were mixed cultures and 12.2% (22/181) were <10,000 cfu/ml. Resistance to at least one of the tested antibiotics was recorded in 61% (33/54) of isolates for which this could be assessed. Comparing the antibiotic prescribed with the isolate and its resistance showed that 55% (36/66) of the isolates tested were susceptible to the antimicrobial prescribed, whereas for 41% (27/66), this was not known. For 3 patients, the cultured organisms tested resistant to the antibiotic prescribed. One of these patients was contacted to change the antibiotic (from co-amoxyclav to nitrofurantoin).

AIA and General Health

Little difference was observed for the AIA score (Table 2); the mean AIA score on day 0 was 6.6 (SD 5.0), and on day 5, it dropped to 4.1 (SD 4.6). A paired two-sided *t* test of patients providing a score on both days showed significance (P=.03), and patients were asked to rate their general health on day 0 (mean 75.6, SD 17.3) and day 5 (mean 76.3, SD 18.2; Figure 3). Their health was rated at 55.8 (SD 20.1) on day 0 and 71.9 (SD 19.0) on day 5. The difference between their rated health on days 0 and 5, as well as the difference in overall health, was significant. On day 5, their health scores were still significantly different from the overall health scores, but the overall health scores on days 0 and 5 were not statistically different.

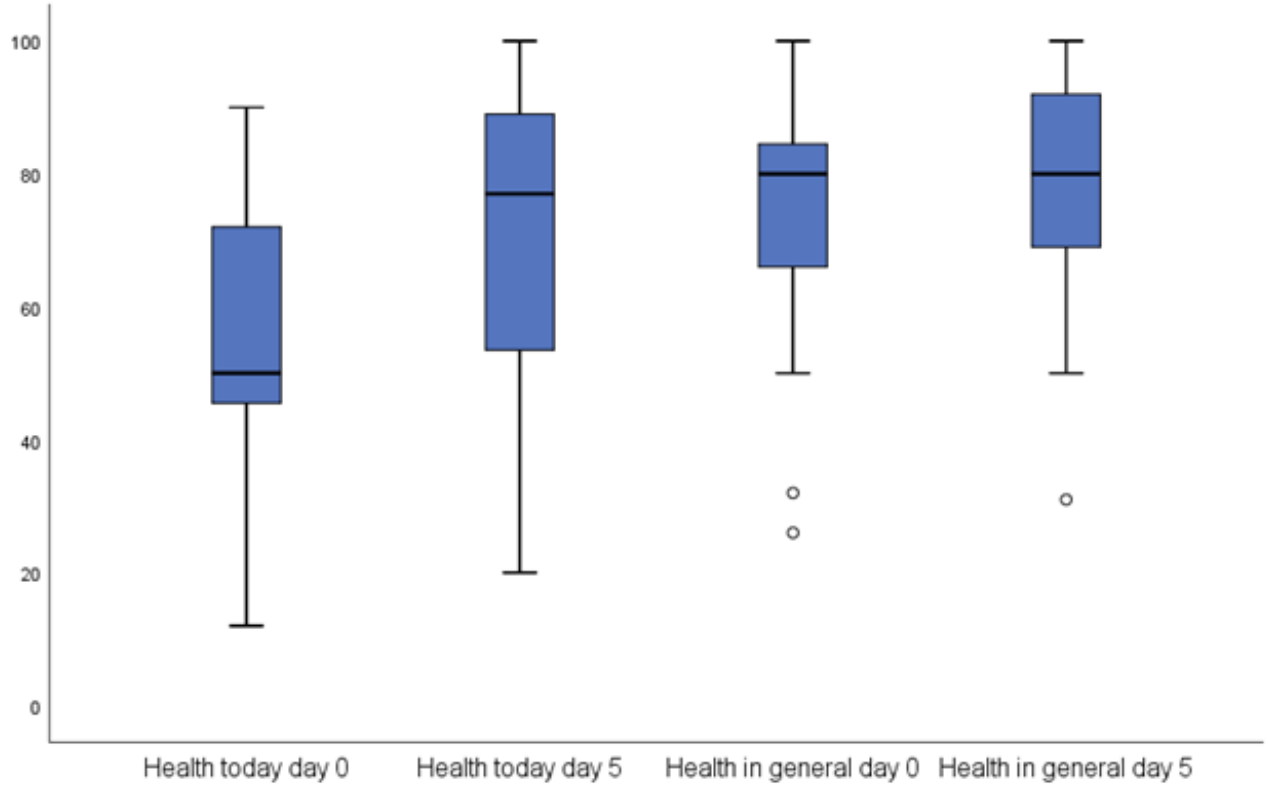
The total symptom score on day 0 was correlated with the AIA (R^2=23%) and similarly with the health score (17%) on day 0 and day 5.

Table 2. Overview of the Activity Impairment Assessment (urinary tract infection–related questionnaire on impairment in daily life activities on the Activity Impairment Assessment) and self-rated health in general and on the particular day (on a visual analog scale from 0 to 100).

	Day of consultation (n=111), mean (SD)	Day 5 (n=96), mean (SD)
AIA[a] item		
Cut down on time at work	0.95 (1.2)	0.6 (1.0)
Accomplished less	1.5 (1.2)	0.8 (1.0)
Limited in type of work	1.2 (1.2)	0.7 (1.0)
Difficulty performing work	1.4 (1.1)	0.8 (1.0)
Interfered with social activity	1.6 (1.2)	1.1 (1.2)
Total AIA score	6.6 (4.9)	4.1 (4.6)
Health score in general	76 (17)	76 (18)
Health score today	56 (20)	72 (19)

[a]AIA: Activity Impairment Assessment.

Figure 3. Health score indicated on a sliding scale (0-100) answered for the questions how is your health today? and how is your health in general on day 0 and day 5. The box represents 50% of the answers, whereas the whiskers cover 100% of the answers. The line in the box represents the mean score. Outliers are indicated by the dots.



Associations

There was no difference in duration of symptoms, total symptom score on the day of consultation, AIA, and health status on the day of consultation between those who did and did not go to university or between students and nonstudents. Only the total symptom score was significantly higher (P=.03) for those not in a relationship compared with participants who were married or with a partner, but no other differences were observed. Age was not correlated with any outcome (duration, total score, days with symptoms before consultation, health score, or AIA).

Day 4 and Day 28 Phone Call

A call was answered by 152 participants on day 4, and 128 patients answered on day 28.

On day 4, 39.8% (71/178) indicated that they felt cured, compared with 70.2% according to the app data. A comparison showed that 67.7% (103/152) indicated the same on the call as they did on the app, 3.9% (6/152) indicated that they were cured over the phone but not on the app, and 41/152 (26.9%) said they were cured in the app but not when asked over the phone. The total symptom score on day 4 according to the phone call was 3.8 days (SD 3.5) according to the call and 2.6 days (SD 2.6)

according to the app, which was significantly different, that is, symptoms were rated more severe when asked over the phone.

When asked about side effects, 28.9% (44/152) reported side effects from the medication. The main side effects recorded were stomach upset or diarrhea (17/152, 11.2%) and thrush (4/152, 2.6%), and 18 did not specify any side effects, and five reported urine discoloration (all patients reported that they had received nitrofurantoin).

Of the patients who answered a call on day 28, 25% (32/128) consulted the GP again for symptoms of UTI during the 28 days after the initial consultation. Patients who consulted their GP had significantly (P=.006) higher symptom duration (5.7 days, SD 4.3 vs 3.9 days, SD 2.6) but no difference in symptom score on the day they consulted. There was no difference in antibiotic or painkiller prescription between those who did and did not reconsult. Patients who reconsulted were significantly older (mean 35.9, SD 17.8 vs mean 27.9, SD 12.4). On day 5, the AIA score (6.9% vs 2.9%) was significantly higher, and the health score (59.7% vs 75.1%) was significantly lower for those who reconsulted.

Evaluation of the App

On the day 4 call, 116 patients provided further comments on their UTI and/or the use of the app, whereas 128 patients provided further comments on their UTI and/or the use of the app on day 28. In general, the comments were positive. Some patients liked the fact that the app reminded them of their medication and improvement of their symptoms:

> *Handy for reminding in morning and evening to take my medication and think about my symptoms more and if they were improving or not.*
>
> *App was good as made me check in with myself, showed me how it was a recovery process.*
>
> *App was good, not time consuming, interesting for herself to track day to day feelings.*

Overall, the app was received well:

> *App was good, quick and easy to use.*

However, a few comments were made in relation to the repetition of questions:

> *Felt like they were the same questions over and over again.*

Discussion

Principal Findings

It was shown that the use of a smartphone app to track patients' symptoms of UTI is an efficient and effective approach to replace paper diaries. Our estimates of the mean and median duration of symptoms from the date of the consultation were 4.2 days and 4 days, respectively. This is comparable with the observations made in previous studies of the natural course of infection in which paper diaries were used [2,12]. However, the mean and median duration of individual symptoms were rated lower in the app than in the paper diaries from both studies. Comparing our results with other trials using paper diaries showed a mean duration of moderate to worse symptoms at 4.6 days [16] compared with 4.3 in our study. The *do you feel cured* question as the main outcome in the Scandinavian trial was also included in our study; 74% indicated cured on day 4 after the consultation when taking the antibiotic, whereas in our study, 71.8% indicated cured in the group where an antibiotic was prescribed. In general, the outcomes are similar to the outcomes of other studies. The main difference and challenge in comparing different studies and trials is the day on which outcomes are measured, that is, whether this includes the day of consultation or not. Inclusion of the method of measuring in a core outcome set for UTI should be clearly defined [35].

UTIs are considered self-limiting infections [4,5], but antibiotics are generally prescribed empirically [6-8]. The prescription of antimicrobials for UTI accounts for a substantial number of prescriptions in primary care, contributing to an increased burden of resistant infections [9,10] as well as increased costs [11]. Our study showed that 90% of the patients presenting were prescribed an antibiotic, of whom 81.6% received a first-line antibiotic [8]. Only 38.7% of the patients had bacteriological confirmation of UTI. Very similar figures of prescribing were confirmed in Wales, England, and Spain, whereas the Netherlands showed much lower prescribing (59%); however, confirmed UTI was approximately 17% in Wales and England, 42% in Spain, and 64% in the Netherlands. Although more literature is emerging suggesting that symptomatic treatment of UTI is often sufficient and acceptable [16,17], most GPs are hesitant to do so. Compared with the last 10 years, GPs are prescribing more in line with guidelines. In 2010, in a study conducted in the same region, findings suggested only 26% GPs prescribing a first-line antibiotic [7]. The efforts made since then to decrease the use of co-amoxyclav in general practice and encourage the use of (first line) prescribing guidelines seem to be having an effect [30,36].

Our overall AIA score on the day of the consultation was 6.6 and 4.1 on day 4, compared with the trial of Gagyor et al [16], in which 8.9 and 1.1 were recorded, respectively. However, it is unclear why such large differences were observed. The validation of the AIA score showed a total score of 7.6 on the day of consultation, which is still higher than our finding, but they do not provide a follow-up score [33]. The inclusion of the health score might better measure general well-being on the day of the consultation, and it shows high concordance between the general health score on day 0 and day 4. This general health score may provide a better insight into the severity of the UTI and its impact on well-being.

The day 4 phone call was included to see how the smartphone scores compared with the answer provided over the phone, and 39.8% indicated that they felt cured, compared with 70.2% indicating to be cured when recorded on their smartphone. The total symptom score according to the app was also lower than that according to the phone call. Again, it is not clear why people report lower scores on the app than on a phone call.

During the follow-up period, 25% reported a reconsultation and 28.9% reported side effects of some sort, which is in line with the reported reconsultations in the four-country study.

In relation to the use of a smartphone app to collect data and considering the comments as well as the analysis of the data, a number of conclusions can be made (Textbox 1).

Textbox 1. Conclusion made based on the use of a smartphone app to collect data, comments, and analysis of the data.

Conclusions

- A smartphone app is an efficient and effective tool for collecting real-time data in research on infections.
- Our app asked for daily scores twice, which were decided to capture differences between morning and afternoon scores. Many did not fill out the app twice a day, and those who did, did not show major differences between morning and afternoon. It also introduced an additional step in the analysis, using the mean score for the day, or if only one score was provided, to use this score.
- Patients were interested and committed to using their app at the start of the study. At this stage, a few additional questions could be included to improve its use:
- The app should request at the start what time of the day is most convenient to fill out the app. This should then be the trigger for the reminder on the phone to fill out their app.
- Type and duration of antibiotic prescribed: On the day after their course is intended to finish, patients can be asked if they have any medication left. This would avoid repeatedly asking if they had taken their antibiotic.
- For painkiller use, this should only be included if it is part of particular interest to the main research question.

Strengths and Limitations

To our knowledge, this is the first prospective study to describe the natural course of a UTI using a smartphone app. Smartphone apps are frequently used for chronic conditions, where patients have more time and are not in distress when asked to enroll in a study. The acceptability and ease of use of the smartphone app have been shown for both patient use and completeness of data. However, participants may have been selectively rather than randomly invited to participate, but no reliable logs of potentially eligible patients were retained. We encouraged GPs to enroll any age group, and the age range (18-76 years), clearly shows that age was not a limitation to participate in our study. The overall mean age of participants was, however, younger than previous studies by our group, in which the mean age was 56.1 years (SD 20.7), and is probably due to the inclusion of a student GP center for enrolment of patients [30]. In the pilot study of our smartphone app to record symptoms of UTI, we found GPs to favor inviting younger patients to participate, which we subsequently and in this study, aimed to avoid by asking GPs to include all ages [31]. The restrictions to enrollment are most likely because of the additional time necessary to consent patients to the study, particularly as UTI consultation is usually considered straightforward and short.

A few questions arise with the results obtained, particularly in relation to discrepancies between phone call scores and smartphone scores from the same day. Other research into the course of infections using a smartphone app may be performed in the future, which may shed more light on this. However, no other study has identified the use of smartphones for infection symptom scores, so our findings could not be confirmed. However, the similarities between our outcomes and those of studies with paper diaries are encouraging. In particular, for trials where consent will be required anyway, smartphone diaries may be considered in the future, which could significantly reduce the energy and time spent to collect outcomes after a patient leaves the practice.

Conclusions

Smartphone diaries for symptom scores over the course of infections are an efficient and acceptable means of collecting data in research.

Acknowledgments

Funding was obtained as part of the SATIN (Symptomatic Versus Antibiotics Treatment of Urinary Tract Infections in Women) which was a work package of the Irish Primary Care Clinical Trials Award from the Health Research Board, Ireland (CTN-2014-011). The authors thank the members of the SATIN steering committee and the Clinical Research Facility Galway for their input in the early development of the smartphone app. The authors would like to thank the GP practitioners and the patients who participated in the study.

Authors' Contributions

AV designed, managed, and supervised the study; performed the statistical analyses and interpretation; and wrote the manuscript. KF and RF organized GP and patient enrollment; collection of smartphone data; and follow-up phone calls. DH and USF provided the microbiological information. MC provided support during the design of the study, helped to interpret the results, and corrected the manuscript. All authors read the drafts of the manuscript and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mean duration of each symptom and total symptom scored moderate to severe for each symptom (dysuria, frequency of urination, urgency of urination, and lower abdominal pain), and duration for any of these symptoms (average total duration).
[PNG File , 10 KB - jmir_v23i9e25364_app1.png]

Multimedia Appendix 2
The day a patient indicates to be cured from the day of consultation.
[PNG File , 176 KB - jmir_v23i9e25364_app2.png]

References

1. Foxman B. Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. Dis Mon 2003 Feb;49(2):53-70. [doi: 10.1067/mda.2003.7] [Medline: 12601337]
2. Little P, Merriman R, Turner S, Rumsby K, Warner G, Lowes JA, et al. Presentation, pattern, and natural course of severe symptoms, and role of antibiotics and antibiotic resistance among patients presenting with suspected uncomplicated urinary tract infection in primary care: observational study. Br Med J 2010 Feb 05;340(feb05 1):b5633 [FREE Full text] [doi: 10.1136/bmj.b5633] [Medline: 20139213]
3. Gupta K, Hooton T, Naber K, Wullt B, Colgan R, Miller L, Infectious Diseases Society of America, European Society for MicrobiologyInfectious Diseases. International clinical practice guidelines for the treatment of acute uncomplicated cystitis and pyelonephritis in women: A 2010 update by the Infectious Diseases Society of America and the European Society for Microbiology and Infectious Diseases. Clin Infect Dis 2011 Mar 01;52(5):103-120. [doi: 10.1093/cid/ciq257] [Medline: 21292654]
4. Christiaens TC, De Meyere M, Verschraegen G, Peersman W, Heytens S, De Maeseneer JM. Randomised controlled trial of nitrofurantoin versus placebo in the treatment of uncomplicated urinary tract infection in adult women. Br J Gen Pract 2002 Sep;52(482):729-734 [FREE Full text] [Medline: 12236276]
5. Little P, Moore MV, Turner S, Rumsby K, Warner G, Lowes JA, et al. Effectiveness of five different approaches in management of urinary tract infection: randomised controlled trial. Br Med J 2010 Feb 05;340(feb05 1):c199 [FREE Full text] [doi: 10.1136/bmj.c199] [Medline: 20139214]
6. Hooton TM, Besser R, Foxman B, Fritsche TR, Nicolle LE. Acute uncomplicated cystitis in an era of increasing antibiotic resistance: a proposed approach to empirical therapy. Clin Infect Dis 2004 Jul 01;39(1):75-80. [doi: 10.1086/422145] [Medline: 15206056]
7. Vellinga A, Cormican M, Hanahoe B, Bennett K, Murphy AW. Antimicrobial management and appropriateness of treatment of urinary tract infection in general practice in Ireland. BMC Fam Pract 2011 Oct 03;12(1):108 [FREE Full text] [doi: 10.1186/1471-2296-12-108] [Medline: 21967276]
8. Management of infection guidance for primary care in Ireland. Health Service Executive. URL: http://www.antibioticprescribing.ie/ [accessed 2021-07-19]
9. Vellinga A, Tansey S, Hanahoe B, Bennett K, Murphy AW, Cormican M. Trimethoprim and ciprofloxacin resistance and prescribing in urinary tract infection associated with Escherichia coli: a multilevel model. J Antimicrob Chemother 2012 Oct 22;67(10):2523-2530. [doi: 10.1093/jac/dks222] [Medline: 22729920]
10. Costelloe C, Metcalfe C, Lovering A, Mant D, Hay AD. Effect of antibiotic prescribing in primary care on antimicrobial resistance in individual patients: systematic review and meta-analysis. Br Med J 2010 May 18;340(may18 2):c2096. [doi: 10.1136/bmj.c2096] [Medline: 20483949]
11. Coast J, Smith RD. Antimicrobial resistance: cost and containment. Expert Rev Anti Infect Ther 2003 Aug 10;1(2):241-251. [doi: 10.1586/14787210.1.2.241] [Medline: 15482119]
12. Butler CC, Hood K, Verheij T, Little P, Melbye H, Nuttall J, et al. Variation in antibiotic prescribing and its impact on recovery in patients with acute cough in primary care: prospective study in 13 countries. Br Med J 2009 Jun 23;338(jun23 2):b2242 [FREE Full text] [doi: 10.1136/bmj.b2242] [Medline: 19549995]
13. Little P, Turner S, Rumsby K, Warner G, Moore M, Lowes JA, et al. Dipsticks and diagnostic algorithms in urinary tract infection: development and validation, randomised trial, economic analysis, observational cohort and qualitative study. Health Technol Assess 2009 Mar;13(19):iii-iv, ix [FREE Full text] [doi: 10.3310/hta13190] [Medline: 19364448]
14. Little P, Turner S, Rumsby K, Warner G, Moore M, Lowes JA, et al. Developing clinical rules to predict urinary tract infection in primary care settings: sensitivity and specificity of near patient tests (dipsticks) and clinical scores. Br J Gen Pract 2006 Aug;56(529):606-612 [FREE Full text] [Medline: 16882379]
15. McNulty CA, Richards J, Livermore DM, Little P, Charlett A, Freeman E, et al. Clinical relevance of laboratory-reported antibiotic resistance in acute uncomplicated urinary tract infection in primary care. J Antimicrob Chemother 2006 Nov 06;58(5):1000-1008. [doi: 10.1093/jac/dkl368] [Medline: 16998209]
16. Gágyor I, Bleidorn J, Kochen MM, Schmiemann G, Wegscheider K, Hummers-Pradier E. Ibuprofen versus fosfomycin for uncomplicated urinary tract infection in women: randomised controlled trial. Br Med J 2015 Dec 23;351:h6544 [FREE Full text] [doi: 10.1136/bmj.h6544] [Medline: 26698878]

17. Vik I, Bollestad M, Grude N, Bærheim A, Damsgaard E, Neumark T, et al. Ibuprofen versus pivmecillinam for uncomplicated urinary tract infection in women—a double-blind, randomized non-inferiority trial. PLoS Med 2018 May 15;15(5):e1002569. [doi: 10.1371/journal.pmed.1002569]
18. Butler CC, Francis N, Thomas-Jones E, Llor C, Bongard E, Moore M, et al. Variations in presentation, management, and patient outcomes of urinary tract infection: a prospective four-country primary care observational cohort study. Br J Gen Pract 2017 Nov 20;67(665):830-841. [doi: 10.3399/bjgp17x693641]
19. Holm A, Cordoba G, Sørensen T, Jessen L, Frimodt-Møller N, Siersma V, et al. Effect of point-of-care susceptibility testing in general practice on appropriate prescription of antibiotics for patients with uncomplicated urinary tract infection: a diagnostic randomised controlled trial. BMJ Open 2017 Oct 16;7(10):e018028 [FREE Full text] [doi: 10.1136/bmjopen-2017-018028] [Medline: 29042390]
20. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. Control Clin Trials 2003 Apr;24(2):182-199. [doi: 10.1016/s0197-2456(02)00320-3]
21. Litcher-Kelly L, Kellerman Q, Stone AA, Hanauer SB. Feasibility and utility of an electronic diary to assess self-report symptoms in patients with inflammatory bowel disease. Ann Behav Med 2007 Jun;33(2):207-212. [doi: 10.1007/bf02879902]
22. Quinn P, Goka J, Richardson H. Assessment of an electronic daily diary in patients with overactive bladder. BJU Int 2003 May;91(7):647-652. [doi: 10.1046/j.1464-410x.2003.04168.x] [Medline: 12699477]
23. Kupczyk M, Haque S, Sterk PJ, Niżankowska-Mogilnicka E, Papi A, Bel EH, BIOAIR investigators. Detection of exacerbations in asthma based on electronic diary data: results from the 1-year prospective BIOAIR study. Thorax 2013 Jul 06;68(7):611-618. [doi: 10.1136/thoraxjnl-2012-201815] [Medline: 23564399]
24. Kay M, Santos J, Takane M. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. In: Global Observatory for eHealth series - Volume 3. Geneva: World Health Organisation; 2011. URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf
25. Stinson J, Petroz G, Tait G, Feldman B, Streiner D, McGrath P, et al. e-Ouch: usability testing of an electronic chronic pain diary for adolescents with arthritis. Clin J Pain 2006;22(3):295-305. [doi: 10.1097/01.ajp.0000173371.54579.31] [Medline: 16514331]
26. Kristjánsdóttir OB, Fors EA, Eide E, Finset A, Stensrud TL, van Dulmen S, et al. A smartphone-based intervention with diaries and therapist-feedback to reduce catastrophizing and increase functioning in women with chronic widespread pain: randomized controlled trial. J Med Internet Res 2013 Jan 07;15(1):e5 [FREE Full text] [doi: 10.2196/jmir.2249] [Medline: 23291270]
27. Belisario JS, Huckvale K, Greenfield G, Car J, Gunn LH. Smartphone and tablet self management apps for asthma. Cochrane Database Syst Rev 2013 Nov 27(11):CD010013 [FREE Full text] [doi: 10.1002/14651858.CD010013.pub2] [Medline: 24282112]
28. Årsand E, Muzny M, Bradway M, Muzik J, Hartvigsen G. Performance of the first combined smartwatch and smartphone diabetes diary application study. J Diabetes Sci Technol 2015 May 14;9(3):556-563 [FREE Full text] [doi: 10.1177/1932296814567708] [Medline: 25591859]
29. Park J, Chu MK, Kim J, Park S, Cho S. Analysis of trigger factors in episodic migraineurs using a smartphone headache diary applications. PLoS One 2016 Feb 22;11(2):e0149577 [FREE Full text] [doi: 10.1371/journal.pone.0149577] [Medline: 26901341]
30. Vellinga A, Galvin S, Duane S, Callan A, Bennett K, Cormican M, et al. Intervention to improve the quality of antimicrobial prescribing for urinary tract infection: a cluster randomized trial. Can Med Asso J 2016 Feb 02;188(2):108-115 [FREE Full text] [doi: 10.1503/cmaj.150601] [Medline: 26573754]
31. Duane S, Tandan M, Murphy AW, Vellinga A. Using mobile phones to collect patient data: lessons learned from the SIMPle Study. JMIR Res Protoc 2017 Apr 25;6(4):e61 [FREE Full text] [doi: 10.2196/resprot.6389] [Medline: 28442451]
32. Vik I, Bollestad M, Grude N, Bærheim A, Mölstad S, Bjerrum L, et al. Ibuprofen versus mecillinam for uncomplicated cystitis--a randomized controlled trial study protocol. BMC Infect Dis 2014 Dec 17;14(1):693 [FREE Full text] [doi: 10.1186/s12879-014-0693-y] [Medline: 25516016]
33. Wild D, Clayson D, Keating K, Gondek K. Validation of a patient-administered questionnaire to measure the activity impairment experienced by women with uncomplicated urinary tract infection: the Activity Impairment Assessment (AIA). Health Qual Life Outcomes 2005 Jul 15;3:42 [FREE Full text] [doi: 10.1186/1477-7525-3-42] [Medline: 16022727]
34. Greiner W, Weijnen T, Nieuwenhuizen M, Oppe S, Badia X, Busschbach J, et al. A single European currency for EQ-5D health states. Results from a six-country study. Eur J Health Econ 2003 Sep 1;4(3):222-231. [doi: 10.1007/s10198-003-0182-5] [Medline: 15609189]
35. Duane S, Vellinga A, Murphy AW, Cormican M, Smyth A, Healy P, et al. COSUTI: a protocol for the development of a core outcome set (COS) for interventions for the treatment of uncomplicated urinary tract infection (UTI) in adults. Trials 2019 Feb 07;20(1):106 [FREE Full text] [doi: 10.1186/s13063-019-3194-x] [Medline: 30732617]
36. O'Connor N, Breen R, Carton M, Mc Grath I, Deasy N, Collins C, et al. Improving the quality of antibiotic prescribing through an educational intervention delivered through the out-of-hours general practice service in Ireland. Eur J Gen Pract 2020 Dec 06;26(1):119-124 [FREE Full text] [doi: 10.1080/13814788.2020.1784137] [Medline: 32757866]

Abbreviations

AIA: Activity Impairment Assessment
EUCAST: European Committee on Antimicrobial Susceptibility Testing
GP: general practitioner
SATIN: Symptomatic Versus Antibiotics Treatment of Urinary Tract Infections in Women
SIMPle: Supporting the Improvement and Management of Prescribing for Urinary Tract Infection
UTI: urinary tract infection

Edited by R Kukafka; submitted 29.10.20; peer-reviewed by O Gbinigie, R Vilela; comments to author 11.12.20; revised version received 04.01.21; accepted 16.03.21; published 03.09.21.

Please cite as:
Vellinga A, Farrell K, Fallon R, Hare D, Sutton-Fitzpatrick U, Cormican M
Presentation, Treatment, and Natural Course of Severe Symptoms of Urinary Tract Infections Measured by a Smartphone App: Observational and Feasibility Study
J Med Internet Res 2021;23(9):e25364
URL: https://www.jmir.org/2021/9/e25364
doi:10.2196/25364
PMID:34477567

Viewpoint

How Should Oncologists Choose an Electronic Patient-Reported Outcome System for Remote Monitoring of Patients With Cancer?

Fabrice Denis[1], MD, PhD; Ivan Krakowski[2], MD

[1]ELSAN, Jean Bernard Institute, Le Mans, France

[2]Association Francophone pour les Soins Oncologiques de Support, Bègles, France

Corresponding Author:
Fabrice Denis, MD, PhD
ELSAN
Jean Bernard Institute
9 Rue Beauverger
Le Mans, 72000
France
Phone: 33 0684190480
Email: fabrice.denis.home@gmail.com

Abstract

Electronic patient-reported outcome (ePRO) systems for symptom monitoring in patients with cancer have shown quality of life and survival benefits in controlled trials. They are beginning to be used in routine oncology practice. Many software developers provide software solutions for clinicians, but how should clinicians decide which system to use? We propose a synthesis of the main questions regarding the effectiveness, safety, and functionality of an ePRO system that a clinician should ask software providers to assist in the selection of a software product in order to obtain the best value tools for their patients and their practice.

(J Med Internet Res 2021;23(9):e30549) doi:10.2196/30549

KEYWORDS

ePRO; cancer; remote monitoring; quality; effectiveness; security; digital monitoring; digital health; cancer patients; patients with cancer; oncology

Background

Electronic patient-reported outcome (ePRO) systems for symptom monitoring of patients with cancer have had demonstrated quality of life and survival benefits in controlled trials (Table 1). They also result in a reduction of emergency hospitalization and have favorable cost-effectiveness and clinical utility; in addition, users have good perceptions of ePRO systems [1-9].

The practical integration of these ePRO systems into patient care is increasing and the positive results of trials have contributed to their increased use. The organization of medical teams is challenging because it requires the presence of a nurse dedicated to initial alert management and a dedicated time for the physician to respond to the alert.

It is challenging for clinicians and patients to identify systems that will add value to patient care in real life because the number of ePRO tools available is rapidly increasing, making it difficult for physicians to choose just one.

Table 1. Randomized studies of remote monitoring of patients with cancer by electronic patient-reported outcome systems.

Authors	Number of patients	Indications	Questionnaires used	Multicentric trial	Type of cancer	Improved outcome
Basch, 2017 [1] and 2016 [9]	766	Toxicity monitoring	NCI-CTCAE[a]	No	All	Quality of life, survival, reduced emergency use
Mir, 2020 [2]	609	Toxicity monitoring	PRO-CTCAE[b]	No	All	Dose intensity
Basch, 2020 [3]	1191	Toxicity monitoring, follow-up, supportive care	PRO-CTCAE	Yes	All	Quality of life, symptom control
Absolom, 2021 [4]	508	Toxicity monitoring	NCI-CTCAE	No	All	Quality of life, symptom control
Berry, 2014 [5]	752	Toxicity monitoring, supportive care	SDS-15[c]	Yes	All	Symptom control
Strasser, 2016 [6]	264	Toxicity monitoring, supportive care	ESAS[d]	Yes	All	Symptom control
Denis, 2019 [7]	133	Follow-up	Not reported	Yes	Lung	Survival
Moony, 2021 [8]	252	Toxicity monitoring	Not reported	No	All	Quality of life, symptom control

[a]NCI-CTCAE: National Cancer Institute version of the Common Terminology Criteria for Adverse Events.

[b]PRO-CTCAE: Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events.

[c]SDS-15: Symptom Distress Scale-15.

[d]ESAS: Edmonton Symptom Assessment System.

Many criteria can be used to select ePRO systems, although there is little scientific literature focused on ePRO systems in oncology. However, some criteria on effectiveness, security, and functionality are usually included in telehealth or digital therapeutics assessments [10-12].

In general, ePRO systems must meet specific criteria for both the patient's and clinician's benefit. Here, we propose some questions that physicians may direct toward system providers to help them choose a relevant ePRO software.

Effectiveness Criteria

1. Notifications must be sent to the medical team. This factor appears to be more important than sending a notification to the patient on their smartphone and allowing them to call the medical center. This was included in randomized studies of ePRO systems for patients with cancer [1-9].
2. The solution must be a medical device with quality marking such as a Class II CE mark [13-15]. This ensures that the product complies with the essential requirements of the relevant European Union legislation in which all devices must be evaluated for clinical efficacy and any side effects, if applicable, by means of preclinical and clinical evaluation. As ePRO systems are Class IIa medical devices, manufacturers must provide full quality assurance such as product and postmarket surveillance (materiovigilance) to obtain conformity. The management of materiovigilance is included in Class IIa CE marking and is mandatory.
3. The solution must use validated algorithms. This validation aims to assess reliability and the performance of algorithms in detecting events; algorithms' performance should be published in journals [1-9,16,17].
4. The solution should contain algorithms that allow adverse events monitoring of the main drugs used in oncology (ie, chemotherapy, immunotherapy, hormonotherapy, targeted therapy, radiotherapy) [1-6,8,16,17].
5. As early supportive care improves survival in oncology, the solution should contain algorithms to detect symptoms that make a patient eligible for early supportive care [3,5,6,18].
6. The solution should allow for follow-up of the patient to detect complications or symptomatic relapse early, which improves clinical utility and patient outcomes [3,7].
7. Clinicians should use solutions with up-to-date algorithms, especially for patients receiving therapies such as maintenance or long-course treatment. As the main new drugs (eg, immunotherapy or targeted therapies) have only been widely used for fewer than 5 years, solutions using algorithms older than this should be avoided. This is mandatory for Class IIa CE marking of medical devices [13-15].
8. Validated questionnaires such as the Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) items from the National Cancer Institute or the Edmonton Symptom Assessment System (ESAS) should be used in the patient questionnaire. The PRO-CTCAE measurement system was developed to capture symptomatic adverse events in patients in cancer clinical trials and is now recommended for ePRO remote monitoring of symptoms because it evaluates the symptom attributes of frequency, severity, and interference. It has also been linguistically validated in more than 30 languages [3,19]. ESAS is one of the most used patient-reported outcome scales for symptom assessment in palliative care and oncology in the past 25 years. Although it only uses unidimensional scales to assess symptom intensity, ESAS has been psychometrically validated, has been translated into numerous languages, and is freely available. A change

of one point was found to be the optimal cutoff for both improvement and deterioration for all 10 symptoms included in ESAS using a sensitivity-specificity approach. As it enables rapid, pragmatic assessment of multiple symptoms simultaneously, ESAS is used extensively in the clinical setting for symptom screening and monitoring worldwide. However, it cannot be used to monitor all treatment toxicities because only 10 symptoms are assessed; the ESAS does not include digestive symptoms (eg, vomiting, diarrhea), cutaneous symptoms, fluid retention (eg, edema), or sepsis. However, its routine use was associated with a 6% increase in 1-year overall survival in a retrospective matched cohort study of 128,893 patients [20-22].

Safety Criteria

1. A real-life study assessment of the solution should be done to confirm that performance and compliance are in line with that reported in premarket studies and to assess the postmarket security of the solution. It is especially important to assess the satisfaction of patients in a real-life study [13].
2. In Europe, the solution must comply with the General Data Protection Regulation (GDPR) that protects and secures patient data. Data protection regulations strengthened the management of alerts and the security of data transmitted to the medical team [23].
3. Oncologists should avoid a solution that allows clinicians to set alert thresholds themselves because the lack of data about the safety and reliability of modified thresholds can result in physician responsibility if a patient experiences an adverse event (eg, following a false negative event).
4. Solutions that allow the creation of "homemade" algorithms should be avoided for the same reasons as described in the previous point.
5. Oncologists should avoid solutions developed by software providers about which a critical security alert was issued by health authorities (eg, risk of a serious adverse event due to critical software dysfunction). Many countries are already publishing these notifications on their health agency's portal (eg, the French national health security agency [24]).

Functionality

1. As not all patients have smartphones, the software should allow for the sending of forms by mail or SMS text message to patients [3].
2. The software must allow different health professionals (physicians and nurses and their colleagues) to manage notifications [3].
3. There should be one software for all major cancers, covering the entire course of treatment (active treatment and surveillance) to avoid the use of multiple software for one patient. This would reduce the technical burden placed on professionals and result in higher levels of satisfaction among patients and health professionals in routine use [3].
4. A single software containing an all-in-one algorithm enabling the detection of toxicities, follow-up management, and supportive care should be prioritized to avoid the use of multiple software with different algorithms. Too many different rules triggering notifications to physicians reduce understandability among users [3].
5. A free-text window triggering notifications when used should be integrated into the patient form to improve communication of other symptoms that are not present on the form or to allow patients to ask health professionals questions [3,7].
6. A software that allows the integration of patient reports into the electronic health record of the health center would be useful to improve the care team's management of notifications [4].
7. A software that contains cancer patient education modules would assist in optimizing treatment tolerance and compliance [25].

These criteria are summarized in Figure 1.

As ePRO systems will become a standard of quality of care for patients with cancer, the use of these solutions will increase, but their efficacy, security, and functionality should be warranted by the software providers to give the same benefit to real-life patients to allow reimbursement for these tools by health authorities.

Figure 1. Criteria synthesis of ePRO systems for remote monitoring of patients with cancer by oncologists. EMR: electronic medical record; ePRO: electronic patient-reported outcome; ESAS: Edmonton Symptom Assessment System; PRO-CTCAE: Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events.

Must have	To be privileged if has	Better if has	Avoid if
Validated algorithms in published trial	All-in-one algorithms (toxicities, follow-up, and supportive care algorithms)	Cancer patient education module	Oncologists can personally set alert thresholds
Class IIa CE marking	One software for all cancers, encompassing the full course of treatment (active treatment and surveillance)	EMR integration	Software developer had a critical security alert issued by health authorities in the past 3 years (eg, serious adverse event due to software dysfunction)
Notification management for physicians and nurses	Adverse events monitoring algorithms for main therapies		Algorithms >5 years without update
Complies with General Data Protection Regulation	PRO-CTCAE / ESAS items		"homemade" algorithms
A single instrument for all types of cancer	Free-text window triggering notification when used		
	Follow-up algorithms		
	Supportive care algorithms		
	Real-life study validation		
	Option to send patient form via mail or SMS text message		

Acknowledgments

No one received compensation for their contribution. We thank Professor Claire Snyder, PhD (Johns Hopkins School of Medicine) and Professor Ethan Basch, MD, MSc (The University of North Carolina at Chapel Hill) for their relevant suggestions.

Authors' Contributions

All authors contributed to the critical revision of the manuscript.

Conflicts of Interest

None declared.

References

1. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment. JAMA 2017 Jul 11;318(2):197-198 [FREE Full text] [doi: 10.1001/jama.2017.7156] [Medline: 28586821]
2. Mir O, Ferrua M, Fourcade A, Mathivon D, Duflot-Boukobza A, Dumont SN, et al. Intervention combining nurse navigators (NNs) and a mobile application versus standard of care (SOC) in cancer patients (pts) treated with oral anticancer agents (OAA): Results of CapRI, a single-center, randomized phase III trial. JCO 2020 May 20;38(15_suppl):2000-2000. [doi: 10.1200/jco.2020.38.15_suppl.2000]
3. Basch E, Stover AM, Schrag D, Chung A, Jansen J, Henson S, et al. Clinical Utility and User Perceptions of a Digital System for Electronic Patient-Reported Symptom Monitoring During Routine Cancer Care: Findings From the PRO-TECT Trial. JCO Clinical Cancer Informatics 2020 Oct(4):947-957. [doi: 10.1200/cci.20.00081]

4. Absolom K, Warrington L, Hudson E, Hewison J, Morris C, Holch P, et al. Phase III Randomized Controlled Trial of eRAPID: eHealth Intervention During Chemotherapy. JCO 2021 Mar 01;39(7):734-747. [doi: 10.1200/jco.20.02015]
5. Berry DL, Hong F, Halpenny B, Partridge AH, Fann JR, Wolpin S, et al. Electronic Self-Report Assessment for Cancer and Self-Care Support: Results of a Multicenter Randomized Trial. JCO 2014 Jan 20;32(3):199-205. [doi: 10.1200/jco.2013.48.6662]
6. Strasser F, Blum D, von Moos R, Cathomas R, Ribi K, Aebi S, Swiss Group for Clinical Cancer Research (SAKK). The effect of real-time electronic monitoring of patient-reported symptoms and clinical syndromes in outpatient workflow of medical oncologists: E-MOSAIC, a multicenter cluster-randomized phase III study (SAKK 95/06). Ann Oncol 2016 Mar;27(2):324-332 [FREE Full text] [doi: 10.1093/annonc/mdv576] [Medline: 26646758]
7. Denis F, Basch E, Septans A, Bennouna J, Urban T, Dueck AC, et al. Two-Year Survival Comparing Web-Based Symptom Monitoring vs Routine Surveillance Following Treatment for Lung Cancer. JAMA 2019 Jan 22;321(3):306-307 [FREE Full text] [doi: 10.1001/jama.2018.18085] [Medline: 30667494]
8. Mooney K, Iacob E, Wilson CM, Lloyd J, Nielson H, Ward JH. Randomized trial of remote cancer symptom monitoring during COVID-19: Impact on symptoms, QoL, and unplanned health care utilization. JCO 2021 May 20;39(15_suppl):12000-12000. [doi: 10.1200/jco.2021.39.15_suppl.12000]
9. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom Monitoring With Patient-Reported Outcomes During Routine Cancer Treatment: A Randomized Controlled Trial. J Clin Oncol 2016 Mar 20;34(6):557-565 [FREE Full text] [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
10. Leigh S, Ouyang J, Mimnagh C. Effective? Engaging? Secure? Applying the ORCHA-24 framework to evaluate apps for chronic insomnia disorder. Evid Based Ment Health 2017 Nov 25;20(4):e20-e20. [doi: 10.1136/eb-2017-102751] [Medline: 28947676]
11. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020 Jan 13;3(1):4 [FREE Full text] [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
12. van Haasteren A, Vayena E, Powell J. The Mobile Health App Trustworthiness Checklist: Usability Assessment. JMIR mHealth uHealth 2020 Jul 21;8(7):e16844 [FREE Full text] [doi: 10.2196/16844] [Medline: 32706733]
13. Medical Devices: Guidance document - Classification of medical devices. European Commission DG Health and Consumer. URL: https://ec.europa.eu/docsroom/documents/10337/attachments/1/translations/en/renditions/pdf [accessed 2021-08-30]
14. Kramer DB, Xu S, Kesselheim AS. How does medical device regulation perform in the United States and the European union? A systematic review. PLoS Med 2012 Jul 31;9(7):e1001276 [FREE Full text] [doi: 10.1371/journal.pmed.1001276] [Medline: 22912563]
15. French-Mowat E, Burnett J. How are medical devices regulated in the European Union? J R Soc Med 2012 Apr 16;105 Suppl 1(1_suppl):S22-S28 [FREE Full text] [doi: 10.1258/jrsm.2012.120036] [Medline: 22508970]
16. Iivanainen S, Alanko T, Vihinen P, Konkola T, Ekstrom J, Virtanen H, et al. Follow-Up of Cancer Patients Receiving Anti-PD-(L)1 Therapy Using an Electronic Patient-Reported Outcomes Tool (KISS): Prospective Feasibility Cohort Study. JMIR Form Res 2020 Oct 28;4(10):e17898 [FREE Full text] [doi: 10.2196/17898] [Medline: 33112242]
17. Iivanainen S, Alanko T, Peltola K, Konkola T, Ekström J, Virtanen H, et al. ePROs in the follow-up of cancer patients treated with immune checkpoint inhibitors: a retrospective study. J Cancer Res Clin Oncol 2019 Mar 21;145(3):765-774 [FREE Full text] [doi: 10.1007/s00432-018-02835-6] [Medline: 30666409]
18. Temel JS, Greer JA, Muzikansky A, Gallagher ER, Admane S, Jackson VA, et al. Early Palliative Care for Patients with Metastatic Non–Small-Cell Lung Cancer. N Engl J Med 2010 Aug 19;363(8):733-742. [doi: 10.1056/nejmoa1000678]
19. Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). NIH National Cancer Institute. URL: https://healthcaredelivery.cancer.gov/pro-ctcae [accessed 2021-08-30]
20. Mackler E, Petersen L, Severson J, Blayney DW, Benitez LL, Early CR, et al. Implementing a Method for Evaluating Patient-Reported Outcomes Associated With Oral Oncolytic Therapy. JOP 2017 Apr;13(4):e395-e400. [doi: 10.1200/jop.2016.018390]
21. Hui D, Bruera E. The Edmonton Symptom Assessment System 25 Years Later: Past, Present, and Future Developments. J Pain Symptom Manage 2017 Mar;53(3):630-643 [FREE Full text] [doi: 10.1016/j.jpainsymman.2016.10.370] [Medline: 28042071]
22. Barbera L, Sutradhar R, Seow H, Mittmann N, Howell D, Earle CC, et al. The impact of routine Edmonton Symptom Assessment System (ESAS) use on overall survival in cancer patients: Results of a population-based retrospective matched cohort analysis. Cancer Med 2020 Oct 14;9(19):7107-7115 [FREE Full text] [doi: 10.1002/cam4.3374] [Medline: 32794634]
23. Article 35 EU GDPR 'Data protection impact assessment'. URL: https://www.privacy-regulation.eu/en/35.htm [accessed 2021-08-30]
24. Informations de sécurité. ANSM. URL: https://ansm.sante.fr/informations-de-securite/ [accessed 2021-08-30]
25. Larson JL, Rosen AB, Wilson FA. The Effect of Telehealth Interventions on Quality of Life of Cancer Patients: A Systematic Review and Meta-Analysis. Telemed J E Health 2018 Jun;24(6):397-405. [doi: 10.1089/tmj.2017.0112] [Medline: 29112484]

Abbreviations

ePRO: electronic patient-reported outcome
ESAS: Edmonton Symptom Assessment System
NCI-CTCAE: National Cancer Institute version of the Common Terminology Criteria for Adverse Events
PRO-CTCAE: Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events

Edited by G Eysenbach; submitted 16.06.21; peer-reviewed by K Ledu, C Helissey; comments to author 08.07.21; revised version received 08.07.21; accepted 02.08.21; published 09.09.21.

Please cite as:
Denis F, Krakowski I
How Should Oncologists Choose an Electronic Patient-Reported Outcome System for Remote Monitoring of Patients With Cancer?
J Med Internet Res 2021;23(9):e30549
URL: https://www.jmir.org/2021/9/e30549
doi:10.2196/30549
PMID:34499046

Review

Application of Artificial Intelligence in Community-Based Primary Health Care: Systematic Scoping Review and Critical Appraisal

Samira Abbasgholizadeh Rahimi[1,2], BEng, PhD; France Légaré[3,4], MD, PhD; Gauri Sharma[5], BA; Patrick Archambault[3,4], MD; Herve Tchala Vignon Zomahoun[4,6], PhD; Sam Chandavong[7], BA; Nathalie Rheault[4,6], MSc; Sabrina T Wong[8,9], RN, PhD; Lyse Langlois[10,11], PhD; Yves Couturier[12], PhD; Jose L Salmeron[13], PhD; Marie-Pierre Gagnon[14], PhD; Jean Légaré[15], PhD

[1]Department of Family Medicine, Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada
[2]Mila-Quebec AI Institute, Montreal, QC, Canada
[3]Department of Family Medicine and Emergency Medicine, Université Laval, Quebec City, QC, Canada
[4]VITAM - Centre de recherche en santé durable, Université Laval, Quebec City, QC, Canada
[5]Faculty of Engineering, Dayalbagh Educational Institute, Agra, India
[6]Quebec SPOR-Support Unit, Quebec City, QC, Canada
[7]Faculty of Science and Engineering, Université Laval, Quebec City, QC, Canada
[8]School of Nursing, University of British Columbia, Vancouver, BC, Canada
[9]Center for Health Services and Policy Research, University of British Columbia, Vancouver, BC, Canada
[10]Department of Industrial Relations, Université Laval, Quebec City, QC, Canada
[11]OBVIA - Quebec International Observatory on the social impacts of AI and digital technology, Quebec City, QC, Canada
[12]School of Social Work, University of Sherbrooke, Sherbrooke, QC, Canada
[13]Department of Data Science, University Pablo de Olavide, Seville, Spain
[14]Faculty of Nursing, Université Laval, Quebec City, QC, Canada
[15]Arthritis Alliance of Canada, Montreal, QC, Canada

Corresponding Author:
Samira Abbasgholizadeh Rahimi, BEng, PhD
Department of Family Medicine, Faculty of Medicine and Health Sciences
McGill University
5858 Côte-des-Neiges Road, Suite 300
Montreal, QC
Canada
Phone: 1 514 399 9218
Email: samira.rahimi@mcgill.ca

Abstract

Background: Research on the integration of artificial intelligence (AI) into community-based primary health care (CBPHC) has highlighted several advantages and disadvantages in practice regarding, for example, facilitating diagnosis and disease management, as well as doubts concerning the unintended harmful effects of this integration. However, there is a lack of evidence about a comprehensive knowledge synthesis that could shed light on AI systems tested or implemented in CBPHC.

Objective: We intended to identify and evaluate published studies that have tested or implemented AI in CBPHC settings.

Methods: We conducted a systematic scoping review informed by an earlier study and the Joanna Briggs Institute (JBI) scoping review framework and reported the findings according to PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analysis-Scoping Reviews) reporting guidelines. An information specialist performed a comprehensive search from the date of inception until February 2020, in seven bibliographic databases: Cochrane Library, MEDLINE, EMBASE, Web of Science, Cumulative Index to Nursing and Allied Health Literature (CINAHL), ScienceDirect, and IEEE Xplore. The selected studies considered all populations who provide and receive care in CBPHC settings, AI interventions that had been implemented, tested, or both, and assessed outcomes related to patients, health care providers, or CBPHC systems. Risk of bias was assessed using the Prediction Model Risk of Bias Assessment Tool (PROBAST). Two authors independently screened the titles and abstracts of the identified records, read the selected full texts, and extracted data from the included studies using a validated extraction

form. Disagreements were resolved by consensus, and if this was not possible, the opinion of a third reviewer was sought. A third reviewer also validated all the extracted data.

Results: We retrieved 22,113 documents. After the removal of duplicates, 16,870 documents were screened, and 90 peer-reviewed publications met our inclusion criteria. Machine learning (ML) (41/90, 45%), natural language processing (NLP) (24/90, 27%), and expert systems (17/90, 19%) were the most commonly studied AI interventions. These were primarily implemented for diagnosis, detection, or surveillance purposes. Neural networks (ie, convolutional neural networks and abductive networks) demonstrated the highest accuracy, considering the given database for the given clinical task. The risk of bias in diagnosis or prognosis studies was the lowest in the participant category (4/49, 4%) and the highest in the outcome category (22/49, 45%).

Conclusions: We observed variabilities in reporting the participants, types of AI methods, analyses, and outcomes, and highlighted the large gap in the effective development and implementation of AI in CBPHC. Further studies are needed to efficiently guide the development and implementation of AI interventions in CBPHC settings.

(J Med Internet Res 2021;23(9):e29839) doi:10.2196/29839

KEYWORDS

artificial intelligence; machine learning; community-based primary health care; systematic scoping review

Introduction

The use of artificial intelligence (AI) in primary health care has been widely recommended [1]. AI systems have been increasingly used in health care, in general [2], given the hope that such systems may help develop and augment the capacity of humans in such areas as diagnostics, therapeutics, and management of patient-care and health care systems [2]. AI systems have the capability to transform primary health care by, for example, improving risk prediction, supporting clinical decision making, increasing the accuracy and timeliness of diagnosis, facilitating chart review and documentation, augmenting patient–physician relationships, and optimizing operations and resource allocation [3].

Community-based primary health care (CBPHC) is a society-wide approach to primary health care that involves a broad range of prevention measures and care services within communities, including health promotion, disease prevention and management, home care, and end-of-life care [4]. CBPHC incorporates health service delivery from personal to community levels and is the first and most frequent point of contact for the patients with health care systems for patients in many countries, including Canada [4]. In addition to providing comprehensive health care and its importance within healthcare systems, CBPHC has also been identified as essential in formulating evidence-informed public health policies [5]. Given the growing role of primary health care and CBPHC in our society [6], it is important to develop strategies that address the limitations of the existing health care system and enhance the overall quality of care delivered alongside all other aspects of CBPHC. This includes efforts for reducing the growing health care burden of CBPHC providers as well as the burden of chronic diseases, decreasing rates of misclassification and misdiagnosis, reducing cases of mismanaged diseases, and increasing accessibility to care [7-17].

Indeed, integration of AI into CBPHC could help in a variety of ways, including identifying patterns, optimizing operations, and gaining insights from clinical big data and community-level data that are beyond the capabilities of humans. Over time, using AI in CBPHC could lessen the excessive workload for health care providers by integrating large quantities of data and knowledge into clinical practice and analyzing these data in ways humans cannot, thus yielding insights that could not otherwise be obtained. This will allow health care providers to devote their time and energy to the more human aspects of health care [18]. Several studies have reported early successes of AI systems for facilitating diagnosis and disease management in different fields, including radiology [19], ophthalmology [20], cardiology [21], orthopedics [22], and pathology [23]. However, the literature also raises doubts about using and implementing AI in health care [24,25]. Aspects including privacy and consent, explainability of the algorithms, workflow disruption, and the “Frame Problem” that is defined as unintended harmful effects from issues not directly addressed for patient care [26].

Despite the potential advantages, disadvantages, and doubts, there is no comprehensive knowledge synthesis that clearly identifies and evaluates AI systems that have been tested or implemented in CBPHC. Thus, we performed a systematic scoping review aiming to (1) summarize existing studies that have tested or implemented AI methods in CBPHC; (2) report evidence regarding the effects of different AI systems’ outcomes on patients, health care providers, or health care systems, and (3) critically evaluate current studies and provide future directions for AI-CBPHC researchers.

Methods

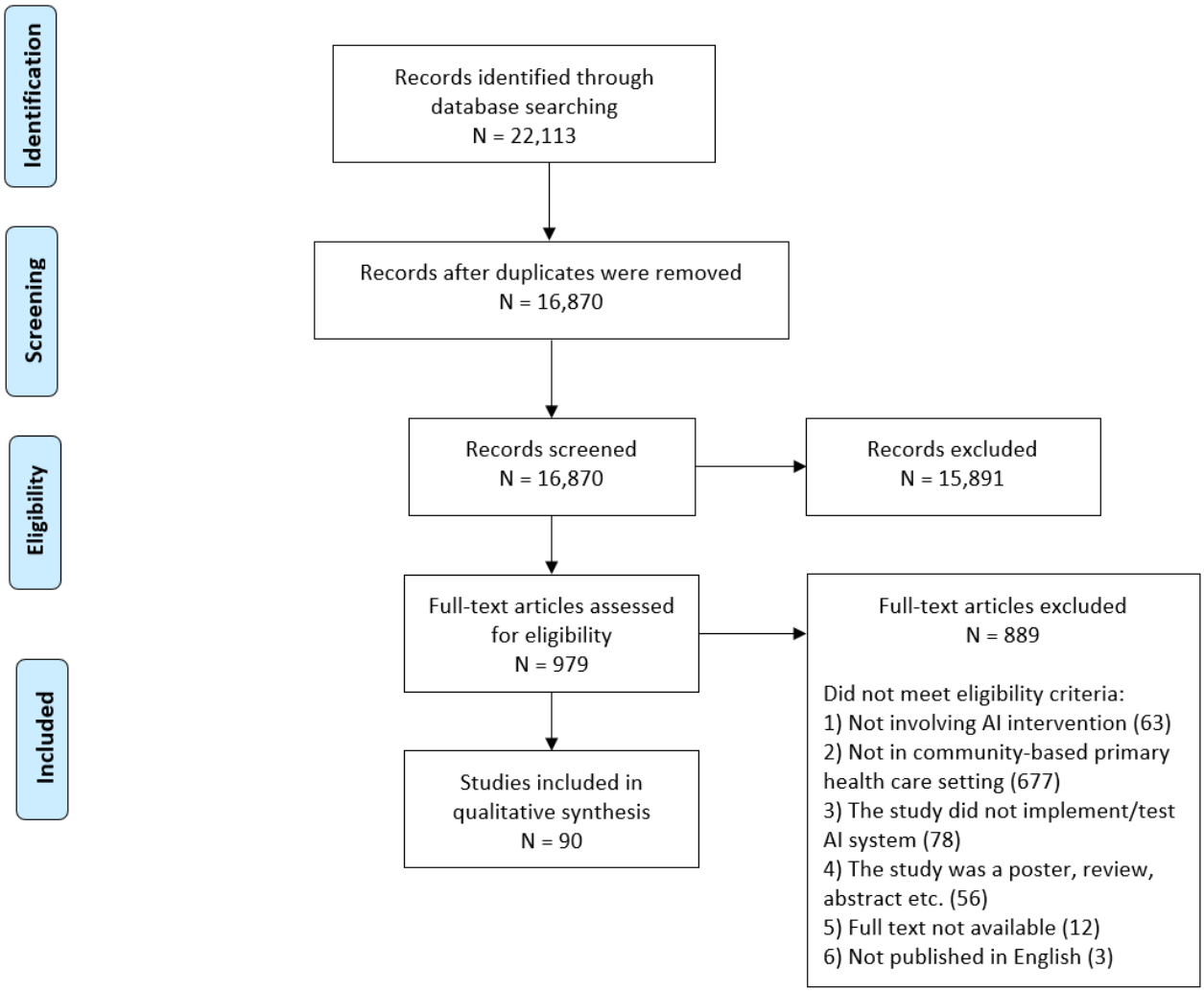
Study Design

Based on the scoping review methodological framework proposed by Levac et al [27], and the Joanna Briggs Institute (JBI) methodological guidance for scoping reviews [28], we developed a protocol with the following steps: (1) clarifying the purpose of the review and linking it to a research question, (2) identifying relevant studies and balancing feasibility with breadth and comprehensiveness, (3) working in a team to iteratively select studies and extract their data, (4) charting the extracted data, incorporating a numerical summary, (5) collating, summarizing, and reporting the results, and (6) consulting the results regularly with stakeholders throughout regarding emerging and final results. This protocol is registered and

available on the JBI website and the Open Science Framework (OSF) websites. We completed this review as per the published protocol.

We formed a multidisciplinary committee of experts in public health, primary health care, AI and data science, knowledge translation, and implementation science, as well as a patient partner and an industry partner (with expertise in the AI-health domain) with whom we consulted during all the steps of the scoping review. This helped us to interpret the results. The screening process is shown in Figure 1. Our review is reported according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analysis-Scoping Reviews) reporting guideline for reporting the study [29] (see Multimedia Appendix 1). Studies that did not report their study design are categorized by methodology according to the classification outlined by the National Institute for Health and Care Excellence [30].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the selection procedure. AI: artificial intelligence.



We used the Prediction Model Risk of Bias Assessment Tool (PROBAST) tool for assessing the risk of bias, which includes 20 signaling questions to facilitate structured judgment of risk of bias organized in four domains of potential biases related to the following: (1) participants (covers potential sources of bias related to participant selection methods and data sources); (2) predictor variables (covers potential sources of bias related to the definition and measurement of predictors evaluated for inclusion in the model); (3) outcomes (covers potential sources of bias related to the definition and measurement of the outcomes predicted by the model); and (4) analyses (covers potential sources of bias in the statistical analysis methods) [31]. Risk of bias was judged as low, high, or unclear. If one or more domains were judged as having high risk of bias, the overall judgment was "high risk" [31].

Eligibility Criteria

We defined our bibliographic database search strategy for peer-reviewed publications in English or French using the Population, Intervention, Comparison, Outcomes, Setting and Study (PICOS) design components [32].

Population

Studies about any population that provides health care services, including nurses, social workers, pharmacists, dietitians, public health practitioners, physicians, and community-based workers (an unregulated type of provider) were included, as were those

about any populations who receive CBPHC services. We adhered to the definition of CBPHC provided by the Canadian Institutes of Health Research (CIHR) (ie, the broad range of primary prevention measures including public health, and primary care services within the community, including health promotion and disease prevention; the diagnosis, treatment, and management of chronic and episodic illness; rehabilitation support; and end-of-life care) [4]. Studies that took place in any CBPHC points of care, including community health centers, primary care networks, clinics, and outpatient departments of hospitals, were also included. Studies conducted in emergency departments were excluded.

Intervention

Only studies that "tested" or "implemented" or "tested and implemented" AI methods, such as computer heuristics, expert systems, fuzzy logic, knowledge representation, automated reasoning, data mining, and machine learning (eg, support vector machines, neural networks, and Bayesian networks) were included. Studies related to robot-assisted care were excluded.

Comparison

No inclusion or exclusion criteria were considered.

Outcomes

The primary outcomes of interest were those related to individuals receiving care (eg, cognitive outcomes, health outcomes, behavioral outcomes), providers of care (eg, cognitive outcomes, health outcomes, behavioral outcomes), and health care systems (eg, process outcomes). Moreover, we analyzed the outcomes of the AI systems for their accuracy and impact on the outcomes of care.

Analysis Methods

All study designs using qualitative, quantitative, or mixed methods were eligible for inclusion. In particular, we included experimental and quasi-experimental studies (randomized controlled trials, quasi-randomized controlled trials, nonrandomized clinical trials, interrupted time series, and controlled before-and-after studies), and observational (cohort, case control, cross- sectional, and case series), qualitative (ethnography, narrative, phenomenological, grounded theory, and case studies), and mixed methods studies (sequential, convergent).

Information Sources and Search Criteria

An information specialist with an epidemiologist, an AI-healthcare researcher, and a family doctor developed a comprehensive search strategy and Medical Subject Headings (MeSH) mediated by the National Library of Medicine. The systematic search was conducted from inception until February 2020 in seven bibliographic databases: Cochrane Library, MEDLINE, EMBASE, Web of Science, Cumulative Index to Nursing and Allied Health Literature (CINAHL), ScienceDirect, and IEEE Xplore. Retrieved records were managed with EndNote X9.2 (Clarivate) and imported into the DistillerSR review software (Evidence Partners, Ottawa, ON) to facilitate the selection process (see Multimedia Appendix 2 for the search strategies used on each database).

Study Selection Process

Title and Abstract Screening (Level 1)

Using DistillerSR, two independent reviewers conducted a pilot screening session using a questionnaire based on our eligibility criteria to test the screening tool and to reach a common understanding. Then, the two reviewers independently screened the titles and abstracts of the remaining records. A third reviewer resolved disagreements between the two reviewers.

Full-Text Screening (Level 2)

Using DistillerSR and the abovementioned questionnaire, the same two reviewers independently assessed the full texts selected at level 1 for their eligibility to be included in the review. A third reviewer resolved conflicting decisions. For those references for which we did not have full-text access, we attempted to obtain access through the interlibrary loan mechanism at the McGill University Library. Studies that met the eligibly criteria were included for full data extraction.

Data Collection

We used a data extraction form, approved by our consultative committee, that we designed based on the Cochrane Effective Practice and Organisation of Care Review Group (EPOC) data collection checklist [33]. Specifically, we extracted study characteristics (eg, design and country of the corresponding author); population characteristics (eg, number of participants and type of disease or treatment); intervention characteristics (eg, AI methods used); and outcome characteristics, including outcomes related to the patients (eg, cognitive outcomes, health outcomes, behavioral outcomes), providers of care (eg, cognitive outcomes, health outcomes, behavioral outcomes), and health care systems (eg, process outcomes).

Assessment of Risk of Bias in the Included Studies

Two reviewers independently appraised the included studies using the criteria outlined in PROBAST to evaluate the risk of bias in each included study that was eligible for evaluation using PROBAST [31]. A third reviewer verified their appraisals.

Synthesis

We performed a descriptive synthesis [34] to describe the studies in terms of their population (patient, primary care providers), interventions (AI systems, evaluated parameters), and outcomes. The results were arranged according to the PICOS format. The tools and techniques for developing a preliminary synthesis included textual descriptions of the studies, grouping and clustering, and tabulation.

Consultation

Throughout the steps of the review, we regularly updated all members of the research team and requested their feedback. We also presented our preliminary results during a workshop at Université Laval, Québec, Canada, with a multidisciplinary group of experts (in public health, primary care, AI and data science, knowledge translation, implementation science, as well as a patient partner, and an industry partner) and collected their comments and feedback.

Patient Involvement

Using a patient-centered approach, our team co-developed the protocol, conducted the review, and reported the results of this study. We integrated patients' priorities within our research questions, search strategy terms, and outcomes of interest. Our patient partner was involved in each step of the research process, including the definition of the objectives, main analysis, descriptive synthesis, interpretation of preliminary and final results, and dissemination of the results obtained in this study.

Results

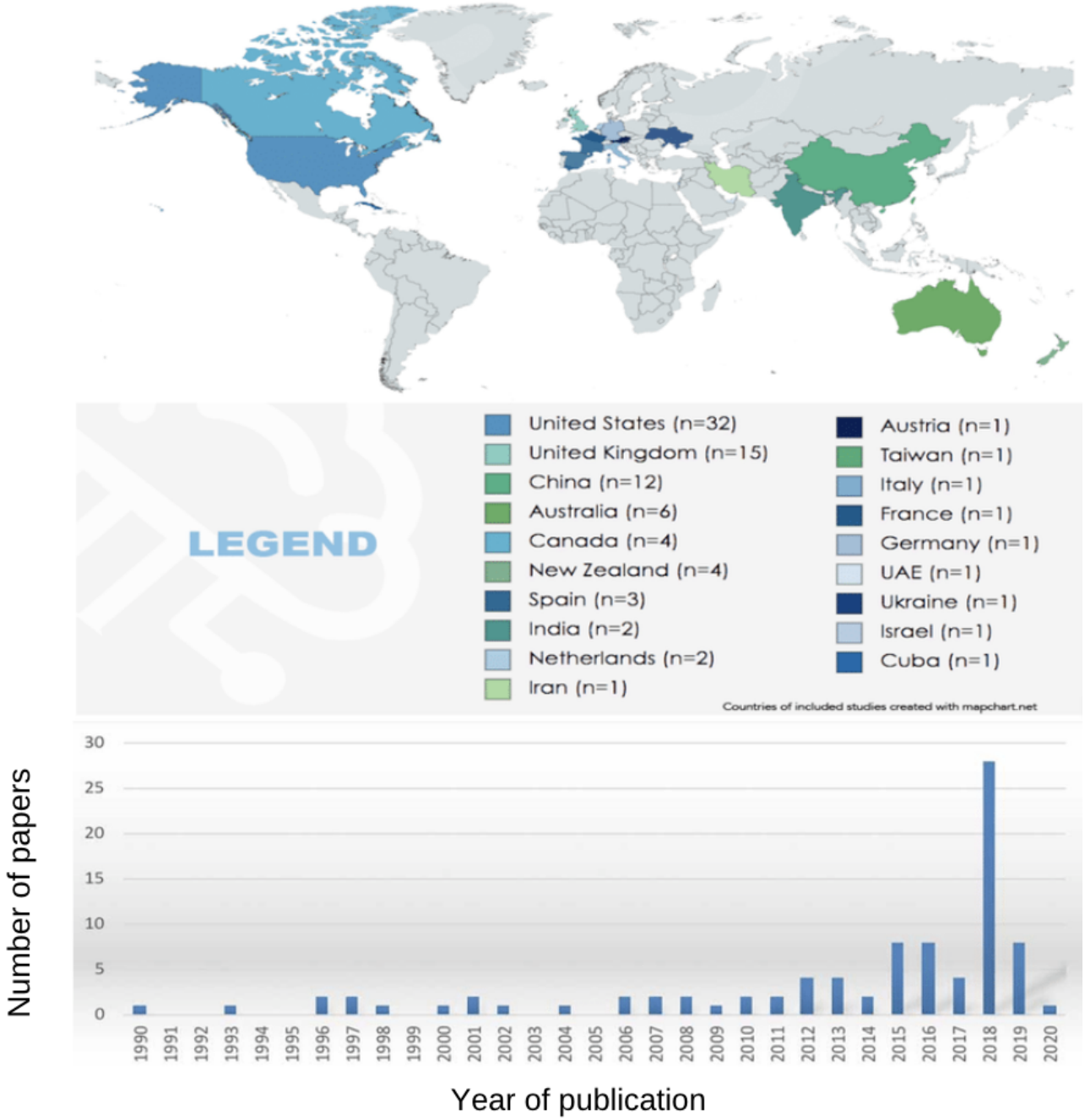
We identified 16,870 unique records. After screening their titles and abstracts, 979 studies remained for full-text review. Ultimately, 90 studies met our inclusion criteria (Figure 1).

Study Characteristics

Countries and Publication Dates

The number of studies published annually has increased gradually since 1990, especially since 2015. Figure 2 shows the timeline of the AI-based studies. Moreover, the four countries publishing a high number of studies are the United States (32/90, 36%), the United Kingdom (15/90, 17%), China (12/90, 13%), and Australia (6/90, 7%). The remaining are New Zealand (4/90, 5%), Canada (4/90, 5%), Spain (3/90, 3%), India (2/90, 2%), and the Netherlands (2/90, 2%), followed by Iran, Austria, Taiwan, Italy, France, Germany, the United Arab Emirates, Ukraine, Israel, and Cuba publishing 1 study each (1%). North America accounts for the highest number of studies (37/90, 41%) followed by Europe (25/90, 28%), Asia (18/90, 20%), and Oceania (10/90, 11%).

Figure 2. Distribution and timeline showing the publication of studies based on artificial intelligence.



Aims of the Included Studies

The included studies sought to describe and test or implement either a novel AI model in CBPHC (16/90, 18%) or an off-the-shelf AI model, which is a modified or improved version of existing AI models in CBPHC (74/90, 82%).

Conceptual Frameworks

Among the 90 studies, 2 (2%) reported using a sociocognitive theoretical framework [35,36]. One of these used the I-change model [35], a model that evolved from several cognitive models, explores the process of behavioral change and the determinants that relate to the change, and focuses on individuals' intentions for adopting innovations [35,37]. In the first study [35] using the I-change model, the authors investigated the cognitive determinants associated with Dutch general practitioners' intention to adopt a smoking cessation expert AI system in their respective practices and found that workload and time constraints are important barriers.

The second study used a continuing medical education framework [38] and compared traditional expert-led training (control group) with an online multimedia-based training activity supplemented with an AI-driven simulation feedback system (treatment group) [36]. Diagnosis accuracy significantly improved in the treatment group when compared to the control group, providing evidence supporting the efficacy of AI medical training methods.

Time Frame of the Collected Data Sets

Among the included studies, 25% (23/90) used data collected over a period of 1 year or less, 20% (17/90) used data collected over a period between 1 and 5 years, 12% (11/20) used data collected over a period between 5 and 10 years, and 9% (8/90) used data collected during more than a 10-year period. One study (1%) used three data sets, collected data from three different sites with over three different time periods (<1 year, 1-5 years, >10 years) [39]. The remaining studies (30/90, 33%) did not specify the time frames of their data set collections.

Population Characteristics

Patients

Sample Size

Overall, 88% (79/90) of the included studies reported their sample size. A total of 21,325,250 patients participated in the testing, training, or validation of the AI systems.

Sex, Gender, and Age

Among the 79 studies reporting their sample size, 46 (58%) reported the sex distribution and none of the studies reported on gender-relevant indicators. Further, 32 (41%) reported the participants' mean age and standard deviation. Overall, the mean age of the participants in these studies was 60.68 (±12.15) years. Age was reported as a range in 21% (17/79) of the studies reporting the sample size, and the remaining 38% (30/79) did not report the age of their participants.

Ethnicity

Among all the included studies, 22% (19/79) reported the participants' ethnic origins, which included Caucasian, Asian-Middle eastern, South Asian, African, American Indian, Alaskan Native, Hispanic, Pacific Islander, Māori, and mixed (Table 1).

Table 1. Characteristics of the participants in the included studies (N=90).

Participant characteristics	Value
Patients	
Total number	21,325,250
Female	2,087,374
Male	1,814,912
Did not report the sex	17,422,964
Age (years), mean (SD)	60.68 (12.15)
Number of studies reporting the sample size of patients (n)	79
Health care providers	
Total number	2,581
Female	467
Male	224
Did not report the sex	1,890
Age (years), mean (SD)	48.50 (7.59)
Number of studies reporting the sample size of health care providers (n)	17
Ethnicities reported for patients (number)	
Caucasian	814,467
Asian	8550
African	42,057
American Indian/Alaskan native	13
Hispanic	5066
Mixed ethnicity	11
Unknown	2,241,937
Number of studies reporting patients' ethnicities (n)	19
Number of studies reporting health care providers ethnicities (n)	0

Other Sociodemographic Information

Only 27% (25/90) of the included studies reported other sociodemographic characteristics of their participants. Socioeconomic status (ie, income level) was the most commonly reported (12/90, 13%). Other characteristics reported were educational status, marital status, area of residence, employment status, smoking status, and insurance status.

Health Care Providers

Among the 90 included studies, 55 (61%) reported the involvement of primary health care providers. Further, 41 of these 55 studies (75%), involved general practitioners, 5 (9%) included nurses, 1 (2%) involved psychiatrists, 1 (2%) involved occupational therapists, and 1 (2%) involved an integrated care specialist. Six studies (7%) involved general practitioners together with other types of health care providers, specifically nurses (3/55, 5%), physician assistants, (1/55 2%), nurses, surgeons, and non-surgeon specialists, (1/55, 2%) and respirologists (1/55; 2%).

Sample Size

Among these 55 studies, 17 (31%) reported the sample size. The data pertaining to 2581 primary health care providers were collected in these studies.

Five of these studies (29%) reported the sex distribution and none reported on gender-relevant indicators. Moreover, 2 (12%) studies reported the age of the primary health care provider participants. The mean age and SD obtained in all the studies for which we collected information is 48.50 (±7.59) years (Table 1).

Sociodemographic Information

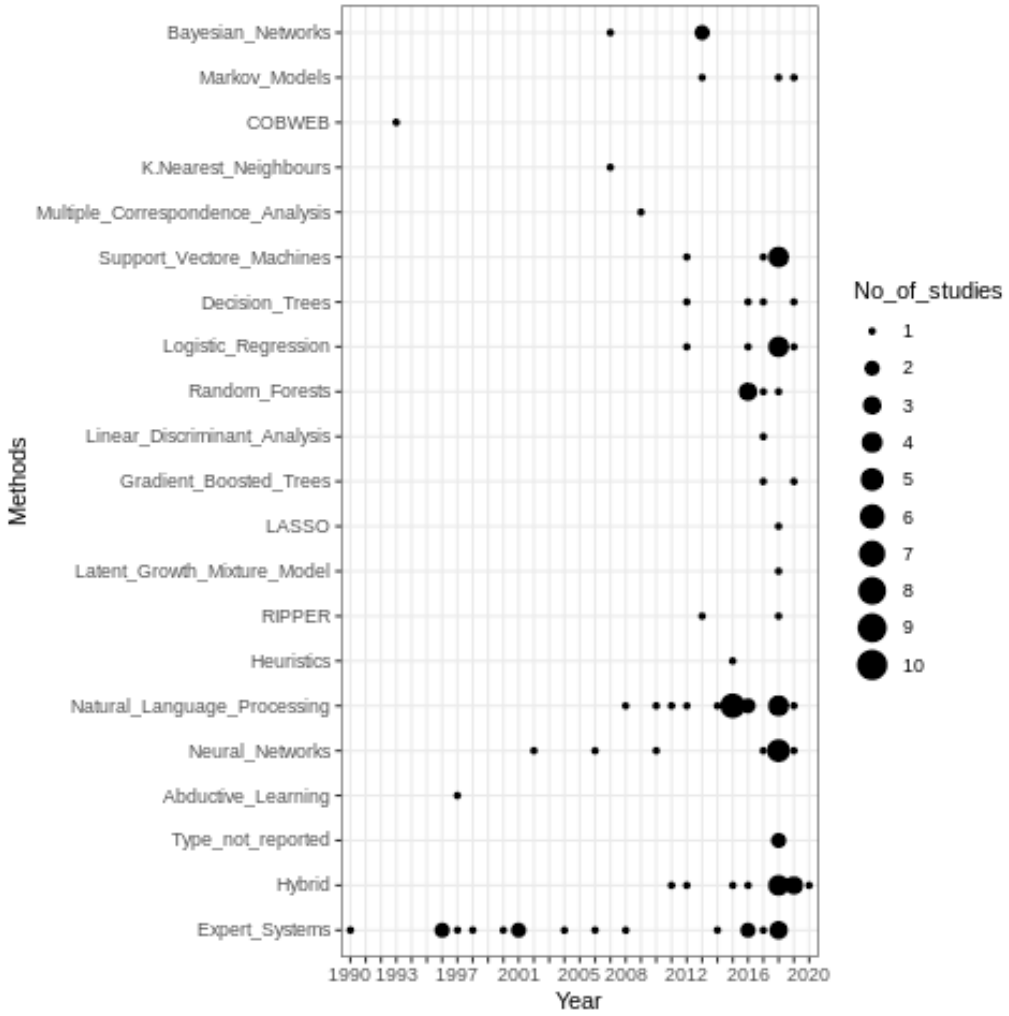
Out of 17 studies, only 1 (5%) reported the primary health care providers' locations of practice. Among the 120 providers in this study, 57 providers practiced in rural areas and 63 practiced in urban areas.

Intervention

AI Methods

Most of the included studies (78/90, 86%), used a single AI method (non-hybrid) and the remaining 14% (n=12) used hybrid AI models—meaning that they integrated multiple AI methods. The most commonly used methods were machine learning (ML) (41/90, 45%) and natural language processing (NLP), including applied ML for NLP (24/90, 27%), and expert systems (17/90, 19%). Figure 3 illustrates the number of studies published according to the type of AI method and year of publication (see Multimedia Appendices 3 and 4 for details regarding the AI methods).

Figure 3. Number of studies published according to the artificial intelligence method used and years of publication.



Performance Measures of AI Interventions

In terms of evaluating the performance of AI models, we considered the following performance metrices: True positive (TP), True negative (TN), False positive (FP), False negative (FN), sensitivity, specificity, precision, F1 score (ie, the weighted average of precision and recall, and area under the curve [AUC]). Among the 90 included studies, 31 (34%) did not report the performance of their models. Among the 59 studies that reported model performance, 13 (22%) used 2 or more performance measures and the remaining 46 (78%) used one measure (see Multimedia Appendix 4 for detailed information on studies' AI methods used in the included studies and their performance measures).

Generated Knowledge

Most of the included studies (81/90, 91%) were either diagnosis- or prognosis-related or focused on surveillance, and the remaining involved operational aspects (eg, resource allocation, system- level decisions) (see Multimedia Appendix 4 for detailed information).

Health Conditions

The majority of the 90 included studies (68/90, 76%) investigated the use of AI in relation to a specific medical condition. Conditions studied were vascular diseases including hypertension, hypercholesteremia, peripheral arterial disease, and congestive heart failure (10/90, 11%) [40-49]; infectious diseases including influenza, herpes zoster, tuberculosis, urinary tract infections, and subcutaneous infections (8/90, 9%) [50-57]; type 2 diabetes (5/90, 6%) [58-62]; respiratory disorders including chronic obstructive pulmonary disease and asthma (6/90, 8%) [63-69]; orthopedic disorders including rheumatoid arthritis, gout, and lower back pain (5/90, 5%) [36,39,70-72]; neurological disorders including stroke, Parkinson disease, Alzheimer disease [73-75], and cognitive impairments (6/90, 5%) [76,77]; cancer including colorectal cancer, and head and neck cancer (4/90, 4%) [78-81]; psychological disorders including depression and schizophrenia (3/90, 3%) [82-84]; diabetic retinopathy (3/90, 3%) [85-87]; suicidal ideations (2/90, 2%) [88,89]; tropical diseases including malaria (2/90, 2%) [90,91]; renal disorders (2/90, 2%) [92,93]; autism spectrum disorder (2/90, 2%) [94,95]; venous disorders including deep vein thrombosis and venous ulcers (2/90, 2%) [96,97]; and other health conditions (8/90, 8%) [98-105].

Data Sets (Training, Testing, and Validation)

In this section, we briefly explain the training, testing, and validation of the data sets, and then present our results. The training data set is the subset of the data that are used to fit in the initial AI model and to train it. The testing data set is the subset of the data used to evaluate the model that fits the initial training data set. The validation data set is a subset of the data used to conduct an unbiased evaluation of the model that fits the training data set, while simultaneously optimizing the model's hyperparameters, namely the parameters whose values are used to control the learning process [106]. The evaluation of these parameters is important because it provides information about the accuracy of predictions made by the AI model, and the prospective effects of hyperparameter tuning [107].

Among the 90 included studies, 9 (10%) reported on all three data sets, 33 (36%) reported on the training and testing data sets, and 36 (40%) reported on the training and validation data sets. No descriptions of these data sets were provided in 49 (54%) of the included studies.

Legal Information and Data Privacy

Legal information concerning privacy was mentioned in 4% (4/90) of the studies in our review. Although health care records were anonymized to protect participants' information in all four of these studies, only one explicitly reported ensuring data collection, storage, and sharing security. The remining studies did not report on data privacy and other legal information.

Involvement of Users

Development

Two of the 90 included studies (2%) reported about the AI developers, all of whom were engineers [60,86]. None of the studies reported the involvement of the end users, including health care providers and patients, in the development stage.

Testing and Validation

Seven out of the 90 (8%) included studies reported information about those who participated in testing or validating the AI. This included general practitioners and nurses [86], engineers [60], general practitioners [51,81], occupational therapists [74], respirologists [64], and nurses [108].

Outcomes

Extraction of the data related to the benefits for patients, primary health care providers, and the health system explained in this section was conducted according to what the authors of the included studies clearly reported as specific benefits to each of these categories.

Potential Benefits for Patients

Included studies reported the following potential benefits of implementing AI in CBPHC: improvements in treatment adherence, person-centered care, quality of life, timeliness of high-risk patient identification, screening speed and cost-effectiveness, enhanced predictability of morbidities and risk factors, benefits related to early diagnosis, as well as early prevention of diseases for the elderly, and facilitated referrals.

Potential Benefits for Primary Health Care Providers

The included studies reported the following information regarding primary health care provider-related benefits of implementing AI in CBPHC: enhanced interprofessional communication and quality of primary care delivery, reduced workload of these providers, and facilitation of referrals and patient-centered care.

Other benefits included benefits with respect to use of AI as a reminder system, application of AI tools to inform commissioning health care priorities, the benefit of an AI system as a quality improvement intervention by generating warnings in electronic medical records and analyzing clinical reports, facilitating monitoring of the diseases, and using AI to reduce health risks.

Potential Benefits for the Health Care System

Studies in our review found that AI can play a role in improving individual patient care and population-based surveillance, can be beneficial by providing predictions to inform and facilitate policy makers decisions regarding the effective management of hospitals, benefits to community-level care, cost-effectiveness, and reducing burden at the system level.

Economic Aspects

Only one study (1%) among the included 90 papers assessed the cost-effectiveness of the AI system studied. The Predicting Out-of-Office Blood Pressure in the Clinic [PROOF-BP] system that the study authors developed for the diagnosis of hypertension in primary care was found to be cost-effective compared to conventional blood pressure diagnostic options in primary care [49].

Challenges of Implementing AI in CBPHC

Our results suggest that challenges of using AI in CBPHC include complications related to the variability of patient data as well as barriers to use AI systems or to participate in AI research owing to the age or cognitive abilities of patients.

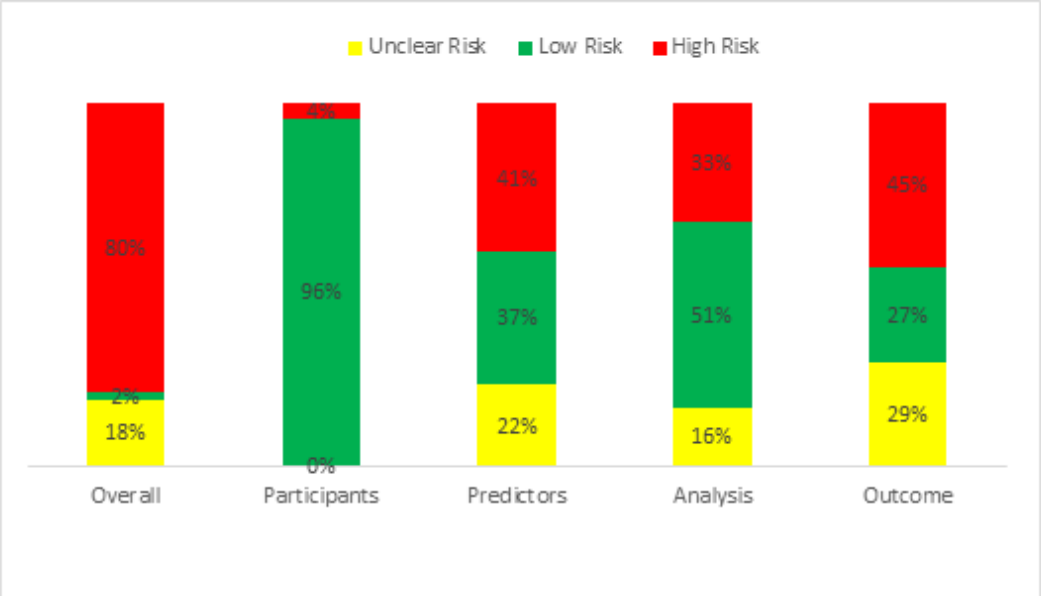
With respect to the health care system, our review found challenges related to how information is recorded (eg, the use of abbreviations in medical records), poor interprofessional communication between nurses and physicians, inconsistent medical tests, and a lack of event recording in cases of communication failures. The included studies also mentioned problems with respect to the restricted resources and administrative aspects such as legislations and administrative approvals, as well as challenges with respect to the lack of digital or computer literacy among the primary health care providers.

In the included studies, other challenges were reported at the level of the health care system such as the data available for use with AI as well as challenges at the level of AI itself (eg, complexity of the system and difficulty in interpretation). The following were identified as the main barriers regarding the data: (1) insufficient data to train, test, and validate AI systems, leading to negative impacts on the robustness of AI models and the accuracy of their predictions; (2) poor quality data, inaccuracies in the data, misclassifications, and lack of representative data; (3) deidentification of protected medical data; and (4) variability in the data sets and combining different data sets. Regarding AI, computational complexity and difficulties in interpreting or explaining some AI model compositions were among the barriers at the AI level.

Risk of Bias

We identified the studies that were eligible to be evaluated using PROBAST. Among our included studies, 54% (49/90) were eligible to be evaluaeted using the PROBAST tool and most (39/49, 80%) were at high risk of bias according to our assessment with PROBAST (Figure 4). With respect to risk of bias for each of the four domains assessed, few studies presented risks regarding participants, (2/49, 4%), whereas 45% (22/49) studies exhibited risks of bias regarding outcomes. See Multimedia Appendices 5 and 6 for details on common causes of risks in each study).

Figure 4. Risk of bias graph: assessing risk of bias in five categories namely overall, participants, predictors, analysis, and outcome (presented as percentages).



Discussion

Principal Findings

We conducted a comprehensive systematic scoping review that included 90 studies on the use of AI systems in CBPHC and provided a critical appraisal of the current studies in this area. Our results highlighted an explosion in the number of studies since 2015. We observed variabilities in reporting the participants, type of AI methods, analysis, and outcomes, and highlighted the large gap in the effective development and implementation of AI in CBPHC. Our review led us to make the following main observations.

AI Models, Their Performance, and Risk of Bias

ML, NLP, and expert systems were the most commonly used in CBPHC. Convolutional neural networks and abductive networks were the methods with the highest performance accuracy within the given data sets for the given task. We observed that a small number of studies reported on the development and testing or implementation of a new AI model in their study, and most of the included studies (74/90, 82%) reported on the usage and testing or implementation of an off-the-shelf AI model. Previous work has demonstrated how off-the-shelf models cannot be directly used in all clinical applications [109]. We observed a high risk of overall bias in the diagnosis- and prognosis-related studies. The highest risk of bias was in the outcome, predictor, and analysis categories of the included studies; validation of studies (external and internal) was poorly reported, and calibration was rarely assessed. A high risk of bias implies that the performance of these AI models in a new data set might not be as optimal as it was reported in these studies. Given the high risk of bias observed in the included studies, AI models used in other settings (ie, with other data) may not exhibit the same level of prediction accuracy as observed.

Where to Use AI?

Primary health care providers are more likely to use AI systems for system-level support in administrative or health care tasks and for operational aspects, rather than for clinical making decisions [1]. However, our results show that few AI systems have been used for these purposes in CBPHC. Rather, the existing AI systems are mostly diagnosis- or prognosis-related, and used for disease detection, risk identification, or surveillance. Further studies in this regard are needed to evaluate the reason behind this tendency in addition to studies for proving the efficiency and accuracy of AI models for assisting in clinical decision making within CBPHC settings. In our review, we found that only 2 of the 90 studies used a (sociocognitive) theoretical framework. Future research needs to use knowledge, attitudes, and behavior theories to expand AI usage for clinical decision making, and more efforts are required to develop and validate frameworks guiding effective development and implementation of AI in CBPHC.

Consideration of Age, Sex, and Gender

Our results show that AI-CBPHC research rarely considers sex, gender, age, and ethnicity. In general, the effect of age is rarely investigated in the AI field and ageism is often ignored in the analysis of discrimination. In health research, AI studies that have evaluated facial and expression recognition methods identified bias toward older adults [109]. This bias could negatively affect the accuracy of the predictions made by AI systems that are commonly used by health care providers.

Furthermore, sex and gender are sources of variations in clinical conditions, affecting different aspects including prognosis, symptomatology manifestation, and treatment effectiveness, among others [110,111]. Despite this importance, big data analytics research focusing on health through the sex and gender lens has shown that current data sets are biased given they are incomplete with respect to gender-relevant indicators with sex-disaggregated data. Indeed, less than 35% of the indicators in international databases have full disaggregation with respect to sex [112]. Our results are consistent with this observation, as we found just half of the AI-CBPHC research with patient participants and nearly one-third with health care provider participants described the sex distribution. Moreover, no AI-CBPHC research has reported on gender-relevant indicators. These are important aspects that need to be considered in the future AI-based CBPHC studies to avoid potential biases in the AI systems.

Consideration of Ethnicity and Geographical Location

Less than one quarter of included studies have reported patient participants' ethnicities, with no discussion on the ethnicities of participating health care providers. Moreover, for those studies that reported patient ethnicity, we observed that the collected data were related to causation populations, thus raising questions regarding the representativeness of the data set, leading to biases. Such biases could result in the AI system making predictions that discriminate against marginalized and vulnerable patient populations, ultimately leading to undesirable patient outcomes.

According to our results, most of the AI research in CBPHC has taken place in North American and Europe-centric settings. Several factors contribute to ethnoracial biases when using AI, including not accounting for ethnoracial information, thereby ignoring the different effects illnesses can have on different populations [113]. Consequently, studies can yield results with historical biases as well as biases related to over- or under-representation of population characteristics in data sets and in the knowledge, bases used to build AI systems. In turn, stereotypes and undesirable outcomes may be amplified. Ensuring ethnic diversity in study populations and accounting for this diversity in analyses is an imperative for developing AI systems that result in equitable CBPHC.

Involvement of Users

Despite the many potential benefits of AI to humans, the development of AI systems is often based on "technology-centered" design approaches instead of "human-centered" approaches [114]. Our results indicate that no AI-CBPHC study has involved any end users in the system development stage and involving primary health care professional users during the validation or testing stages has been rare. This results in AI systems that do not meet the needs of health care providers and patients; they suffer from poor usage scenarios and eventually fail during implementation in clinical practice. A recent assessment of the current user-centered design methods showed that most of the existing user-centered design methods were primarily created for non-AI systems and do not effectively address the unique issues in AI systems [115]. Further efforts are needed to include health care providers and patients as users of the developed AI systems in the design, development, validation, and implementation stages in CBPHC. Nevertheless, effectively involving these users in the development, testing, and validation of AI systems remains a challenge; further studies are required to overcome them.

Ethical and Legal Aspects

Ethical and legal challenges related to the use of AI in health care include, but are not limited to, informed consent to use AI, safety and transparency of personal data, algorithmic fairness, influenced by the aforementioned biases, liability, data protection, and data privacy. Our results indicate that ethical and legal aspects have rarely been addressed in AI-CBPHC research, except with respect to privacy and data security issues. There is a need to address all legal and ethical aspects and considerations within AI-CBPHC studies to facilitate implementation of AI in CBPHC settings. For instance, to increase the use of AI systems by CBPHC providers, clarifying scenarios in which informed consent is required could be useful, as would clarifying providers' responsibilities regarding the use of AI systems. To improve patient outcomes related to AI use in CBPHC, defining the responsibilities of providers and researchers regarding the development and implementation of AI-health literacy programs for patients may be necessary, together with gaining an understanding of how and when patients need to be informed about the results that AI systems yield.

Economic Aspects

AI systems can provide solutions to rising health care costs; however, only one (1%) AI-CBPHC study has addressed this issue by conducting a cost-effectiveness analysis of AI use. This is consistent with other study results showing that the cost-effectiveness of using AI in health care is rarely and inadequately reported [116,117]. Thus, further research analyzing cost-effectiveness is needed for identifying the economic benefits of AI in CBPHC in terms of treatment, time and resource management, and mitigation of human error; this would be valuable as it could influence decisions for or against implementing AI in CBPHC.

AI in Clinical Practice

Our results show different barriers and facilitators for implementing AI in clinical practice. Aspects related to the data were among mostly mentioned ones. For instance, the lack of high amounts of quality data, specifically when using modern AI methods (eg, deep learning), is a challenge commonly faced when developing AI systems for use in CBPHC. The promotion of AI-driven innovation in any setting, including CBPHC, is closely linked to data governance, open data directives, and other data initiatives, as they help to establish trustworthy mechanisms and services for sharing, reusing, and pooling data [118] that are required for the development of high-quality data-driven AI systems.

In addition, some data security and privacy laws can create a bottleneck, limiting the use of AI systems in CBPHC and the sharing of health care information that is required for developing high- performance AI systems. To facilitate the implementation and adoption of high-quality AI systems in CBPHC and ensuring benefits to patients, providers and the health care system, research providing insights for addressing these implementation challenges is needed.

Limitations of the Study

Our review has some limitations. Firstly, given that we used the Canadian Institute of Health Research's definition of CBPHC to determine our inclusion criteria and given that the definition of CBPHC differs from one country to another, our search strategy may not have captured all relevant records. Secondly, we excluded studies conducted in emergency care settings. In many countries, emergency departments are the points of access to community-based care. The European Commission recently released a legal framework (risk-based approach) for broad AI governance among EU member states [118] and categorized emergency care and first aid services as "high risk." Requirements of high-quality data, documentation and traceability, transparency, human oversight, and model accuracy and robustness are cited as being strictly necessary to mitigate the risks in these settings [118].

Conclusion

In this systematic scoping review, we have demonstrated the extent and variety of AI systems being tested and implemented in CBPHC, critically evaluated these AI systems, showed that this field is growing exponentially, and exposed knowledge gaps that remain and that should be prioritized in future studies.

Acknowledgments

This study was funded by Canadian Institutes of Health Research's Planning/Dissemination Grants (Principal Investigators [PIs]: SAR and FL), Québec SPOR SUPPORT Unit (PI: SAR) and start-up fund from McGill University (PI: SAR). We acknowledge the support from these institutions. SAR receives salary support from a Research Scholar Junior 1 Career Development Award from the Fonds de Recherche du Québec-Santé (FRQS), and her research program is supported by the Natural Sciences and Engineering Research Council (NSERC) Discovery (grant 2020-05246). FL is a Tier 1 Canada Research Chair in Shared Decision Making and Knowledge Translation. We thank the Québec-SPOR SUPPORT Unit for their methodological support and Paula L. Bush, PhD, for her revisions of the draft of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 104 KB - jmir_v23i9e29839_app1.pdf]

Multimedia Appendix 2
Full search strategy.
[PDF File (Adobe PDF File), 518 KB - jmir_v23i9e29839_app2.pdf]

Multimedia Appendix 3
Timeline of artificial intelligence implementation in community-based primary health care between 1990 and 2020.
[PDF File (Adobe PDF File), 668 KB - jmir_v23i9e29839_app3.pdf]

Multimedia Appendix 4
Data extracted from the included studies.
[PDF File (Adobe PDF File), 324 KB - jmir_v23i9e29839_app4.pdf]

Multimedia Appendix 5
Details on the risk of bias in each evaluated study.
[PDF File (Adobe PDF File), 162 KB - jmir_v23i9e29839_app5.pdf]

Multimedia Appendix 6
Risk of bias graph based on authors' judgments about each risk of bias item presented as percentages.
[PDF File (Adobe PDF File), 262 KB - jmir_v23i9e29839_app6.pdf]

References

1. Liyanage H, Liaw S, Jonnagaddala J, Schreiber R, Kuziemsky C, Terry AL, et al. Artificial intelligence in primary health care: perceptions, issues, and challenges. Yearb Med Inform 2019 Aug;28(1):41-46 [FREE Full text] [doi: 10.1055/s-0039-1677901] [Medline: 31022751]
2. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan;25(1):30-36 [FREE Full text] [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
3. Lin SY, Mahoney MR, Sinsky CA. Ten ways artificial intelligence will transform primary care. J Gen Intern Med 2019 Aug;34(8):1626-1630 [FREE Full text] [doi: 10.1007/s11606-019-05035-1] [Medline: 31090027]
4. Community-Based Primary Health Care. URL: https://cihr-irsc.gc.ca/e/43626.html [accessed 2021-07-28]
5. Canadian Institutes of Health Research. What is community-based primary health care? CBPHC Overview. 2015. URL: https://cihr-irsc.gc.ca/e/44079.html [accessed 2021-07-28]
6. Adashi EY, Geiger HJ, Fine MD. Health care reform and primary care—the growing importance of the community health center. N Engl J Med 2010 Jun;362(22):2047-2050. [doi: 10.1056/nejmp1003729]
7. Bodenheimer T, Chen E, Bennett HD. Confronting the growing burden of chronic disease: Can the U.S. health care workforce do the job? Health Aff (Millwood) 2009 Jan;28(1):64-74. [doi: 10.1377/hlthaff.28.1.64] [Medline: 19124856]
8. Howard J, Clark EC, Friedman A, Crosson JC, Pellerano M, Crabtree BF, et al. Electronic health record impact on work burden in small, unaffiliated, community-based primary care practices. J Gen Intern Med 2013 Jan;28(1):107-113 [FREE Full text] [doi: 10.1007/s11606-012-2192-4] [Medline: 22926633]
9. The Lancet Global Health. Adding quality to primary care. The Lancet Global Health 2018 Nov;6(11):e1139. [doi: 10.1016/s2214-109x(18)30457-1]
10. Cecil E, Bottle A, Majeed A, Aylin P. Patient and health-care factors associated with potentially missed acute deterioration in primary care: a retrospective observational study of linked primary and secondary care data. The Lancet 2019 Nov;394:S30. [doi: 10.1016/s0140-6736(19)32827-2]
11. de Lusignan S, Sadek N, Mulnier H, Tahir A, Russell-Jones D, Khunti K. Miscoding, misclassification and misdiagnosis of diabetes in primary care. Diabet Med 2012 Feb;29(2):181-189. [doi: 10.1111/j.1464-5491.2011.03419.x] [Medline: 21883428]
12. Casas Herrera A, Montes de Oca M, López Varela MV, Aguirre C, Schiavi E, Jardim JR, PUMA Team. COPD underdiagnosis and misdiagnosis in a high-risk primary care population in four Latin American countries. a key to enhance disease diagnosis: the PUMA study. PLoS One 2016 Apr;11(4):e0152266 [FREE Full text] [doi: 10.1371/journal.pone.0152266] [Medline: 27073880]
13. Lanzarotto F, Crimí F, Amato M, Villanacci V, Pillan NM, Lanzini A, Brescia Coeliac Disease Study Group. Is under diagnosis of celiac disease compounded by mismanagement in the primary care setting? a survey in the Italian province of Brescia. Minerva Gastroenterol Dietol 2004 Dec;50(4):283-288. [Medline: 15788984]
14. Statham MO, Sharma A, Pane AR. Misdiagnosis of acute eye diseases by primary health care providers: incidence and implications. Med J Aust 2008 Oct;189(7):402-404. [doi: 10.5694/j.1326-5377.2008.tb02091.x] [Medline: 18837685]
15. Shah TI, Clark AF, Seabrook JA, Sibbald S, Gilliland JA. Geographic accessibility to primary care providers: comparing rural and urban areas in Southwestern Ontario. Can Geogr/Le Géographe canadien 2019 Aug;64(1):65-78. [doi: 10.1111/cag.12557]

16. Forget C. The case of the vanishing Québec physicians: how to improve access to care. CD Howe Institute Commentary(410) 2014 May:1-20.
17. Haggerty J, Pineault R, Beaulieu M, Brunelle Y, Gauthier J, Goulet F, et al. Room for improvement: patients' experiences of primary care in Quebec before major reforms. Can Fam Physician 2007 Jun;53(6):1056-1057 [FREE Full text] [Medline: 17872786]
18. Verghese A, Shah NH, Harrington RA. What this computer needs is a physician: humanism and artificial intelligence. JAMA 2018 Jan;319(1):19-20. [doi: 10.1001/jama.2017.19198] [Medline: 29261830]
19. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer 2018 Aug;18(8):500-510 [FREE Full text] [doi: 10.1038/s41568-018-0016-5] [Medline: 29777175]
20. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, et al. Artificial intelligence and deep learning in ophthalmology. Br J Ophthalmol 2019 Feb;103(2):167-175 [FREE Full text] [doi: 10.1136/bjophthalmol-2018-313173] [Medline: 30361278]
21. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, et al. Artificial intelligence in cardiology. J Am Coll Cardiol 2018 Jun;71(23):2668-2679 [FREE Full text] [doi: 10.1016/j.jacc.2018.03.521] [Medline: 29880128]
22. Olczak J, Fahlberg N, Maki A, Razavian AS, Jilert A, Stark A, et al. Artificial intelligence for analyzing orthopedic trauma radiographs. Acta Orthop 2017 Dec;88(6):581-586 [FREE Full text] [doi: 10.1080/17453674.2017.1344459] [Medline: 28681679]
23. Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. Lancet Oncol 2019 May;20(5):e253-e261. [doi: 10.1016/S1470-2045(19)30154-8] [Medline: 31044723]
24. Sun TQ, Medaglia R. Mapping the challenges of artificial intelligence in the public sector: evidence from public healthcare. Government Information Quarterly 2019 Apr;36(2):368-383. [doi: 10.1016/j.giq.2018.09.008]
25. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial intelligence and the implementation challenge. J Med Internet Res 2019 Jul;21(7):e13659 [FREE Full text] [doi: 10.2196/13659] [Medline: 31293245]
26. Yu K, Kohane IS. Framing the challenges of artificial intelligence in medicine. BMJ Qual Saf 2019 Mar;28(3):238-241. [doi: 10.1136/bmjqs-2018-008551] [Medline: 30291179]
27. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
28. Khalil H, Bennett M, Godfrey C, McInerney P, Munn Z, Peters M. Evaluation of the JBI scoping reviews methodology by current users. Int J Evid Based Healthc 2020 Mar;18(1):95-100. [doi: 10.1097/XEB.0000000000000202] [Medline: 31567603]
29. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
30. National Institute for Health and Care Excellence. The NICE Public Health Guidance Development Process. URL: https://pubmed.ncbi.nlm.nih.gov/27905713/ [accessed 2021-07-28]
31. Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a Tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019 Jan;170(1):51-58. [doi: 10.7326/m18-1376]
32. Stone PW. Popping the (PICO) question in research and evidence-based practice. Appl Nurs Res 2002 Aug;15(3):197-198. [doi: 10.1053/apnr.2002.34181] [Medline: 12173172]
33. Cochrane Effective Practice and Organisation of Care Group (EPOC). Data Collection Checklist. URL: https://methods.cochrane.org/sites/methods.cochrane.org.bias/files/public/uploads/EPOC%20Data%20Collection%20Checklist.pdf [accessed 2021-07-28]
34. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews: a product from the ESRC methods Programme. 2006. URL: https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf
35. Hoving C, Mudde AN, de Vries H. Intention to adopt a smoking cessation expert system within a self-selected sample of Dutch general practitioners. Eur J Cancer Prev 2006 Feb;15(1):82-86. [doi: 10.1097/01.cej.0000186633.81753.8b] [Medline: 16374236]
36. McFadden P, Crim A. Comparison of the effectiveness of interactive didactic lecture versus online simulation-based CME programs directed at improving the diagnostic capabilities of primary care practitioners. J Contin Educ Health Prof 2016;36(1):32-37. [doi: 10.1097/CEH.0000000000000061] [Medline: 26954243]
37. de Vries H, Mudde A, Leijs I, Charlton A, Vartiainen E, Buijs G, et al. The European Smoking Prevention Framework Approach (EFSA): an example of integral prevention. Health Educ Res 2003 Oct;18(5):611-626. [doi: 10.1093/her/cyg031] [Medline: 14572020]
38. Collins J. The continuing professional development of physicians: from research to practice. J Am Coll Radiol 2004 Aug;1(8):609-610. [doi: 10.1016/j.jacr.2003.12.038]
39. Zhou S, Fernandez-Gutierrez F, Kennedy J, Cooksey R, Atkinson M, Denaxas S, UK Biobank Follow-upOutcomes Group, et al. Defining disease phenotypes in primary care electronic health records by a machine learning approach: a case study

in identifying rheumatoid arthritis. PLoS One 2016;11(5):e0154515 [FREE Full text] [doi: 10.1371/journal.pone.0154515] [Medline: 27135409]
40. Chaudhry AP, Afzal N, Abidian MM, Mallipeddi VP, Elayavilli RK, Scott CG, et al. Innovative informatics approaches for peripheral artery disease: current state and provider survey of strategies for improving guideline-based care. Mayo Clin Proc Innov Qual Outcomes 2018 Jun;2(2):129-136 [FREE Full text] [doi: 10.1016/j.mayocpiqo.2018.02.001] [Medline: 30035252]
41. Goldstein MK, Coleman RW, Tu SW, Shankar RD, O'Connor MJ, Musen MA, et al. Translating research into practice: organizational issues in implementing automated decision support for hypertension in three medical centers. J Am Med Inform Assoc 2004;11(5):368-376 [FREE Full text] [doi: 10.1197/jamia.M1534] [Medline: 15187064]
42. Hill NR, Ayoubkhani D, McEwan P, Sugrue DM, Farooqui U, Lister S, et al. Predicting atrial fibrillation in primary care using machine learning. PLoS One 2019;14(11):e0224582 [FREE Full text] [doi: 10.1371/journal.pone.0224582] [Medline: 31675367]
43. Pakhomov SV, Jacobsen SJ, Chute CG, Roger VL. Agreement between patient-reported symptoms and their documentation in the medical record. Am J Manag Care 2008 Aug;14(8):530-539 [FREE Full text] [Medline: 18690769]
44. Pesko MF, Gerber LM, Peng TR, Press MJ. Home health care: nurse-physician communication, patient severity, and hospital readmission. Health Serv Res 2018 Apr;53(2):1008-1024 [FREE Full text] [doi: 10.1111/1475-6773.12667] [Medline: 28217974]
45. Press MJ, Gerber LM, Peng TR, Pesko MF, Feldman PH, Ouchida K, et al. Postdischarge communication between home health nurses and physicians: measurement, quality, and outcomes. J Am Geriatr Soc 2015 Jul;63(7):1299-1305. [doi: 10.1111/jgs.13491] [Medline: 26115315]
46. Safarova MS, Liu H, Kullo IJ. Rapid identification of familial hypercholesterolemia from electronic health records: the SEARCH study. J Clin Lipidol 2016;10(5):1230-1239. [doi: 10.1016/j.jacl.2016.08.001] [Medline: 27678441]
47. Selskyy P, Vakulenko D, Televiak A, Veresiuk T. On an algorithm for decision-making for the optimization of disease prediction at the primary health care level using neural network clustering. Family Med Prim Care Rev 2018;20(2):171-175. [doi: 10.5114/fmpcr.2018.76463]
48. Vijayakrishnan R, Steinhubl SR, Ng K, Sun J, Byrd RJ, Daar Z, et al. Prevalence of heart failure signs and symptoms in a large primary care population identified through the use of text and data mining of the electronic health record. J Card Fail 2014 Jul;20(7):459-464 [FREE Full text] [doi: 10.1016/j.cardfail.2014.03.008] [Medline: 24709663]
49. Monahan M, Jowett S, Lovibond K, Gill P, Godwin M, Greenfield S, et al. Predicting out-of-office blood pressure in the clinic for the diagnosis of hypertension in primary care: an economic evaluation. Hypertension 2018 Feb;71(2):250-261. [doi: 10.1161/hypertensionaha.117.10244]
50. Balas EA, Li ZR, Spencer DC, Jaffrey F, Brent E, Mitchell JA. An expert system for performance-based direct delivery of published clinical evidence. J Am Med Inform Assoc 1996;3(1):56-65 [FREE Full text] [doi: 10.1136/jamia.1996.96342649] [Medline: 8750390]
51. Gu Y, Kennelly J, Warren J, Nathani P, Boyce T. Automatic detection of skin and subcutaneous tissue infections from primary care electronic medical records. Stud Health Technol Inform 2015;214:74-80. [Medline: 26210421]
52. MacRae J, Love T, Baker MG, Dowell A, Carnachan M, Stubbe M, et al. Identifying influenza-like illness presentation from unstructured general practice clinical narrative using a text classifier rule-based expert system versus a clinical expert. BMC Med Inform Decis Mak 2015 Oct;15:78 [FREE Full text] [doi: 10.1186/s12911-015-0201-3] [Medline: 26445235]
53. Rennie TW, Roberts W. Data mining of tuberculosis patient data using multiple correspondence analysis. Epidemiol Infect 2009 Dec;137(12):1699-1704. [doi: 10.1017/S0950268809002787] [Medline: 19450381]
54. Tou H, Yao L, Wei Z, Zhuang X, Zhang B. Automatic infection detection based on electronic medical records. BMC Bioinformatics 2018 Apr;19(Suppl 5):117 [FREE Full text] [doi: 10.1186/s12859-018-2101-x] [Medline: 29671399]
55. Turner NM, MacRae J, Nowlan ML, McBain L, Stubbe MH, Dowell A. Quantifying the incidence and burden of herpes zoster in New Zealand general practice: a retrospective cohort study using a natural language processing software inference algorithm. BMJ Open 2018 May;8(5):e021241 [FREE Full text] [doi: 10.1136/bmjopen-2017-021241] [Medline: 29858420]
56. Zheng C, Luo Y, Mercado C, Sy L, Jacobsen SJ, Ackerson B, et al. Using natural language processing for identification of herpes zoster ophthalmicus cases to support population-based study. Clin Exp Ophthalmol 2019 Jan;47(1):7-14. [doi: 10.1111/ceo.13340] [Medline: 29920898]
57. Burton RJ, Albur M, Eberl M, Cuff SM. Using artificial intelligence to reduce diagnostic workload without compromising detection of urinary tract infections. BMC Med Inform Decis Mak 2019 Aug;19(1):171 [FREE Full text] [doi: 10.1186/s12911-019-0878-9] [Medline: 31443706]
58. Hertroijs DFL, Elissen AMJ, Brouwers MCGJ, Schaper NC, Köhler S, Popa MC, et al. A risk score including body mass index, glycated haemoglobin and triglycerides predicts future glycaemic control in people with type 2 diabetes. Diabetes Obes Metab 2018 Mar;20(3):681-688 [FREE Full text] [doi: 10.1111/dom.13148] [Medline: 29095564]
59. Lappenschaar M, Hommersom A, Lucas PJ, Lagro J, Visscher S, Korevaar JC, et al. Multilevel temporal Bayesian networks can model longitudinal change in multimorbidity. J Clin Epidemiol 2013 Dec;66(12):1405-1416 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.06.018] [Medline: 24035172]

60. Moreno E, Lujan MJA, Rusinol MT, Fernandez PJ, Manrique P, Trivino CA, et al. Type 2 diabetes screening test by means of a pulse oximeter. IEEE Trans Biomed Eng 2017 Feb;64(2):341-351. [doi: 10.1109/TBME.2016.2554661] [Medline: 28113188]
61. Perveen S, Shahbaz M, Keshavjee K, Guergachi A. Prognostic modeling and prevention of diabetes using machine learning technique. Sci Rep 2019 Sep;9(1):13805 [FREE Full text] [doi: 10.1038/s41598-019-49563-6] [Medline: 31551457]
62. Sayadi M, Zibaeenezhad M, Taghi Ayatollahi SM. Simple prediction of type 2 diabetes mellitus via decision tree modeling. Int Cardiovasc Res J 2017;11(2):71-76.
63. Afzal Z, Engelkes M, Verhamme K, Janssens H, Sturkenboom M, Kors J, et al. Automatic generation of case-detection algorithms to identify children with asthma from large electronic health record databases. Pharmacoepidemiol Drug Saf 2013 Aug;22(8):826-833. [doi: 10.1002/pds.3438] [Medline: 23592573]
64. Braido F, Santus P, Corsico A, Di Marco F, Melioli G, Scichilone N, et al. Chronic obstructive lung disease "expert system": validation of a predictive tool for assisting diagnosis. Int J Chron Obstruct Pulmon Dis 2018;13:1747-1753 [FREE Full text] [doi: 10.2147/COPD.S165533] [Medline: 29881264]
65. Gautier V, Rédier H, Pujol JL, Bousquet J, Proudhon H, Michel C, et al. Comparison of an expert system with other clinical scores for the evaluation of severity of asthma. Eur Respir J 1996 Jan;9(1):58-64 [FREE Full text] [doi: 10.1183/09031936.96.09010058] [Medline: 8834335]
66. Hung J, Posey J, Freedman R, Thorton T. Electronic surveillance of disease states: a preliminary study in electronic detection of respiratory diseases in a primary care setting. In: Proceedings/AMIA Annual Symposium. 1998 Presented at: AMIA Annual Symposium; 1998; Florida p. 688-692.
67. Klann JG, Anand V, Downs SM. Patient-tailored prioritization for a pediatric care decision support system through machine learning. J Am Med Inform Assoc 2013 Dec 01;20(e2):e267-e274 [FREE Full text] [doi: 10.1136/amiajnl-2013-001865] [Medline: 23886921]
68. MacRae J, Darlow B, McBain L, Jones O, Stubbe M, Turner N, et al. Accessing primary care Big Data: the development of a software algorithm to explore the rich content of consultation records. BMJ Open 2015 Aug;5(8):e008160 [FREE Full text] [doi: 10.1136/bmjopen-2015-008160] [Medline: 26297364]
69. Morales DR, Flynn R, Zhang J, Trucco E, Quint JK, Zutis K. External validation of ADO, DOSE, COTE and CODEX at predicting death in primary care patients with COPD using standard and machine learning approaches. Respir Med 2018 May;138:150-155 [FREE Full text] [doi: 10.1016/j.rmed.2018.04.003] [Medline: 29724388]
70. Betancourt-Hernandez M, Viera-Lopez G, Serrano-Munoz A. Automatic diagnosis of rheumatoid arthritis from hand radiographs using convolutional neural networks. Revista Cubana De Fisica 2018;35(1):39-43.
71. Jarvik JG, Gold LS, Tan K, Friedly JL, Nedeljkovic SS, Comstock BA, et al. Long-term outcomes of a large, prospective observational cohort of older adults with back pain. Spine J 2018 Sep;18(9):1540-1551. [doi: 10.1016/j.spinee.2018.01.018] [Medline: 29391206]
72. Kerr GS, Richards JS, Nunziato CA, Patterson OV, DuVall SL, Aujero M, et al. Measuring physician adherence with gout quality indicators: a role for natural language processing. Arthritis Care Res (Hoboken) 2015 Feb;67(2):273-279 [FREE Full text] [doi: 10.1002/acr.22406] [Medline: 25047509]
73. Arroyo-Gallego T, Ledesma-Carbayo MJ, Butterworth I, Matarazzo M, Montero-Escribano P, Puertas-Martín V, et al. Detecting motor impairment in early parkinson's disease via natural typing interaction with keyboards: Validation of the neuroQWERTY approach in an uncontrolled at-home setting. J Med Internet Res 2018 Mar;20(3):e89 [FREE Full text] [doi: 10.2196/jmir.9462] [Medline: 29581092]
74. Lee SI, Adans-Dester CP, Grimaldi M, Dowling AV, Horak PC, Black-Schaffer RM, et al. Enabling stroke rehabilitation in home and community settings: a wearable sensor-based approach for upper-limb motor training. IEEE J Transl Eng Health Med 2018;6:1-11. [doi: 10.1109/jtehm.2018.2829208]
75. Tandon R, Adak S, Kaye J. Neural networks for longitudinal studies in Alzheimer's disease. Artif Intell Med 2006 Mar;36(3):245-255. [doi: 10.1016/j.artmed.2005.10.007] [Medline: 16427257]
76. Levy B, Gable S, Tsoy E, Haspel N, Wadler B, Wilcox R, et al. Machine learning detection of cognitive impairment in primary care. J Am Geriatr Soc 2018;66:S111-S111. [doi: 10.1007/978-1-4612-1884-5_7]
77. Ursenbach J, O'Connell ME, Neiser J, Tierney MC, Morgan D, Kosteniuk J, et al. Scoring algorithms for a computer-based cognitive screening tool: an illustrative example of overfitting machine learning approaches and the impact on estimates of classification accuracy. Psychol Assess 2019 Nov;31(11):1377-1382. [doi: 10.1037/pas0000764] [Medline: 31414853]
78. Denny JC, Choma NN, Peterson JF, Miller RA, Bastarache L, Li M, et al. Natural language processing improves identification of colorectal cancer testing in the electronic medical record. Med Decis Making 2012;32(1):188-197. [doi: 10.1177/0272989X11400418] [Medline: 21393557]
79. Hoogendoorn M, Szolovits P, Moons LM, Numans ME. Utilizing uncoded consultation notes from electronic medical records for predictive modeling of colorectal cancer. Artif Intell Med 2016 May;69:53-61 [FREE Full text] [doi: 10.1016/j.artmed.2016.03.003] [Medline: 27085847]
80. Kop R, Hoogendoorn M, Teije AT, Büchner FL, Slottje P, Moons LM, et al. Predictive modeling of colorectal cancer using a dedicated pre-processing pipeline on routine electronic medical records. Comput Biol Med 2016 Sep;76:30-38. [doi: 10.1016/j.compbiomed.2016.06.019] [Medline: 27392227]

81. Lau K, Wilkinson J, Moorthy R. A web-based prediction score for head and neck cancer referrals. Clin Otolaryngol 2018 Aug;43(4):1043-1049. [doi: 10.1111/coa.13098] [Medline: 29543399]
82. Haslam N, Beck AT. Categorization of major depression in an outpatient sample. J Nerv Ment Dis 1993 Dec;181(12):725-731. [doi: 10.1097/00005053-199312000-00003] [Medline: 8254323]
83. Lin Y, Huang S, Simon GE, Liu S. Data-based decision rules to personalize depression follow-up. Sci Rep 2018 Mar;8(1):5064-5068 [FREE Full text] [doi: 10.1038/s41598-018-23326-1] [Medline: 29567970]
84. Patel R, Jayatilleke N, Broadbent M, Chang C, Foskett N, Gorrell G, et al. Negative symptoms in schizophrenia: a study in a large clinical sample of patients using a novel automated method. BMJ Open 2015 Sep;5(9):e007619. [doi: 10.1136/bmjopen-2015-007619]
85. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit Med 2018 Aug;1(1):39-48 [FREE Full text] [doi: 10.1038/s41746-018-0040-6] [Medline: 31304320]
86. Kanagasingam Y, Xiao D, Vignarajan J, Preetham A, Tay-Kearney M, Mehrotra A. Evaluation of artificial intelligence-based grading of diabetic retinopathy in primary care. JAMA Netw Open 2018 Sep;1(5):e182665 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.2665] [Medline: 30646178]
87. Verbraak FD, Abramoff MD, Bausch GC, Klaver C, Nijpels G, Schlingemann RO, et al. Diagnostic accuracy of a device for the automated detection of diabetic retinopathy in a primary care setting. Diabetes Care 2019 Apr 14;42(4):651-656. [doi: 10.2337/dc18-0148] [Medline: 30765436]
88. Anderson HD, Pace WD, Brandt E, Nielsen RD, Allen RR, Libby AM, et al. Monitoring suicidal patients in primary care using electronic health records. J Am Board Fam Med 2015 Jan;28(1):65-71 [FREE Full text] [doi: 10.3122/jabfm.2015.01.140181] [Medline: 25567824]
89. Jordan P, Shedden-Mora MC, Löwe B. Predicting suicidal ideation in primary care: An approach to identify easily assessable key variables. Gen Hosp Psychiatry 2018 Mar;51:106-111. [doi: 10.1016/j.genhosppsych.2018.02.002] [Medline: 29428582]
90. Doukidis GI, Forster D. The potential for computer-aided diagnosis of tropical diseases in developing countries: an expert system case study. Eur J Oper Res 1990 Nov;49(2):271-278. [doi: 10.1016/0377-2217(90)90345-c]
91. Thakur S, Dharavath R. Artificial neural network based prediction of malaria abundances using big data: a knowledge capturing approach. Clin Epidemiol Glob Health 2019 Mar;7(1):121-126. [doi: 10.1016/j.cegh.2018.03.001]
92. Luo L, Small D, Stewart WF, Roy JA. Methods for estimating kidney disease stage transition probabilities using electronic medical records. EGEMS (Wash DC) 2013 Dec;1(3):1040 [FREE Full text] [doi: 10.13063/2327-9214.1040] [Medline: 25848580]
93. Xu G, Player P, Shepherd D, Brunskill NJ. Identifying acute kidney injury in the community--a novel informatics approach. J Nephrol 2016 Feb;29(1):93-98. [doi: 10.1007/s40620-015-0190-4] [Medline: 25779026]
94. Ben-Sasson A, Robins DL, Yom-Tov E. Risk assessment for parents who suspect their child has autism spectrum disorder: machine learning approach. J Med Internet Res 2018 Apr;20(4):e134 [FREE Full text] [doi: 10.2196/jmir.9496] [Medline: 29691210]
95. Thabtah F, Kamalov F, Rajab K. A new computational intelligence approach to detect autistic features for autism screening. Int J Med Inform 2018 Sep;117:112-124. [doi: 10.1016/j.ijmedinf.2018.06.009] [Medline: 30032959]
96. Janssen KJ, Siccama I, Vergouwe Y, Koffijberg H, Debray T, Keijzer M, et al. Development and validation of clinical prediction models: marginal differences between logistic regression, penalized maximum likelihood estimation, and genetic programming. J Clin Epidemiol 2012 Apr;65(4):404-412. [doi: 10.1016/j.jclinepi.2011.08.011] [Medline: 22214734]
97. Taylor R, Taylor A, Smyth J. Using an artificial neural network to predict healing times and risk factors for venous leg ulcers. J Wound Care 2002 Mar;11(3):101-105. [doi: 10.12968/jowc.2002.11.3.26381] [Medline: 11933726]
98. Chen Y, Lin C, Hong C, Lee D, Sun C, Lin H. Design of a clinical decision support system for predicting erectile dysfunction in men using NHIRD dataset. IEEE J Biomed Health Inform 2019 Sep;23(5):2127-2137. [doi: 10.1109/jbhi.2018.2877595]
99. Lin J, Bruni FM, Fu Z, Maloney J, Bardina L, Boner AL, et al. A bioinformatics approach to identify patients with symptomatic peanut allergy using peptide microarray immunoassay. J Allergy Clin Immunol 2012 May;129(5):1321-1328 [FREE Full text] [doi: 10.1016/j.jaci.2012.02.012] [Medline: 22444503]
100. Abdel-Aal R, Mangoud A. Modeling obesity using abductive networks. Comput Biomed Res 1997 Dec;30(6):451-471. [doi: 10.1006/cbmr.1997.1460] [Medline: 9466835]
101. Maizels M, Wolfe WJ. An expert system for headache diagnosis: the Computerized Headache Assessment tool (CHAT). Headache 2008 Jan;48(1):72-78. [doi: 10.1111/j.1526-4610.2007.00918.x] [Medline: 17868352]
102. Knab JH, Wallace MS, Wagner RL, Tsoukatos J, Weinger MB. The use of a computer-based decision support system facilitates primary care physicians' management of chronic pain. Anesth Analg 2001 Sep;93(3):712-720. [doi: 10.1097/00000539-200109000-00035] [Medline: 11524346]
103. Betts K, Kisely S, Alati R. Predicting common maternal postpartum complications: leveraging health administrative data and machine learning. BJOG 2019 May;126(6):702-709. [doi: 10.1111/1471-0528.15607] [Medline: 30628159]
104. Penny K, Smith G. The use of data-mining to identify indicators of health-related quality of life in patients with irritable bowel syndrome. J Clin Nurs 2012 Oct;21(19-20):2761-2771. [doi: 10.1111/j.1365-2702.2011.03897.x] [Medline: 22985319]

105. Tran T, Fang T, Pham V, Lin C, Wang PC, Lo MT. Development of an automatic diagnostic algorithm for pediatric otitis media. Otol Neurotol 2018 Sep;39(8):1060-1065. [doi: 10.1097/MAO.0000000000001897] [Medline: 30015751]
106. Alpaydin E. Introduction to Machine Learning. Cambridge, Massachusetts, United States: MIT Press; 2020.
107. Bishop C. Pattern Recognition and Machine Learning. Cambridge, United Kingdom: Springer; 2006.
108. Tariq A, Westbrook J, Byrne M, Robinson M, Baysari MT. Applying a human factors approach to improve usability of a decision support system in tele-nursing. Collegian 2017 Jun;24(3):227-236. [doi: 10.1016/j.colegn.2016.02.001]
109. Taati B, Zhao S, Ashraf AB, Asgarian A, Browne ME, Prkachin KM, et al. Algorithmic bias in clinical populations—evaluating and improving facial analysis technology in older adults with dementia. IEEE Access 2019;7:25527-25534. [doi: 10.1109/access.2019.2900022]
110. Cirillo D, Catuara-Solarz S, Morey C, Guney E, Subirats L, Mellino S, et al. Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. NPJ Digit Med 2020 Jun;3(1):81 [FREE Full text] [doi: 10.1038/s41746-020-0288-5] [Medline: 32529043]
111. Regitz-Zagrosek V. EMBO Rep 2012 Jun;13(7):596-603 [FREE Full text] [doi: 10.1038/embor.2012.87] [Medline: 22699937]
112. Open Data Watch. Bridging the gap: mapping gender data availability in Latin America and the Caribbean. Data2X. 2020. URL: https://data2x.org/resource-center/bridging-the-gap-mapping-gender-data-availability-in-latin-america-and-the-caribbean/ [accessed 2021-07-28]
113. Pham Q, Gamble A, Hearn J, Cafazzo JA. The need for ethnoracial equity in artificial intelligence for diabetes management: review and recommendations. J Med Internet Res 2021 Feb;23(2):e22320 [FREE Full text] [doi: 10.2196/22320] [Medline: 33565982]
114. Shneiderman B. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. International Journal of Human–Computer Interaction 2020 Mar 23;36(6):495-504. [doi: 10.1080/10447318.2020.1741118]
115. Xu W, Dainoff MJ, Ge L, Gao Z. From human-computer interaction to human-AI Interaction: new challenges and opportunities for enabling human-centered AI. ArXiv Preprint posted online on May 12, 2021. [FREE Full text]
116. Milne-Ives M, de Cock C, Lim E, Shehadeh M, de Pennington N, Mole G, et al. The effectiveness of artificial intelligence conversational agents in health care: systematic review. J Med Internet Res 2020 Oct;22(10):e20346 [FREE Full text] [doi: 10.2196/20346] [Medline: 33090118]
117. Wolff J, Pauling J, Keck A, Baumbach J. The economic impact of artificial intelligence in health care: systematic review. J Med Internet Res 2020 Feb;22(2):e16866 [FREE Full text] [doi: 10.2196/16866] [Medline: 32130134]
118. Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). European Commission. 2021. URL: https://www.europeansources.info/record/proposal-for-a-regulation-laying-down-harmonised-rules-on-artificial-intelligence-artificial-intelligence-act-and-amending-certain-union-legislative-acts/ [accessed 2021-07-28]

Abbreviations

AI: Artificial Intelligence
CBPHC: Community-Based Primary Health Care
CIHR: Canadian Institutes of Health Research
CINAHL: Cumulative Index to Nursing and Allied Health Literature
EPOC: Cochrane Effective Practice and Organisation of Care Review Group
FN: False negative
FP: False positive
JBI: Joanna Briggs Institute
MeSH: Medical Subject Headings
ML: Machine Learning
PICOS: Population, Intervention, Comparison, Outcomes, Setting and Study designs
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analysis-Scoping Reviews
PROBAST: Prediction Model Risk of Bias Assessment Tool
NLP: Natural Language Processing
OSI: Open Science Framework
TN: True negative
TP: True positive

Edited by G Eysenbach; submitted 23.04.21; peer-reviewed by R Hendricks-Sturrup; comments to author 17.05.21; revised version received 29.05.21; accepted 31.05.21; published 03.09.21.

Please cite as:
Abbasgholizadeh Rahimi S, Légaré F, Sharma G, Archambault P, Zomahoun HTV, Chandavong S, Rheault N, T Wong S, Langlois L, Couturier Y, Salmeron JL, Gagnon MP, Légaré J
Application of Artificial Intelligence in Community-Based Primary Health Care: Systematic Scoping Review and Critical Appraisal
J Med Internet Res 2021;23(9):e29839
URL: https://www.jmir.org/2021/9/e29839
doi:10.2196/29839
PMID:34477556

Original Paper

Machine Learning Analysis of Time-Dependent Features for Predicting Adverse Events During Hemodialysis Therapy: Model Development and Validation Study

Yi-Shiuan Liu[1,2,3,4], PhD; Chih-Yu Yang[1,2,5,6*], MD, PhD; Ping-Fang Chiu[7], MD; Hui-Chu Lin[2], MS; Chung-Chuan Lo[8], PhD; Alan Szu-Han Lai[2], BS; Chia-Chu Chang[9,10], MD, PhD; Oscar Kuang-Sheng Lee[1,2,11*], MD, PhD

[1]Institute of Clinical Medicine, National Yang Ming Chiao Tung University School of Medicine, Taipei, Taiwan
[2]Stem Cell Research Center, National Yang Ming Chiao Tung University, Taipei, Taiwan
[3]Department of Physiology and Pharmacology, Chang Gung University College of Medicine, Taoyuan, Taiwan
[4]Department of Plastic and Reconstructive Surgery, Chang Gung Memorial Hospital, Taoyuan, Taiwan
[5]Division of Nephrology, Department of Medicine, Taipei Veterans General Hospital, Taipei, Taiwan
[6]Center for Intelligent Drug Systems and Smart Bio-devices, Hsinchu, Taiwan
[7]Division of Nephrology, Department of Medicine, Changhua Christian Hospital, Changhua, Taiwan
[8]Institute of Systems Neuroscience, National Tsing Hua University, Hsinchu, Taiwan
[9]Department of Medicine, Kuang Tien General Hospital, Taichung, Taiwan
[10]Department of Nutrition, Hungkuang University, Taichung, Taiwan
[11]Department of Orthopedics, China Medical University Hospital, Taichung, Taiwan
[*]these authors contributed equally

Corresponding Author:
Oscar Kuang-Sheng Lee, MD, PhD
Institute of Clinical Medicine
National Yang Ming Chiao Tung University School of Medicine
2F, Shou-Ren Bldg, 155, Sec 2, Li-Nong St, Beitou Dist
Taipei, 11221
Taiwan
Phone: 886 228712121 ext 7391
Email: oscarlee9203@gmail.com

Abstract

Background: Hemodialysis (HD) therapy is an indispensable tool used in critical care management. Patients undergoing HD are at risk for intradialytic adverse events, ranging from muscle cramps to cardiac arrest. So far, there is no effective HD device–integrated algorithm to assist medical staff in response to these adverse events a step earlier during HD.

Objective: We aimed to develop machine learning algorithms to predict intradialytic adverse events in an unbiased manner.

Methods: Three-month dialysis and physiological time-series data were collected from all patients who underwent maintenance HD therapy at a tertiary care referral center. Dialysis data were collected automatically by HD devices, and physiological data were recorded by medical staff. Intradialytic adverse events were documented by medical staff according to patient complaints. Features extracted from the time series data sets by linear and differential analyses were used for machine learning to predict adverse events during HD.

Results: Time series dialysis data were collected during the 4-hour HD session in 108 patients who underwent maintenance HD therapy. There were a total of 4221 HD sessions, 406 of which involved at least one intradialytic adverse event. Models were built by classification algorithms and evaluated by four-fold cross-validation. The developed algorithm predicted overall intradialytic adverse events, with an area under the curve (AUC) of 0.83, sensitivity of 0.53, and specificity of 0.96. The algorithm also predicted muscle cramps, with an AUC of 0.85, and blood pressure elevation, with an AUC of 0.93. In addition, the model built based on ultrafiltration-unrelated features predicted all types of adverse events, with an AUC of 0.81, indicating that ultrafiltration-unrelated factors also contribute to the onset of adverse events.

Conclusions: Our results demonstrated that algorithms combining linear and differential analyses with two-class classification machine learning can predict intradialytic adverse events in quasi-real time with high AUCs. Such a methodology implemented

with local cloud computation and real-time optimization by personalized HD data could warn clinicians to take timely actions in advance.

(J Med Internet Res 2021;23(9):e27098) doi:10.2196/27098

KEYWORDS

hemodialysis; intradialytic adverse events; prediction algorithm; machine learning

Introduction

Hemodialysis (HD) therapy has a substantial role in critical care management [1]. Due to oliguria or even anuria, most patients with renal failure require fluid removal during HD therapy to maintain an euvolemic status. The volume-dependent component of hypertension may be corrected by fluid removal, but the ultrafiltration process exposes HD patients to the risks of hemodynamic instability, which may lead to fatal consequences such as cardiac arrest [2]. Intradialytic hypotension is the most frequent complication during HD [3-7] and has been identified as a pivotal cause of reduced HD efficacy [4,8]. Acutely, intradialytic adverse events can be fatal; chronically, frequent intradialytic adverse events increase patient morbidity and long-term all-cause mortality [3,9,10]. Therefore, it is urged to develop a solution for this unmet medical need.

The Crit-Line (Fresenius Medical Care) monitor is a device developed to assist with fluid removal during ultrafiltration by noninvasively monitoring real-time hematocrit, oxygen saturation, and intradialytic volume status, using an optical transmission method [11]. Although uncontrolled studies have suggested that this device reduced intradialytic symptoms [5,12] and assisted in the assessment of target weight [13,14], an unblinded randomized controlled trial showed a higher hospitalization rate in the Crit-Line group than in the control group [15]. Therefore, novel solutions are urged to solve this unmet medical need.

Artificial intelligence has been applied to HD patients to assist clinical practice, including prediction of urea clearance [16-19], dietary protein intake [17,20], volume status [21], erythropoiesis-stimulating agent response [22-26], iron supplement response [22,24], hemoglobin level [27], HD quality [28-31], mortality [32], etc. Although artificial intelligence has also been applied to predict intradialytic hypotension risks, previous studies lack time-series data input [17,33]. Therefore, we aimed to include time-series data in a machine learning model to predict intradialytic adverse events. Herein, we hypothesize that the machine learning method can predict intradialytic adverse events in an unbiased manner. To solve this unmet need, we conducted this study.

Methods

Study Protocol and Subjects

This was a retrospective observational study in a single institution. We reviewed the records of all patients who underwent maintenance HD therapy at Changhua Christian Hospital, a tertiary-care referral center in middle Taiwan, between August 2017 and October 2017. During this period, 129 patients were eligible for enrollment evaluation, and 108 patients completed the 3-month study. HD sessions were excluded for the following three reasons: (1) session interruption due to dialyzer exchange, (2) more than one interruption per session due to patient urination or defecation, and (3) inability of patients to freely express their discomfort during the session. Eventually, a total of 4221 HD sessions from 108 patients were used to build the model. Each patient received either 39 or 40 HD sessions during the 3-month study period.

The Institutional Review Board of our institution approved all protocols in April 2017 before the study began, and the protocols conformed to the ethical guidelines of the Helsinki declaration. The need for informed consent was waived because of the retrospective nature of the study.

Dialysis and Physiological Data Collection

Demographic information from medical records, including age, gender, and years under dialysis treatment, were included for model building. Dialysis and physiological data of the enrolled patients during the 4-hour HD session were included in the study. Physiological data were measured and recorded by medical staff every 30 to 60 minutes approximately. Dialysis data were collected from the dialysis machine automatically. Intradialytic adverse events were documented by medical staff according to physiological measurements or patient complaints, as shown in Table 1. The list of HD machine readouts is presented in Multimedia Appendix 1.

For each HD session i (i=1-4221), the data set HD_i consisted of records $\{Y_{j,k}, T_k\}$, where j (range 1-9) is the index for the dialysis and physiological measurements, and k is the index of time when a measurement is taking. $Y_{j,k}$ is the value of the measurement j at time T_k. HD_i also included additional time-invariant patient-specific information Y_j (j=10-13), including age, gender, years under dialysis treatment, and predialytic weight (Multimedia Appendix 2). According to the manufacturer default setting, the machine-derived dialysis data are recorded from the dialysis machine automatically once the value of venous pressure or transmembranous pressure alters and becomes different from the last measurement at $T=T_{k-1}$. Therefore, the time interval $T_k - T_{k-1}$ between any two consecutive records may not be equal.

Table 1. List of intradialytic adverse events.

Adverse event	Episodes, n
Muscle cramps	138
Blood pressure elevation	108
Low blood pressure	64
Miscellaneous	45
Headache	28
Lightheadedness	26
Chest tightness	23
Vascular access thrombosis	23
Cold sweating	22
Nausea/vomiting	12
Fever	10
Tachycardia	10
Dyspnea	8
Hoarseness	8
Chills	5
Leg pain	5
Low back pain	5
Shoulder pain	5
Altered mental status	4
Chest discomfort	3
Numb hands	3
Tinnitus	3
Vascular access occlusion	3
Abdominal pain	2
Hypersomnia	1
Palpitation	1
Pruritus	1

Feature Extraction

To avoid the artifacts at the beginning of the data due to the different procedures on how the dialysis was set up and started in each HD session, the first data point $Y_{j,1}$ at the beginning of each HD session was excluded if the blood flow rate varied between T_1 and T_2. We also excluded the data point $Y_{j,k}$ when the blood flow rate was equal to or below zero due to dialysis interruption (dialyzer exchange or patient urination/defecation). An entire HD session was excluded from the analysis if the session was interrupted more than once.

In our main analysis, the whole data set $\{Y_{j,k}, T_k\}$ of an HD session was included for feature extraction if no adverse event was registered for the session. On the other hand, for the HD session with adverse events, only data preceding the first adverse event were included for feature extraction, meaning the length of HD was less than 4 hours. Because the time interval between two adjacent records and the length of HD sessions vary, regression analysis is challenging, and we need to include the temporal features of the measured variables in the analyses for classification. To this end, we derived the mean, standard deviation of the mean, and coefficient of variance, as well as the slope and R square of linear regression from the dialysis and physiological measurements $\{Y_{j,k}, T_k\}$. We also derived the maximum, minimum, and mean of change rate (the first-order derivative), as well as the second-order derivative of venous pressure and transmembranous pressure as features for analysis. A total of 84 features $\{X_h\}$ (h=1-84), including those from the raw measurements $\{Y_{j,k}, T_k\}$ and those derived from the temporal aspect of the data as described above, were extracted for each HD session (Multimedia Appendix 3). Feature extraction of data sets was performed using the AWK program (source code available in the format of .awk; Multimedia Appendix 4).

As aforementioned, the dialysis data set $\{Y_{j,k}, T_k\}$ is recorded once the values of venous pressure or transmembranous pressure

change. Therefore, the value of any measurement at the time T_p between two measured time stamps, T_k and T_{k-1}, can be assigned as $\{Y_{j,k}, T_p\} = \{Y_{j,k}, T_{k-1}\}$. Thus, feature extraction of the data in an HD session could be terminated at an arbitrary time (T_p).

Outcome Labeling for Model Building

For outcome labeling, the HD sessions with one or more than one adverse event were labeled as 1, and HD sessions with no adverse event were labeled as 0. We also randomly relabeled 4221 HD sessions regardless of their true outcome as a negative control set while kept the same 0 to 1 ratio as the experimental set. A two-class classification model was built and evaluated by four-fold cross-validation using Azure (Microsoft Inc). At least three repeats were performed by introducing different random numbers for each model building.

Selection for Top Performance Features

To pinpoint which features are more important than others in predicting HD adverse events, we also selected and used key features for model building and compared the results with that by a total of 84 features. The selection of key features was performed using MATLAB (MATrixLABoratory, MathWorks Inc) (source code available in the format of .m; Multimedia Appendix 5). A two-class classification model was built using ensemble random undersampling boosted trees by four-fold cross-validation. The score was given by summing up the percentages of true positives and true negatives.

The key feature selection process started with selecting the top feature according to the scores obtained from the model using a single feature from the 84 features once at a time. Next, the top two-feature combinations were selected from the two-feature combination pool, which was established by combining the top feature from the last step with each of the remaining 83 features. The two-feature combinations that resulted in scores higher than that of the top feature from the last step were kept for the next step. Likewise, the top three-feature combinations could be selected from the pool established by combining the top two-feature combinations with each of the remaining 82 features when the three-feature combinations scored higher than the two-feature combinations. We repeated this procedure until the top 20-feature combinations were selected. Features that most frequently appeared in these 20-feature combinations were defined as key features.

Ethics Approval and Consent to Participate

This study was approved by the Institutional Review Board of National Yang-Ming University (N_105_0132) and the Institutional Review Board of Changhua Christian Hospital (CCH IRB No. 161005).

Results

Demographic Characteristics of the Study Participants

As of November 2017, we enrolled 108 patients. Multimedia Appendix 6 shows the baseline characteristics of the 108 patients at the beginning of the study. The mean age was 63.6 years, 60 (56%) patients were male, and the mean duration on dialysis was 7.7 years. A total of 47 (44%) patients had diabetes mellitus, 69 (64%) patients had hypertension, 11 (10%) patients had coronary artery disease, 12 (11%) patients had congestive heart failure, 7 (7%) patients had a history of stroke, and 2 (2%) patients had malignancy.

The list of intradialytic adverse events and the number of occurrences are shown in Table 1. Four HD sessions had more than three intradialytic adverse events, 19 HD sessions had three adverse events, 106 HD sessions had two adverse events, and 276 HD sessions had a single adverse event. Altogether, there were 406 HD sessions with adverse events out of 4221 total HD sessions (Multimedia Appendix 7).

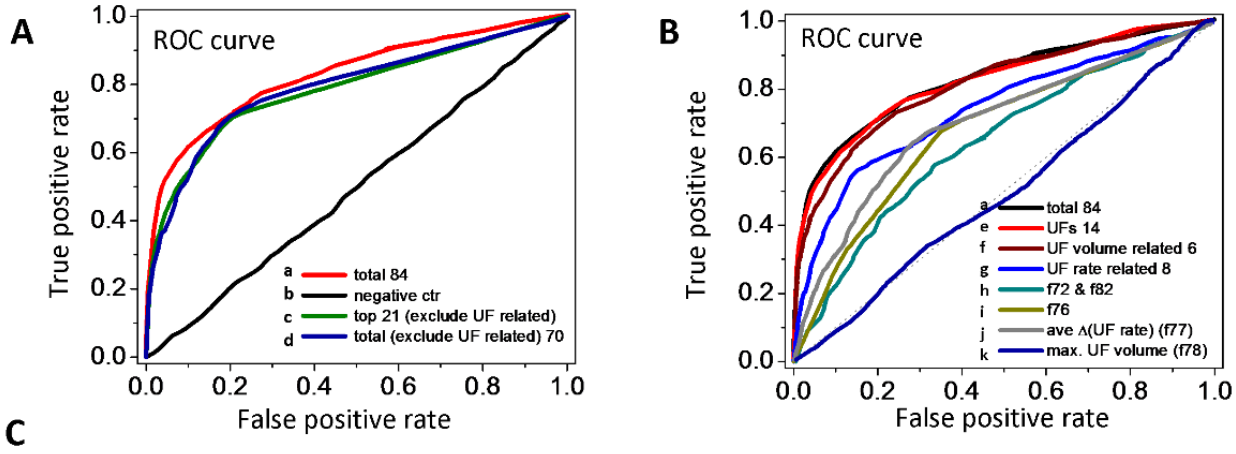
Performance of the Model for Prediction

To increase the outcome 1 to 0 ratios, wherein the session with an adverse event is labeled as 1 and the session without an adverse event is labeled as 0, we categorized the 27 adverse events listed in Table 1 into three groups. The first group was total events but excluded events of blood pressure elevation, vascular access occlusion, and vascular access thrombosis. A total of 323 HD sessions belonged to adverse event group 1. The second group included muscle cramps, and there were 138 HD sessions in this group. The third group included blood pressure elevation, and there were 108 HD sessions in this group.

Group 1: All Events Except Blood Pressure Elevation

A two-class averaged perceptron was used for model building with a learning rate of 20 and maximal iterations of 20. For the 84-feature model, the mean area under the curve (AUC) was 0.83 (SD 0.03), with an F1 score of 0.53, sensitivity of 0.53, and specificity of 0.96 (Figure 1A and 1B curve a). Compared with the negative control, (mean AUC=0.50, SD 0.04; F1=0.15; Figure 1A curve b), the 84-feature model of the two-class averaged perceptron could predict adverse events plausibly. Other algorithms were also tested for the prediction. The mean AUC obtained by two-class support vector machines (SVM) was 0.83 (SD 0.02), with an F1 score of 0.55, sensitivity of 0.53, and specificity of 0.96. The results were similar to those obtained by the averaged perceptron. Compared to averaged perceptron and SVM algorithms, two-class logistic regression and decision forest did not predict the adverse events well. The mean AUC obtained by logistic regression was 0.82 (SD 0.02), with an F1 score of 0.48, and the mean AUC obtained by decision forest was 0.83 (SD 0.02), with an F1 score of 0.46. Additionally, intrapatient partition and interpatient partition for sampling did not show significant difference in prediction (mean AUC=0.83, SD 0.03; mean F1=0.53, SD 0.02 vs mean AUC=0.82, SD 0.04; mean F1=0.50, SD 0.06).

Figure 1. Prediction of all intradialytic adverse events except blood pressure elevation. (A) Machine learning performance represented by receiver operating characteristic (ROC) curves from 84 features (curve a, red), the top 21 features (curve c, green), all features but excluding ultrafiltration-related features (curve d, blue), and the negative control (curve b, black). Each ROC curve shown here is the average of 12 simulated ROC curves. (B). Machine learning performance represented by ROC curves is obtained from ultrafiltration-related features (curves e, f, g, h, i, j, and k) and from 84 features (curve a, black). (C) Area under the curve (AUC) and F1 scores from different feature combinations for predicting all intradialytic adverse events. UF: ultrafiltration.



C

	a	b	c	d	e	f	g	h	i	j	k
	All features	All with out-come randomly labeled	Top features without UF related features	All without UF related features	UF related features	UF volume related features	UF rate related features	Two UF related features	Single UF related feature		
Feature included	1 - 84	1- 84	2,5,6,8,11,14,17,20 ,21,26,29,31,36,47-49, 50-52, 57,59	1-69, 84	70-83	78-83	70-77	72,82	76	77	78
Number of features used	84	84	21	70	14	6	8	2	1		
AUC	0.83	0.50	0.82	0.81	0.83	0.82	0.75	0.65	0.69	0.70	0.48
F1 score	0.53	0.15	0.45	0.45	0.52	0.46	0.36	0.22	0.25	0.28	0.15

Ultrafiltration rate and ultrafiltration volume are important parameters for HD. However, our models indicated that employing a single feature, such as the maximal value of ultrafiltration volume (feature 78) or the mean value of ultrafiltration rate changes (feature 77), cannot predict adverse events properly (Figure 1B curves j and k). The model built by the maximal value of ultrafiltration volume, defined as the ultrafiltration volume recorded at the last time point, had an AUC of 0.48 and F1 score of 0.15, which were similar to the results of the negative control. On the other hand, the model built by the mean value of ultrafiltration rate changes during HD sessions had an AUC of 0.70 and F1 score of 0.28. Combining two ultrafiltration-related features also failed to predict adverse events (Figure 1B curve h). After up to six ultrafiltration volume-related features (features 78-83) were used for prediction, the AUC increased from 0.48 to 0.82, and the F1 score increased from 0.15 to 0.46 (Figure 1B curve f). The model with 14 ultrafiltration features (features 70-83) had an AUC of 0.83 and F1 score of 0.52 (Figure 1B curve e).

Next, the 21 features that most frequently appeared in the 20-feature combinations were selected for the evaluation. The two-class averaged perceptron model based on these top 21 performance features but skipping ultrafiltration-related features showed a mean AUC of 0.82 (SD 0.02) and F1 score of 0.45 (Figure 1A curve c). The increase of one or two features did not enhance the prediction significantly (23 top features model: mean AUC=0.82, SD 0.02; F1=0.46). Compared with the model based on all features but excluding ultrafiltration-related features (AUC=0.81; F1=0.45; Figure 1A curve d; 70 features), the results of the 21 top features model (without ultrafiltration-related features) demonstrated that a quarter of the total 84 features was sufficient to predict adverse events.

The 21 features were age, maximum transmembranous pressure, minimum systolic blood pressure (SBP), minimum diastolic blood pressure (DBP), minimum pulse pressure, minimum blood flow rate, mean SBP, mean venous pressure, mean transmembranous pressure, slope of linear regression of SBP, slope of linear regression of DBP, slope of linear regression of pulse pressure, slope of linear regression of pulse rate, slope of linear regression of transmembranous pressure, standard deviation of the mean of blood flow rate, R-squared of linear regression of pulse pressure, and related parameters to the second-order derivative of venous pressure (features 2, 5, 6, 8, 11, 14, 17, 20, 21, 26, 29, 31, 36, 47-52, 57, and 59) (Multimedia Appendix 3).

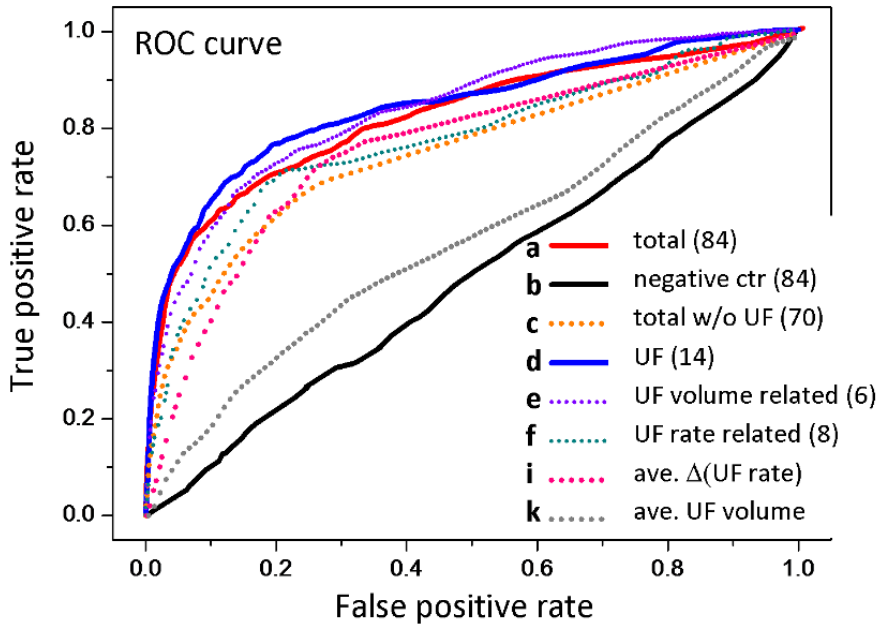
Group 2: Muscle Cramps

The model, which was based on 14 ultrafiltration-related features, had a mean AUC of 0.85 (SD 0.04) and F1 score of 0.45 (Figure 2 curve d) for predicting the occurrence of muscle cramps, and the result is similar to that of the 84-feature model (mean AUC=0.83, SD 0.04; F1=0.42; Figure 2 curve a) and

better than that of the model built based on all features but excluding ultrafiltration-related features (mean AUC=0.79, SD 0.04; F1=0.30; Figure 2 curve c). However, a single ultrafiltration-related feature cannot predict cramps properly (Figure 2 curves i and k). The combination of two ultrafiltration-related features also failed to predict muscle cramps (AUC=0.79 and F1=0.29 for features 70 and 77; AUC=0.84 and F1=0.37 for features 78 and 83). Our results demonstrated that ultrafiltration-related features contribute more than other features to the prediction of muscle cramps.

Figure 2. Prediction of a specific intradialytic adverse event: muscle cramps. (A) Machine learning performance is represented by receiver operating characteristic (ROC) curves from 84 features (curve a, red), all features but excluding ultrafiltration-related features (curve c, orange dot), ultrafiltration-related features (curves d, e, f, i, k), and the negative control (curve b, black). (B) Area under the curve (AUC) and F1 scores from different feature combinations for predicting muscle cramps. UF: ultrafiltration.

A



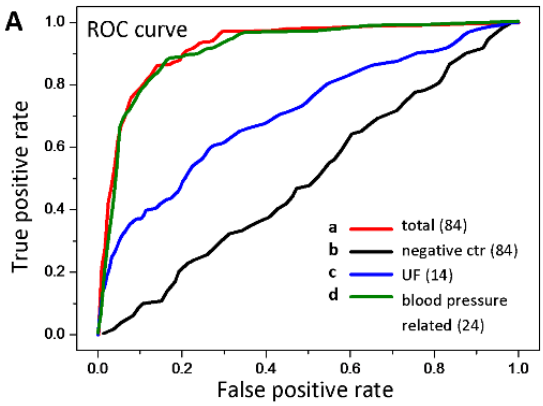
B

	a	b	c	d	e	f	i	j	k
	All features	All with out-come randomly labeled	All without UF related features	UF related features	UF volume related features	UF rate related features	Single UF related feature		
Feature included	1 - 84	1- 84	1-69, 84	70-83	78-83	70-77	77	82	80
Number of features used	84	84	70	14	6	8	1		
AUC	0.82	0.49	0.79	0.85	0.84	0.79	0.76	0.70	0.58
F1 score	0.42	0.07	0.30	0.45	0.37	0.29	0.22	0.16	0.08

Group 3: Blood Pressure Elevation

The model, which was based on a total of 84 features, had a mean AUC of 0.93 (SD 0.02) and F1 score of 0.41 for predicting the occurrence of hypertension (Figure 3 curve a). Compared with the model built based on 14 ultrafiltration-related features (AUC=0.72; F1=0.22; Figure 3 curve c), our results demonstrated that ultrafiltration parameters did not play important roles in predicting intradialytic hypertension. Even though the model based on 24 blood pressure–related features had an AUC higher than 0.9 (AUC=0.92, SD 0.03; F1=0.38; Figure 3 curve d), features other than blood pressure can further contribute to an additional improvement in the F1 score.

Figure 3. Prediction of a specific intradialytic adverse event: blood pressure elevation. (A) Machine learning performance represented by receiver operating characteristic (ROC) curves from 84 features (curve a, red), blood pressure-related features (curve d, green), ultrafiltration-related features (curve c, blue), and the negative control (curve b, black). (B) Area under the curve (AUC) and F1 scores from different feature combinations for predicting blood pressure elevation. UF: ultrafiltration.



B

	a	b	c	d
	All features	All with outcome randomly labeled	UF related features	Features related to blood pressure
Features included	1 - 84	1- 84	70-83	4-27
Number of features used	84	84	14	24
AUC	0.93	0.53	0.72	0.92
F1 score	0.41	0.08	0.22	0.38

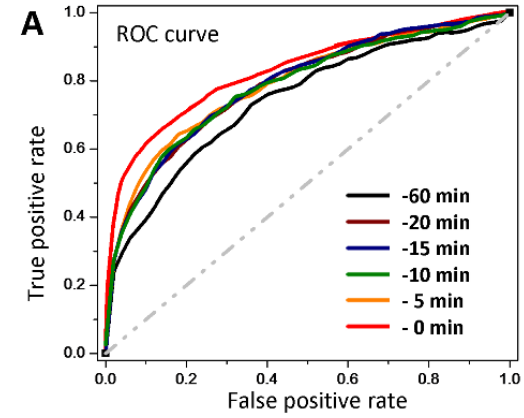
Consistency of the Predicted Probabilities of Adverse Events Over Time

As shown in Figure 1, the time series features were collected throughout the HD session, from the beginning of HD to the time point right before the index adverse event or right before the end of the HD session, which has no adverse event. In this case, we defined the end of feature collection as 0 minutes if this ending time point was right before the occurrence of the index adverse event. In addition to 0 minutes, we also set the cutoff ending time points of feature collection as 5, 10, 15, 20, and 60 minutes before the occurrence of the adverse event to evaluate the prediction accuracy. Machine learning for predicting all intradialytic adverse events, except blood pressure elevation, showed that features of the 0-minute cutoff led to the best AUC and F1 score (AUC=0.83; F1=0.53) compared to those learned from the features of earlier cutoff time points (Figure 4A and 4B), even though AUC scores from the features of the cutoff ending time points 5, 10, 15, and 20 minutes before the index adverse event or before the end of HD sessions without an adverse event were about 0.80 and their F1 scores were all lower than 0.5. The results suggest that while the information embedded in the 20-minute time window before the index adverse event is valuable, the information embedded in the 5-minute time window before the index adverse event is more influential for event prediction.

To further understand the cutoff ending time point dependence of prediction accuracy, 500 HD sessions were randomly selected to compare the prediction probabilities of adverse events obtained from 84 features with cutoff ending time points of 0, 5, 10, 15, and 20 minutes before the index adverse event. As shown in Figure 4C (red circles), five HD sessions possessed strong consistency in the predicted probabilities of adverse events using extracted features based on different cutoff ending time points, and adverse events indeed occurred in these five HD sessions. Since there should be approximately 40 HD sessions developing adverse events among 500 randomly selected HD sessions, the results suggest that at least one-tenth of HD sessions with adverse events can be sighted as early as 20 minutes in advance and can be further confirmed by real-time machine learning using features from subsequent cutoff ending time points (Figure 4C).

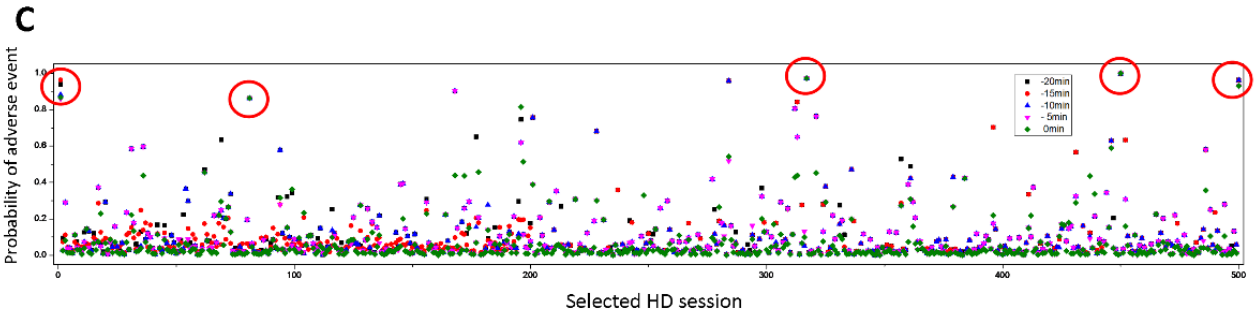
Even though none of the 84 features contained explicit time series information, the linear and differential analyses that feature extraction employed may be affected by the length of HD sessions. Therefore, we truncated the HD sessions with no adverse events (negative ones) and compared the prediction results with those from the untruncated ones. Since the average length of HD sessions with adverse events (positive ones) was 3.3 hours, negative HD sessions were truncated and randomly assigned endpoints (T_{end}) between 3 and 3.5 hours, yet the endpoints of positive ones remained unchanged. The data set $\{Y_{j,k}, T_k\}$ at endpoint T_{end} was defined according to the same method used for $\{Y_{j,k}, T_k\}$ at arbitrary time T_p. Regarding the results, the mean AUC was 0.89 (SD 0.019), F1 score was 0.55, sensitivity was 0.52, and specificity was 0.97. Alternately, the AUC was 0.86 with an F1 score of 0.55 when the endpoints were assigned exactly at 3.3 hours. Compared to the original results obtained from the untruncated negative HD sessions with a duration of about 4 hours (AUC=0.83, F1=0.53, sensitivity=0.53, and specificity=0.96), the prediction results were better when the endpoints were set earlier. Indeed, the AUC was 0.92, with an F1 score of 0.62, sensitivity of 0.61, and specificity of 0.98, when the endpoints were randomly assigned between 2.5 and 3.5 hours for negative HD sessions.

Figure 4. Prediction performance for group 1 intradialytic adverse events using the features of different cutoff ending time points. (A) Machine learning performance represented by receiver operating characteristic (ROC) curves from 84 features extracted from [Yj, Tk]HD terminated at different cutoff ending time points as follows: one time point before an adverse event (noted as 0 minutes), and 5, 10, 15, 20, and 60 minutes before an adverse event or before the end of the hemodialysis (HD) session if no adverse event. (B) Area under the curve (AUC) and F1 scores. (C) Probability of adverse event occurrence in 500 randomly selected HD sessions. The red circle indicates the HD session with adverse events, and the predicted probabilities of adverse events are all higher than 0.8 independent of the cutoff ending time point.



B

Time before the event	True positive	False positive	False negative	True negative	AUC	F1
0 min	168	131	155	3684	0.83	0.53
5 min	149	221	174	3594	0.80	0.43
10 min	135	207	188	3608	0.79	0.41
15 min	124	143	199	3672	0.80	0.42
20 min	134	179	189	3636	0.79	0.42
60 min	102	126	221	3689	0.75	0.37



Discussion

Contributions and Principal Findings

Our findings indicate that algorithms combining linear and differential analyses with two-class classification machine learning predict intradialytic adverse events with high AUCs. We attempted to identify features that contribute the most to predicting all adverse events, except hypertension, (group 1) from a total of 84 features extracted from $[Y_j, T_k]_{HD}$. Among the top 23 features, only feature 76 and feature 82 were related to ultrafiltration (the number of times that the ultrafiltration rate changes and the linear regression slope of ultrafiltration volume). After excluding these two ultrafiltration-related features, we found that the remaining 21 features were sufficient for accurate prediction with good discriminating power, with a slight reduction in the AUC from 0.83 (84 features) to 0.82 (21 features). The model built by 14 ultrafiltration-related features also had a good AUC of 0.83. Therefore, instead of including all 84 features for model building, selecting the top 21 ultrafiltration-unrelated features or integrating a total of 14 ultrafiltration-related features can reduce computing load. Our results also suggest that these two clusters of features (Figure 1A curve c and Figure 1B curve e) may embed similar factors contributing to the onset of adverse events.

In our study, muscle cramp was an adverse event that occurred most frequently during HD treatment. A muscle cramp is a common adverse event that happens during HD therapy, with a prevalence of 28% among all HD sessions [5,34]. Muscle cramps result from ischemia of the skeletal muscle tissue, indicating an early sign of hypotension, and it may lead to premature discontinuation of HD sessions [35,36]. Tissue ischemia during HD is positively related to the ultrafiltration rate [5,37,38]. Indeed, the model built by 14 ultrafiltration-related features had an AUC of 0.85 for predicting the occurrence of muscle cramps in this study. We excluded all ultrafiltration-related features to test the prediction accuracy, including ultrafiltration rate and ultrafiltration volume, and the AUC reduced from 0.82 (84 features) to 0.79 (70 features), indicating that ultrafiltration-related features are important but not necessarily required to predict muscle cramps. The results from machine learning revealed that ultrafiltration-independent features contribute to predicting intradialytic muscle cramps as well.

In general, symptomatic hypotension occurs in 20% to 30% of dialysis sessions [6,39-43]. There are two major pathophysiological mechanisms of intradialytic hypotension. First, when plasma fluid removal through ultrafiltration exceeds the rate of plasma refilling into the blood vessels, blood volume reduces [5]. In the meantime, if the cardiovascular and neurohormone systems fail to compensate for the acute vascular volume depletion during ultrafiltration, hypotension occurs [3,44-47]. Frequent episodes of intradialytic hypotension may cause reduced ultrafiltration, inadequate “dry weight,” increased preload, and impaired heart function that eventually leads to more episodes of hypotension, thus creating a vicious cycle. Meanwhile, frequent intradialytic hypotension disrupts dialysis efficiency and efficacy. It is associated with higher morbidity

and mortality [48-51], which partly contributes to the fact that cardiovascular disease is the leading cause of morbidity and mortality in HD patients [52]. Lin et al recently developed an intelligent system to predict intradialytic hypotension [53]. However, our prediction models not only further precluded ultrafiltration-related features but also examined overall intradialytic adverse events instead of only focusing on the hypotensive episode.

As shown in Multimedia Appendix 8, the top 16 features that majorly contributed to predicting muscle cramps included patient characteristics, venous pressure, transmembranous pressure, ultrafiltration, blood flow rate, and pulse pressure. The minimal value of venous pressure and the mean value of transmembranous pressure were features that had the most hits (20 and 17 hits, respectively). These top two features for predicting muscle cramps were derived from dialysis machine output parameters, indicating that there is a potential to integrate our algorithm into the dialysis machine software to alert clinicians and to adjust dialysis machine settings in advance. Nevertheless, unlike prediction of all adverse events, for which only two out of the top 23 features were related to ultrafiltration, eight out of the top 16 features were related to ultrafiltration in terms of predicting muscle cramps, indicating that ultrafiltration-related parameters are important factors of muscle cramps.

Compared with several two-class classification modules, such as Bayes point machine, boosted decision tree, and SVM, models built by two-class average perceptron had the best AUC and F1 score. We also built models by deep learning (data not shown), but the results from deep learning did not show a good AUC and F1 score, possibly due to the limited number of our HD data sets. As clinicians are now facing the new era of artificial intelligence [54], the integration of computer science and dialysis medicine could be regarded as the first step to improve HD patients' care quality comprehensively. Our study demonstrated the feasibility of this integration. Even though the limited number of data sets and imbalanced data outcomes in our research hinder better prediction accuracy, it is anticipated that increased data sets will further improve the AUC and F1 score. Moreover, integrating machine learning with the dialysis machine and modifying algorithms in real-time by cloud computing with accumulation of data sets could enhance prediction performance.

Several questions may be answered if the size of the HD data set is expanded in future studies. First, how early can we predict adverse events? The consistency in the predicted probabilities of adverse events using features based on different cutoff ending time points could detect about one-tenth of HD sessions with adverse events (Figure 4). We anticipate that an increase in the number of HD sessions with adverse events for model training can improve imbalanced data and possibly bring forward the timing for the alert. Second, since most of the adverse events took place in second-half HD sessions, whether the data sets of second-half HD sessions are sufficient for prediction can be further studied. Finally, if more HD sessions with adverse events are recruited, we can build models for different adverse events instead of grouping the events to reduce imbalanced data outcomes.

Conclusion

In this study, a model of two-class classification was established to predict intradialytic adverse events in quasi-real time, with AUCs higher than 0.8. The consistency in the predicted probabilities of adverse events obtained from the features extracted in the ongoing HD process in real time could have the HD session tagged for forthcoming adverse events. Such a methodology implemented with local cloud computation could warn clinicians to take necessary actions and adjust the HD machine settings in advance.

Acknowledgments

The authors acknowledge financial support from Advantech Foundation (YM105C041). This study was also supported for research purposes by the Ministry of Science and Technology (MOST), Taiwan (MOST 108-2923-B-010-002-MY3, MOST 109-2926-I-010-502, MOST 109-2823-8-010-003-CV, and MOST 110-2923-B-A49A-501-MY3 [OKL]; 109-2314-B-010-053-MY3, 109-2321-B-009-007, MOST 109-2811-B-010-532, MOST 110-2811-B-010-510, MOST 110-2813-C-A49A-551-B, and MOST 110-2321-B-A49-003 [CY]; MOST 108-2633-B-009-001 [CY and OKL]; and MOST 109-2314-B-182-010 [YL]), grants from Taipei Veterans General Hospital, Taipei, Taiwan (V106D25-003-MY3, VGHUST107-G5-3-3, VGHUST109-V5-1-2, and V110C-194) (CY), the "Yin Yen-Liang Foundation Development and Construction Plan" of the School of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan (107F-M01-0504) (CY and OKL), and the Chang Gung Medical Research Foundation (CMRPD1K0601) (YL). Moreover, this work was in part supported for research purposes by the "Center for Intelligent Drug Systems and Smart Bio-devices (IDS^{2}B)" from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan (CY). The funders had no role in study design, data collection, analysis, interpretation, or writing of the manuscript.

Authors' Contributions

CY and OKL contributed equally as Corresponding Authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of hemodialysis machine readouts.
[DOC File , 57 KB - jmir_v23i9e27098_app1.doc]

Multimedia Appendix 2
An example of physiological and dialysis data collected in one hemodialysis (HD) session.
[PNG File , 147 KB - jmir_v23i9e27098_app2.png]

Multimedia Appendix 3
List of 84 features.
[PNG File , 361 KB - jmir_v23i9e27098_app3.png]

Multimedia Appendix 4
Source code: Feature extraction.
[ZIP File (Zip Archive), 4 KB - jmir_v23i9e27098_app4.zip]

Multimedia Appendix 5
Source code: Top performance feature selection.
[ZIP File (Zip Archive), 3 KB - jmir_v23i9e27098_app5.zip]

Multimedia Appendix 6
Demographic characteristics of the study participants (n=108).
[DOC File , 57 KB - jmir_v23i9e27098_app6.doc]

Multimedia Appendix 7
Simple decision tree of hemodialysis (HD) patients and HD sessions used for the study. One hundred twenty-nine patients were eligible after the enrollment evaluation, and 108 patients completed the 3-month study. There were 4221 HD sessions used for model building, and of these, 3815 did not have adverse events and 406 had one or multiple adverse events.
[PNG File , 60 KB - jmir_v23i9e27098_app7.png]

Multimedia Appendix 8
Top 16 features for predicting intradialytic muscle cramps.
[DOC File , 62 KB - jmir_v23i9e27098_app8.doc]

References

1. Ricci Z, Romagnoli S. Renal replacement therapy for critically ill patients: an intermittent continuity. Crit Care 2014 Mar 05;18(2):115 [FREE Full text] [doi: 10.1186/cc13756] [Medline: 24670363]
2. Douvris A, Malhi G, Hiremath S, McIntyre L, Silver SA, Bagshaw SM, et al. Interventions to prevent hemodynamic instability during renal replacement therapy in critically ill patients: a systematic review. Crit Care 2018 Feb 22;22(1):41 [FREE Full text] [doi: 10.1186/s13054-018-1965-5] [Medline: 29467008]
3. Daugirdas JT. Pathophysiology of dialysis hypotension: an update. Am J Kidney Dis 2001 Oct;38(4 Suppl 4):S11-S17. [doi: 10.1053/ajkd.2001.28090] [Medline: 11602456]
4. Garzoni D, Keusch G, Kleinoeder T, Martin H, Dhondt A, Cremaschi L, et al. Reduced complications during hemodialysis by automatic blood volume controlled ultrafiltration. Int J Artif Organs 2007 Jan;30(1):16-24. [doi: 10.1177/039139880703000104] [Medline: 17295189]
5. Steuer RR, Leypoldt JK, Cheung AK, Harris DH, Conis JM. Hematocrit as an indicator of blood volume and a predictor of intradialytic morbid events. ASAIO J 1994;40(3):M691-M696. [doi: 10.1097/00002480-199407000-00087] [Medline: 8555603]
6. Daugirdas JT. Preventing and Managing Hypotension. Seminars in Dialysis 1994;7(4):276-283 [FREE Full text] [doi: 10.1111/j.1525-139X.1994.tb00786.x]
7. Santoro A, Mancini E, Basile C, Amoroso L, Di Giulio S, Usberti M, et al. Blood volume controlled hemodialysis in hypotension-prone patients: a randomized, multicenter controlled trial. Kidney Int 2002 Sep;62(3):1034-1045 [FREE Full text] [doi: 10.1046/j.1523-1755.2002.00511.x] [Medline: 12164888]
8. Donauer J, Kölblin D, Bek M, Krause A, Böhler J. Ultrafiltration profiling and measurement of relative blood volume as strategies to reduce hemodialysis-related side effects. Am J Kidney Dis 2000 Jul;36(1):115-123. [doi: 10.1053/ajkd.2000.8280] [Medline: 10873880]

9. Shoji T, Tsubakihara Y, Fujii M, Imai E. Hemodialysis-associated hypotension as an independent risk factor for two-year mortality in hemodialysis patients. Kidney Int 2004 Sep;66(3):1212-1220 [FREE Full text] [doi: 10.1111/j.1523-1755.2004.00812.x] [Medline: 15327420]
10. Stefánsson BV, Brunelli SM, Cabrera C, Rosenbaum D, Anum E, Ramakrishnan K, et al. Intradialytic hypotension and risk of cardiovascular disease. Clin J Am Soc Nephrol 2014 Dec 05;9(12):2124-2132 [FREE Full text] [doi: 10.2215/CJN.02680314] [Medline: 25376764]
11. Steuer RR, Harris DH, Weiss RL, Biddulph MC, Conis JM. Evaluation of a noninvasive hematocrit monitor: a new technology. Am Clin Lab 1991 Jul;10(6):20-22. [Medline: 10148237]
12. Steuer RR, Leypoldt JK, Cheung AK, Senekjian HO, Conis JM. Reducing symptoms during hemodialysis by continuously monitoring the hematocrit. American Journal of Kidney Diseases 1996 Apr;27(4):525-532. [doi: 10.1016/s0272-6386(96)90163-8] [Medline: 8678063]
13. Sinha AD, Light RP, Agarwal R. Relative plasma volume monitoring during hemodialysis AIDS the assessment of dry weight. Hypertension 2010 Feb;55(2):305-311 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.109.143974] [Medline: 20038754]
14. Rodriguez HJ, Domenici R, Diroll A, Goykhman I. Assessment of dry weight by monitoring changes in blood volume during hemodialysis using Crit-Line. Kidney Int 2005 Aug;68(2):854-861 [FREE Full text] [doi: 10.1111/j.1523-1755.2005.00467.x] [Medline: 16014066]
15. Reddan DN, Szczech LA, Hasselblad V, Lowrie EG, Lindsay RM, Himmelfarb J, et al. Intradialytic blood volume monitoring in ambulatory hemodialysis patients: a randomized trial. J Am Soc Nephrol 2005 Jul;16(7):2162-2169 [FREE Full text] [doi: 10.1681/ASN.2004121053] [Medline: 15930095]
16. Akl AI, Sobh MA, Enab YM, Tattersall J. Artificial intelligence: a new approach for prescription and monitoring of hemodialysis therapy. Am J Kidney Dis 2001 Dec;38(6):1277-1283. [doi: 10.1053/ajkd.2001.29225] [Medline: 11728961]
17. Gabutti L, Vadilonga D, Mombelli G, Burnier M, Marone C. Artificial neural networks improve the prediction of Kt/V, follow-up dietary protein intake and hypotension risk in haemodialysis patients. Nephrol Dial Transplant 2004 May;19(5):1204-1211. [doi: 10.1093/ndt/gfh084] [Medline: 14993478]
18. Goldfarb-Rumyantzev A, Schwenk MH, Liu S, Charytan C, Spinowitz BS. Prediction of single-pool Kt/v based on clinical and hemodialysis variables using multilinear regression, tree-based modeling, and artificial neural networks. Artif Organs 2003 Jun;27(6):544-554. [doi: 10.1046/j.1525-1594.2003.07001.x] [Medline: 12780509]
19. Guh JY, Yang CY, Yang JM, Chen LM, Lai YH. Prediction of equilibrated postdialysis BUN by an artificial neural network in high-efficiency hemodialysis. Am J Kidney Dis 1998 Apr;31(4):638-646. [doi: 10.1053/ajkd.1998.v31.pm9531180] [Medline: 9531180]
20. Gabutti L, Burnier M, Mombelli G, Malé F, Pellegrini L, Marone C. Usefulness of artificial neural networks to predict follow-up dietary protein intake in hemodialysis patients. Kidney Int 2004 Jul;66(1):399-407 [FREE Full text] [doi: 10.1111/j.1523-1755.2004.00744.x] [Medline: 15200449]
21. Chiu J, Chong C, Lin Y, Wu C, Wang Y, Li Y. Applying an artificial neural network to predict total body water in hemodialysis patients. Am J Nephrol 2005;25(5):507-513. [doi: 10.1159/000088279] [Medline: 16155360]
22. Barbieri C, Bolzoni E, Mari F, Cattinelli I, Bellocchio F, Martin JD, et al. Performance of a Predictive Model for Long-Term Hemoglobin Response to Darbepoetin and Iron Administration in a Large Cohort of Hemodialysis Patients. PLoS One 2016;11(3):e0148938 [FREE Full text] [doi: 10.1371/journal.pone.0148938] [Medline: 26939055]
23. Barbieri C, Molina M, Ponce P, Tothova M, Cattinelli I, Ion Titapiccolo J, et al. An international observational study suggests that artificial intelligence for clinical decision support optimizes anemia management in hemodialysis patients. Kidney Int 2016 Aug;90(2):422-429 [FREE Full text] [doi: 10.1016/j.kint.2016.03.036] [Medline: 27262365]
24. Barbieri C, Mari F, Stopper A, Gatti E, Escandell-Montero P, Martínez-Martínez JM, et al. A new machine learning approach for predicting the response to anemia treatment in a large cohort of End Stage Renal Disease patients undergoing dialysis. Comput Biol Med 2015 Jun;61:56-61. [doi: 10.1016/j.compbiomed.2015.03.019] [Medline: 25864164]
25. Escandell-Montero P, Chermisi M, Martínez-Martínez JM, Gómez-Sanchis J, Barbieri C, Soria-Olivas E, et al. Optimization of anemia treatment in hemodialysis patients via reinforcement learning. Artif Intell Med 2014 Sep;62(1):47-60. [doi: 10.1016/j.artmed.2014.07.004] [Medline: 25091172]
26. Gabutti L, Lötscher N, Bianda J, Marone C, Mombelli G, Burnier M. Would artificial neural networks implemented in clinical wards help nephrologists in predicting epoetin responsiveness? BMC Nephrol 2006 Sep 18;7:13 [FREE Full text] [doi: 10.1186/1471-2369-7-13] [Medline: 16981983]
27. Martínez-Martínez JM, Escandell-Montero P, Barbieri C, Soria-Olivas E, Mari F, Martínez-Sober M, et al. Prediction of the hemoglobin level in hemodialysis patients using machine learning techniques. Comput Methods Programs Biomed 2014 Nov;117(2):208-217. [doi: 10.1016/j.cmpb.2014.07.001] [Medline: 25070755]
28. Saadat S, Aziz A, Ahmad H, Imtiaz H, Sohail ZS, Kazmi A, et al. Predicting Quality of Life Changes in Hemodialysis Patients Using Machine Learning: Generation of an Early Warning System. Cureus 2017 Sep 25;9(9):e1713 [FREE Full text] [doi: 10.7759/cureus.1713] [Medline: 29188157]

29. Cattinelli I, Bolzoni E, Chermisi M, Bellocchio F, Barbieri C, Mari F, et al. Computational intelligence for the Balanced Scorecard: studying performance trends of hemodialysis clinics. Artif Intell Med 2013 Jul;58(3):165-173. [doi: 10.1016/j.artmed.2013.04.005] [Medline: 23768423]
30. Bellazzi R, Larizza C, Magni P, Bellazzi R. Temporal data mining for the quality assessment of hemodialysis services. Artif Intell Med 2005 May;34(1):25-39. [doi: 10.1016/j.artmed.2004.07.010] [Medline: 15885564]
31. Guo Y, Hao Z, Zhao S, Gong J, Yang F. Artificial Intelligence in Health Care: Bibliometric Analysis. J Med Internet Res 2020 Jul 29;22(7):e18228 [FREE Full text] [doi: 10.2196/18228] [Medline: 32723713]
32. Sheng K, Zhang P, Yao X, Li J, He Y, Chen J. Prognostic Machine Learning Models for First-Year Mortality in Incident Hemodialysis Patients: Development and Validation Study. JMIR Med Inform 2020 Oct 29;8(10):e20578 [FREE Full text] [doi: 10.2196/20578] [Medline: 33118948]
33. Gabutti L, Machacek M, Marone C, Ferrari P. Predicting intradialytic hypotension from experience, statistical models and artificial neural networks. J Nephrol 2005;18(4):409-416. [Medline: 16245245]
34. Sherman RA, Goodling KA, Eisinger RP. Acute Therapy of Hemodialysis-Related Muscle Cramps. American Journal of Kidney Diseases 1982 Sep;2(2):287-288. [doi: 10.1016/s0272-6386(82)80076-0] [Medline: 7124726]
35. Jenkins PG, Dreher WH. Dialysis-induced muscle cramps: treatment with hypertonic saline and theory as to etiology. Trans Am Soc Artif Intern Organs 1975;21:479-482. [Medline: 1146021]
36. Rocco MV, Burkart JM. Prevalence of missed treatments and early sign-offs in hemodialysis patients. J Am Soc Nephrol 1993 Nov;4(5):1178-1183 [FREE Full text] [doi: 10.1681/ASN.V451178] [Medline: 8305644]
37. de Vries PM, Kouw PM, Olthof CG, Solf A, Schuenemann B, Oe LP, et al. The influence of dialysate sodium and variable ultrafiltration on fluid balance during hemodialysis. ASAIO Trans 1990;36(4):821-824. [doi: 10.1097/00002480-199010000-00008] [Medline: 2268486]
38. Bogaard HJ, de Vries JP, de Vries PM. Assessment of refill and hypovolaemia by continuous surveillance of blood volume and extracellular fluid volume. Nephrol Dial Transplant 1994;9(9):1283-1287. [Medline: 7816290]
39. Zucchelli P, Santoro A. Dialysis-induced hypotension: a fresh look at pathophysiology. Blood Purif 1993;11(2):85-98. [doi: 10.1159/000170102] [Medline: 8274245]
40. Orofino L, Marcén R, Quereda C, Villafruela JJ, Sabater J, Matesanz R, et al. Epidemiology of symptomatic hypotension in hemodialysis: is cool dialysate beneficial for all patients? Am J Nephrol 1990;10(3):177-180. [doi: 10.1159/000168077] [Medline: 2382678]
41. Degoulet P, Réach I, Di Giulio S, Devriès C, Rouby JJ, Aimé F, et al. Epidemiology of dialysis induced hypotension. Proc Eur Dial Transplant Assoc 1981;18:133-138. [Medline: 7329960]
42. Chen I, Chang M, Chiao S, Chen J, Yu C, Yang S, et al. Korean red ginseng improves blood pressure stability in patients with intradialytic hypotension. Evid Based Complement Alternat Med 2012;2012:595271 [FREE Full text] [doi: 10.1155/2012/595271] [Medline: 22645630]
43. Palmer BF, Henrich WL. Recent advances in the prevention and management of intradialytic hypotension. J Am Soc Nephrol 2008 Jan;19(1):8-11 [FREE Full text] [doi: 10.1681/ASN.2007091006] [Medline: 18178796]
44. Yamamoto K, Kobayashi N, Kutsuna T, Ishii A, Matsumoto T, Hara M, et al. Excessive fall of blood pressure during maintenance hemodialysis in patients with chronic renal failure is induced by vascular malfunction and imbalance of autonomic nervous activity. Ther Apher Dial 2012 Jun;16(3):219-225. [doi: 10.1111/j.1744-9987.2011.01053.x] [Medline: 22607564]
45. Perazella MA, Rho M, Brewster UC. Vasopressin insufficiency and intradialytic hypotension. Am J Kidney Dis 2008 Dec;52(6):1197. [doi: 10.1053/j.ajkd.2008.07.045] [Medline: 19026359]
46. Coen G, Mantella D, Sardella D, Beraldi MP, Ferrari I, Pierantozzi A, et al. Asymmetric dimethylarginine, vascular calcifications and parathyroid hormone serum levels in hemodialysis patients. J Nephrol 2009;22(5):616-622. [Medline: 19809994]
47. Rashid G, Bernheim J, Green J, Benchetrit S. Parathyroid hormone stimulates endothelial expression of atherosclerotic parameters through protein kinase pathways. Am J Physiol Renal Physiol 2007 Apr;292(4):F1215-F1218 [FREE Full text] [doi: 10.1152/ajprenal.00406.2006] [Medline: 17190908]
48. Henderson LW. Symptomatic intradialytic hypotension and mortality: an opinionated review. Semin Dial 2012 May;25(3):320-325. [doi: 10.1111/j.1525-139X.2012.01068.x] [Medline: 22452289]
49. Inrig JK, Oddone EZ, Hasselblad V, Gillespie B, Patel UD, Reddan D, et al. Association of intradialytic blood pressure changes with hospitalization and mortality rates in prevalent ESRD patients. Kidney Int 2007 Mar;71(5):454-461 [FREE Full text] [doi: 10.1038/sj.ki.5002077] [Medline: 17213873]
50. Tislér A, Akócsi K, Borbás B, Fazakas L, Ferenczi S, Görögh S, et al. The effect of frequent or occasional dialysis-associated hypotension on survival of patients on maintenance haemodialysis. Nephrol Dial Transplant 2003 Dec;18(12):2601-2605. [doi: 10.1093/ndt/gfg450] [Medline: 14605284]
51. Chou JA, Streja E, Nguyen DV, Rhee CM, Obi Y, Inrig JK, et al. Intradialytic hypotension, blood pressure changes and mortality risk in incident hemodialysis patients. Nephrol Dial Transplant 2018 Jan 01;33(1):149-159 [FREE Full text] [doi: 10.1093/ndt/gfx037] [Medline: 28444336]

52. No authors listed. Causes of death. USRDS. United States Renal Data System. Am J Kidney Dis 1997 Aug;30(2 Suppl 1):S107-S117. [Medline: 9259696]
53. Lin C, Chen C, Wu P, Pan C, Shih H, Huang M, et al. Intelligent system to predict intradialytic hypotension in chronic hemodialysis. J Formos Med Assoc 2018 Oct;117(10):888-893 [FREE Full text] [doi: 10.1016/j.jfma.2018.05.023] [Medline: 29941330]
54. Lemley KV. Machine Learning Comes to Nephrology. JASN 2019 Sep 05;30(10):1780-1781. [doi: 10.1681/asn.2019070664]

Abbreviations

AUC: area under the curve
DBP: diastolic blood pressure
HD: hemodialysis
SBP: systolic blood pressure
SVM: support vector machines

Edited by R Kukafka; submitted 11.01.21; peer-reviewed by D Schwarz, Y Cai; comments to author 20.04.21; revised version received 10.07.21; accepted 12.08.21; published 07.09.21.

Please cite as:
Liu YS, Yang CY, Chiu PF, Lin HC, Lo CC, Lai ASH, Chang CC, Lee OKS
Machine Learning Analysis of Time-Dependent Features for Predicting Adverse Events During Hemodialysis Therapy: Model Development and Validation Study
J Med Internet Res 2021;23(9):e27098
URL: https://www.jmir.org/2021/9/e27098
doi:10.2196/27098
PMID:34491204

Original Paper

Predicting the Mortality and Readmission of In-Hospital Cardiac Arrest Patients With Electronic Health Records: A Machine Learning Approach

Chien-Yu Chi[1*], MD; Shuang Ao[2*], PhD; Adrian Winkler[2], BEng; Kuan-Chun Fu[2], MSc; Jie Xu[2], PhD; Yi-Lwun Ho[3], MD; Chien-Hua Huang[4], MD; Rohollah Soltani[2], MSc

[1]Department of Emergency Medicine, Yunlin Branch, National Taiwan University Hospital, Yunlin, Taiwan

[2]Knowtions Research, Toronto, ON, Canada

[3]Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan

[4]Department of Emergency Medicine, National Taiwan University Hospital, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Chien-Hua Huang, MD
Department of Emergency Medicine
National Taiwan University Hospital
#7 Chung-Shan South Road
Taipei, 100
Taiwan
Phone: 886 0972651304
Email: chhuang5940@ntu.edu.tw

Abstract

Background: In-hospital cardiac arrest (IHCA) is associated with high mortality and health care costs in the recovery phase. Predicting adverse outcome events, including readmission, improves the chance for appropriate interventions and reduces health care costs. However, studies related to the early prediction of adverse events of IHCA survivors are rare. Therefore, we used a deep learning model for prediction in this study.

Objective: This study aimed to demonstrate that with the proper data set and learning strategies, we can predict the 30-day mortality and readmission of IHCA survivors based on their historical claims.

Methods: National Health Insurance Research Database claims data, including 168,693 patients who had experienced IHCA at least once and 1,569,478 clinical records, were obtained to generate a data set for outcome prediction. We predicted the 30-day mortality/readmission after each current record (ALL-mortality/ALL-readmission) and 30-day mortality/readmission after IHCA (cardiac arrest [CA]-mortality/CA-readmission). We developed a hierarchical vectorizer (HVec) deep learning model to extract patients' information and predict mortality and readmission. To embed the textual medical concepts of the clinical records into our deep learning model, we used Text2Node to compute the distributed representations of all medical concept codes as a 128-dimensional vector. Along with the patient's demographic information, our novel HVec model generated embedding vectors to hierarchically describe the health status at the record-level and patient-level. Multitask learning involving two main tasks and auxiliary tasks was proposed. As CA-mortality and CA-readmission were rare, person upsampling of patients with CA and weighting of CA records were used to improve prediction performance.

Results: With the multitask learning setting in the model learning process, we achieved an area under the receiver operating characteristic of 0.752 for CA-mortality, 0.711 for ALL-mortality, 0.852 for CA-readmission, and 0.889 for ALL-readmission. The area under the receiver operating characteristic was improved to 0.808 for CA-mortality and 0.862 for CA-readmission after solving the extremely imbalanced issue for CA-mortality/CA-readmission by upsampling and weighting.

Conclusions: This study demonstrated the potential of predicting future outcomes for IHCA survivors by machine learning. The results showed that our proposed approach could effectively alleviate data imbalance problems and train a better model for outcome prediction.

(J Med Internet Res 2021;23(9):e27798) doi:10.2196/27798

KEYWORDS

in-hospital cardiac arrest; 30-day mortality; 30-day readmission; machine learning; imbalanced dataset

Introduction

Background

In the United States, approximately 209,000 patients experience in-hospital cardiac arrest (IHCA) each year [1]. The rate of survival to hospital discharge is around 14%, and only 7% of IHCA patients could regain an independent life or a partially independent life [2]. In order to reduce the severe effect of IHCA on personal life or society, identifying measures to improve IHCA outcomes is crucial.

Prior Work

Prognostic factors and prediction tools for survivors of IHCA and their neurologic outcomes have been identified in previous studies [3-8]. However, the evidence of an early warning system for predicting the mortality of IHCA survivors is limited. Current early warning scoring systems using physiologic track-and-trigger systems (TTSs) have been developed for identifying patients at risk for IHCA or other serious outcomes including mortality [9-11]. Most of TTSs rely on the routine observations of vital signs carried out by ward staff. Although many patients could be monitored with this approach, the quality of evidence underpinning the use of TTSs is poor. Specifically, most TTSs have low sensitivity, low positive predictive values, and high specificity [12,13]. In addition to the high mortality after IHCA, readmission after IHCA has a significant cost burden and is associated with comorbidities. Predicting readmission events provides the chance for appropriate interventions and reducing health care costs, including further readmission [14,15].

Our Study

Here, we first extracted the IHCA cohort from the National Health Insurance Research Database (NHIRD). We assessed their risk based on historical electronic health records (EHRs) in the NHIRD. To provide a long enough window for clinical intervention, we used the 30-day mortality and readmission after IHCA as our prediction targets. In contrast to TTSs, EHRs are prepared by physicians, and they contain several important medical information, including the diagnosis and management of patients. To achieve a better performance, we developed a novel deep learning model, hierarchical vectorizer (HVec), to analyze the patients' historical EHRs and predict mortality and readmission. This study aimed to demonstrate that with the proper data set and learning strategies, we can predict the outcome of IHCA patients based on their historical claims and help clinicians design more effective intervention programs.

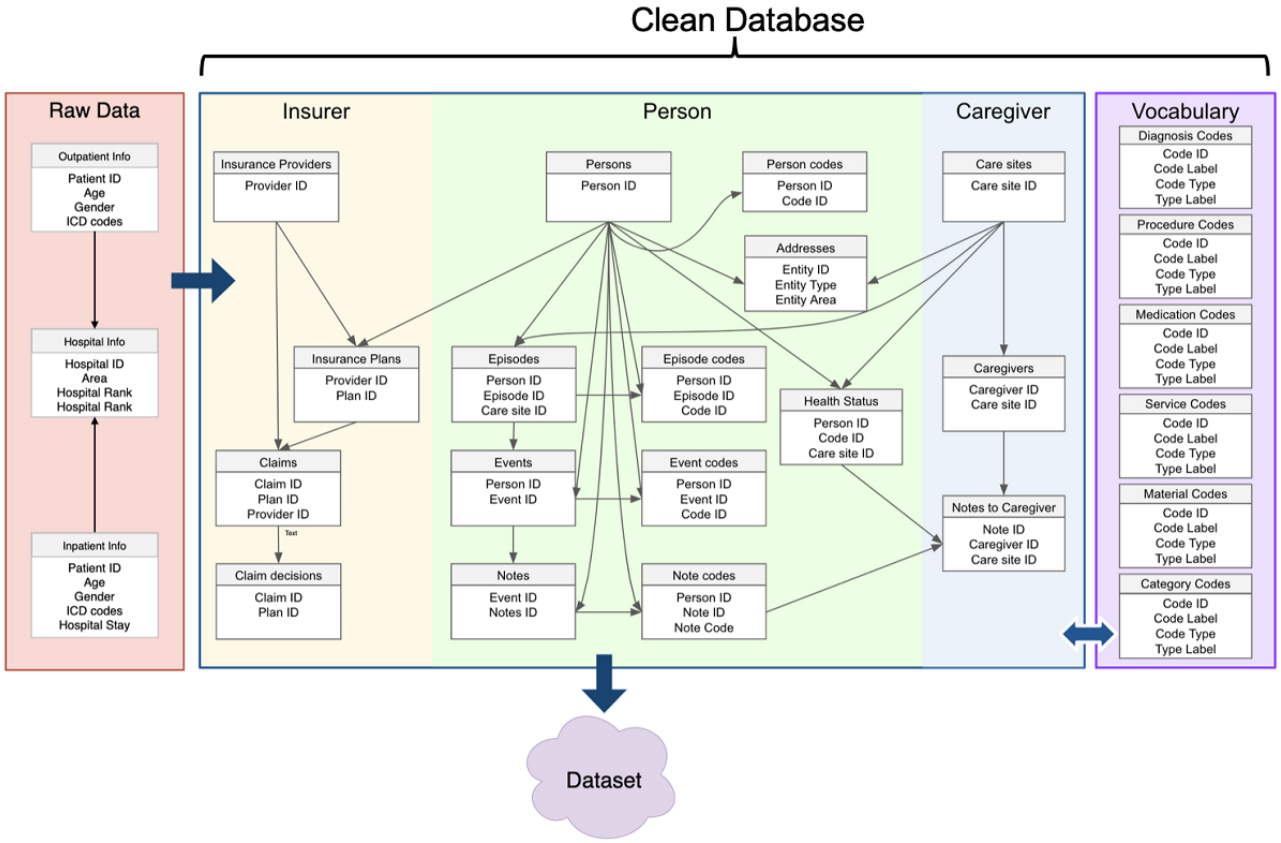
Methods

Data Collection

This study was approved by the Institutional Review Board of National Taiwan University Medical College. The IHCA cohort extracted from the NHIRD consisted of 168,693 patients who had at least one IHCA event over 9 years (between January 1, 2002, and December 31, 2010). The Taiwan National Health Insurance program is the only health insurance scheme in Taiwan and covers up to 99.99% of Taiwan's population [16]. The NHIRD contains all health records in inpatient and outpatient settings (clinic or emergency department); however, the records cannot be specifically linked to each patient. International Classification of Disease, 9th Revision (ICD-9) was used during the study period for diagnosis and medical procedures. The NHIRD contains medical information, including gender, age, diagnosis, medical procedure, operation, medication, laboratory test, care site, discharge status, and cost of each hospital visit. Laboratory test results and bedside information, including vital signs, blood pressure, and physical examination, are not included in the NHIRD.

The IHCA population was defined by inpatient records with the ICD-9 procedure codes 99.60 (cardiopulmonary resuscitation, not otherwise specified) and 99.63 (closed-chest cardiac massage) [17]. We used the extract, transform, and load (ETL, see Figure 1) procedure to process raw data into a clean database by eliminating records with missing or invalid information. Raw data in the cleaned database were re-grouped into three major categories (insurer, person, and caregiver) to improve data organization. In addition, vocabulary tables were constructed based on extracted concepts that were used in the raw data.

Figure 1. ETL process for converting raw NHIRD data into the data set. The raw data are extracted, transformed, and loaded into the cleaned database after cohort selection and eliminating the invalid data. ETL: extract, transform, load; ICD: International Classification of Disease; NHIRD: National Health Insurance Research Database.



Experimental Data Set

The 168,693 patients in the data set were split into three data sets: train, validation, and test. The training data set (70% of the data set) was used to train each model. The remaining 30% of patients were split between the validation and test data set evenly to tune the hyperparameters (Table 1) and evaluate model performance, respectively.

For comparison, we trained two single-task models for both mortality and readmission. Person upsampling and event weighting were only performed on the training data set so as not to affect the distribution of the validation and test data set. The F1 score and area under the receiver operating characteristic (AUROC) were our main evaluation metrics.

Table 1. Hyperparameter settings.

Hyperparameter	Value
Visit embed size	200
RNN[a] output size	128
Learning rate	1e-3
Dropout rate	0.5
l2_weights	0.01
Code embed size	128

[a]RNN: recurrent neural network.

Input Features

In the retrospective review of the data set, each person may have multiple clinical records (inpatient/outpatient visit) to the hospital within a 9-year period. Each clinical record was set as a unit of analysis. For each clinical record, information was extracted and grouped into input and target features (see Tables 2 and 3). The input features consisted of five major groups described as follows:

1. Medical records consisted of five types of codes, and all codes in the clinical records were mapped to over 400,000 clinical concepts. The health status of the patient in the current record can be determined from this group.

2. Demographics included the age, gender, and information of the targeted patients.
3. Care site information included information of the clinical institution where the patient received treatment.
4. Record statistics provided information on the length (by day) of the record, the number of codes in each International Classification of Disease (ICD) code category, and the total monetary cost involved. This could help the model estimate the severity of the patient's disease at the record level.
5. Historical information described previous hospital stays and admissions. It was used to estimate the overall health status of the patient in our model.

Table 2. Summary of the input features of the model.

Group and feature name	Feature description	Dimension
Medical records		
Diagnosis	Diagnosis codes from the health record	128
Procedures	Procedure codes from the health record	128
Meds	Medication codes from the health record	128
Tests	Lab test codes from the health record	128
Other	Other codes from the health record	128
Demographics		
Claim type	Inpatient or outpatient	2
Age	Age at the event	1
Gender	Male or female	2
Care site information		
Care site type	Type of site (public, corporate, or private)	21
Care site specialization	Medical center, community hospital, district hospital, regional hospital, or clinic	5
Care site rank	Rank of the care site	17
Record statistics		
Hospital stay	Duration of current hospital stay	1
Total cost	Monetary cost of each of the five ICD[a] codes	5
Total count	Counts of each of the five ICD codes	5
Historical information		
Past hospitalization duration	The number of days a person spends in the hospital within 3, 6, 12, and 24 months	4
Past admission count	The number of times a person is admitted to the hospital within 3, 6, 12, and 24 months	4

[a]ICD: International Classification of Disease.

Table 3. Summary of the prediction targets.

Group and feature name	Feature description	Dimension
Main target		
Mortality	Whether this event would lead to another mortality event in (within 1 to 30) days	1
Readmission	Whether this event would lead to another readmission event in (within 1 to 30) days	1

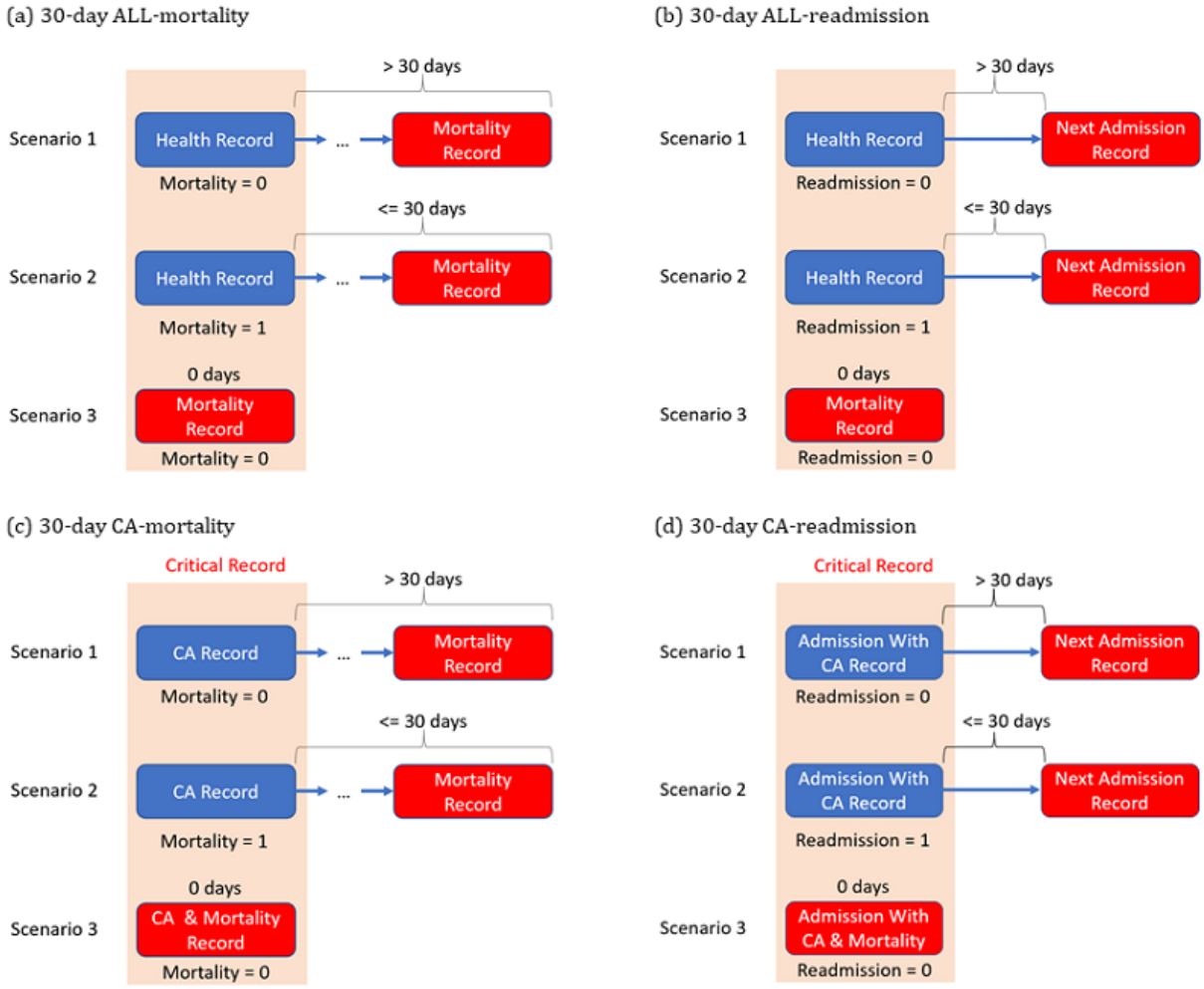
Targets

A threshold of 30 days was set to predict whether a person would die or readmit within 30 days. Mortality was defined when the patient had an inpatient or outpatient record of mortality or was discharged under critical condition following IHCA. For readmission, whether the patient is readmitted within 1 to 30 days from hospital discharge was predicted. In contrast to other mortality studies, records with mortality (0-day mortality record) were excluded in our study. The main purpose of this strategy was to reduce the "leakage" of features in these records. Our initial results showed that the features of mortality records usually contain information (eg, respiratory failure) explicitly indicating patient mortality. Indeed, these features are significant factors for predicting mortality. However, such cases are not beneficial to our model as the severe condition of these patients makes it hard to treat them with any intervention. Moreover, the high degree of correlation of these features and mortality would cause the model to rely on them and underestimate other potential predictive factors. In order to avoid leakage and let the model focus on other predictive factors,

mortality records were set as the negative class, and previous records of mortality (within 30 days) were the positive class.

In clinical practice, the 30-day outcomes of patients after IHCA and discharge from hospitalization are of great interest. The 30-day mortality or readmission after cardiac arrest (CA; CA-mortality/CA-readmission) is a subset of 30-day mortality or readmission. In the rest of the paper, ALL-mortality/ALL-readmission will be used to represent the 30-day mortality or readmission for all records. CA-mortality/CA-readmission refers to the 30-day mortality or readmission following CA (Figure 2).

Figure 2. ALL/CA-mortality and ALL/CA-readmission. 3 scenarios of 30-days mortality & readmission after cardiac arrest. Events in red are the outcomes we want to predict. ALL-mortality: 30-day mortality after all records; ALL-readmission: 30-day readmission after all records; CA: cardiac arrest; CA-mortality: 30-day mortality after cardiac arrest records; CA-readmission: 30-day readmission after cardiac arrest records.

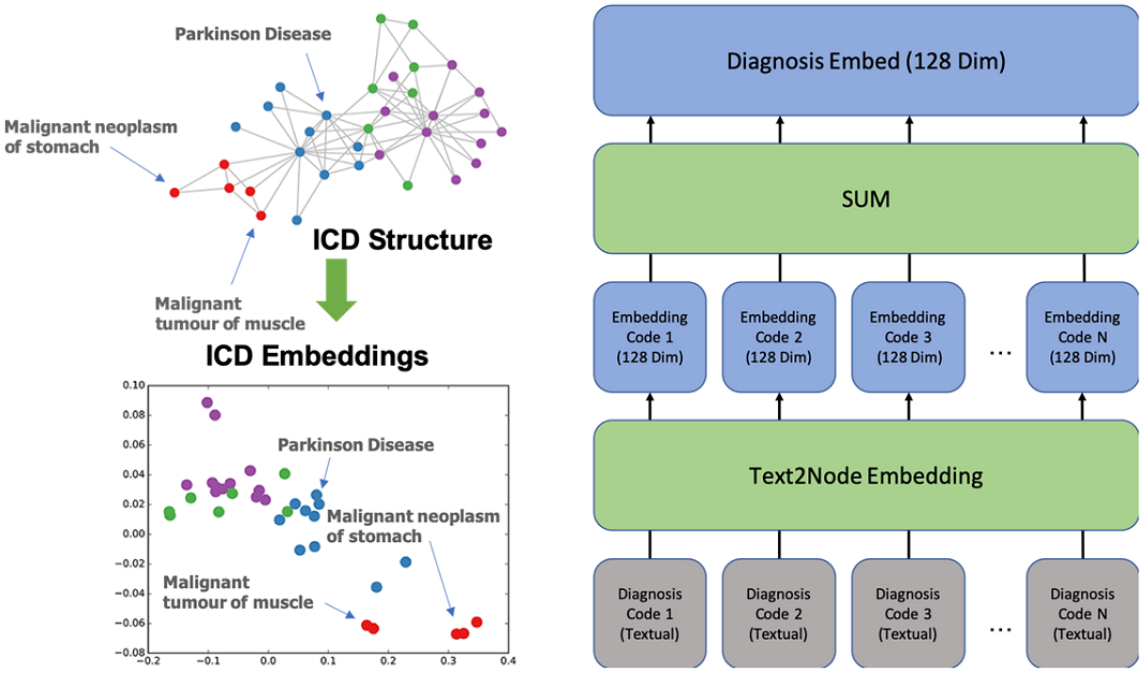


ALL-mortality/ALL-readmission is more common than CA-mortality/CA-readmission as most people would not survive after the first CA event. This means that when we want to predict the future outcome of a recovered CA patient, we do not have enough positive cases for analysis.

Hierarchical Vectorizer (HVec)

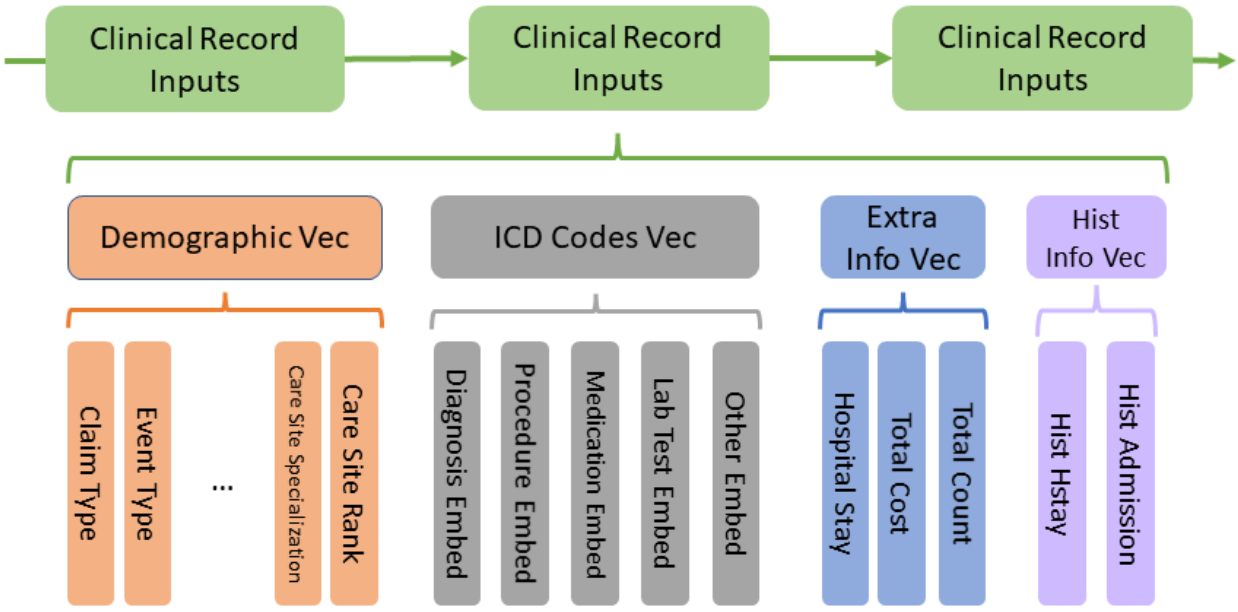
Each record was constructed into a 707-dimension vector for further training. Based on Table 1, all features except for ICD code features (textual features) can be vectorized with one-hot encoding. The features of ICD codes were extracted directly from the health record. A medical knowledge embedding system called Text2Node was used to embed the textual features into vectors [18]. Each of the five categories of ICD codes could contain many ICD codes from a single record, and all codes were added together as a single code for a given category (see Figure 3 as an example). Trained from a substantial medical knowledge database, Text2Node can effectively transform the textual medical concepts into a latent space while preserving the relationship of similar concepts.

Figure 3. Example of Text2Node embedding [18] and code embedding for diagnosis codes in a clinical record. Dim: dimension; ICD: International Classification of Disease.



For each clinical record, by concatenating all feature vectors into group vectors hierarchically (see Figure 4), the clinical record vector was obtained. After sorting each clinical record vector according to the date, time series techniques were used to train a model to predict the outcomes of each record.

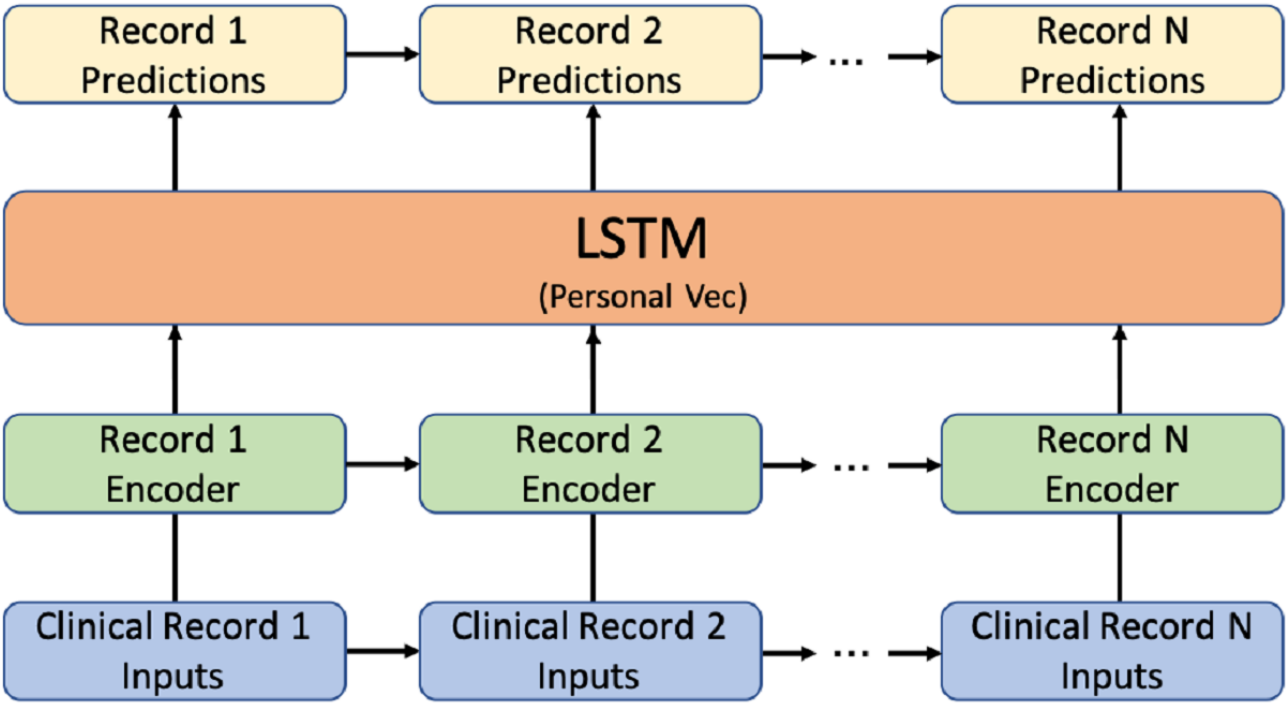
Figure 4. Feature concatenation to generate clinical record vectors for time series analysis. ICD: International Classification of Disease; Vec: vector.



Model Architecture

Deep recurrent neural networks (RNNs) have been proven to be a powerful tool for predicting time series data. In clinical research, the use of RNNs, especially long short-term memory (LSTM), for clinical prediction has been widely investigated [19-21]. Inspired by Choi's work [19], we proposed an HVec model using LSTM networks (Figure 5). In this framework, the record encoder was a fully-connected layer that generated the record embedding for each clinical record independently. The record embedding was a latent vector that contained all the information representing the current clinical record. This latent vector was used as the input of the LSTM to update the person vector (ie, patient status). This person vector was then used to predict our targets.

Figure 5. HVec model using LSTM networks. HVec: hierarchical vectorizer; LSTM: long short-term memory; Vec: vector.



However, most previous studies were based on a relatively balanced data set. As we pointed out earlier, the distribution of CA-mortality/CA-readmission is extremely imbalanced. Training LSTM networks with an extremely imbalanced data set is always challenging because, without a carefully designed training strategy, the model could be biased (ie, predicting the negative class for all records). In order to address this problem, two different training strategies have been proposed to alleviate model bias during the training step:

1. The multitask learning framework was introduced to combine several related learning tasks to regularize the gradient and alleviate data imbalance problems during training.
2. At the person level, CA-mortality/CA-readmission records were upsampled, and at the record level, higher weights were assigned to CA records.

Multitask Learning

The multitask learning framework was proposed, and several auxiliary related tasks were added to the HVec outputs (Figure 6). In this framework, the main functions were mortality and readmission. Here, instead of dividing each main task output into ALL-mortality/ALL-readmission and CA-mortality/CA-readmission independently, an output was considered to cover both because the latter is a subset of the former. Although the distribution was different, they still can be achieved simultaneously with the proposed person upsampling and CA record weighting (described in the next section). Inspired by a previous study [22], three auxiliary autoencoder tasks were introduced to help the model learn the embeddings (Figure 6). Two self-supervised regression tasks were also introduced to allow the embedding to "memorize" the current cost and predict the future cost. Furthermore, another classification task was implemented to predict whether a record is an IHCA record considering that we observed the correlation of IHCA to mortality in the previous analysis.

To monitor the gradients of different tasks and regularize the learning process with auxiliary tasks, Theorem 1 was adopted from Du et al [23].

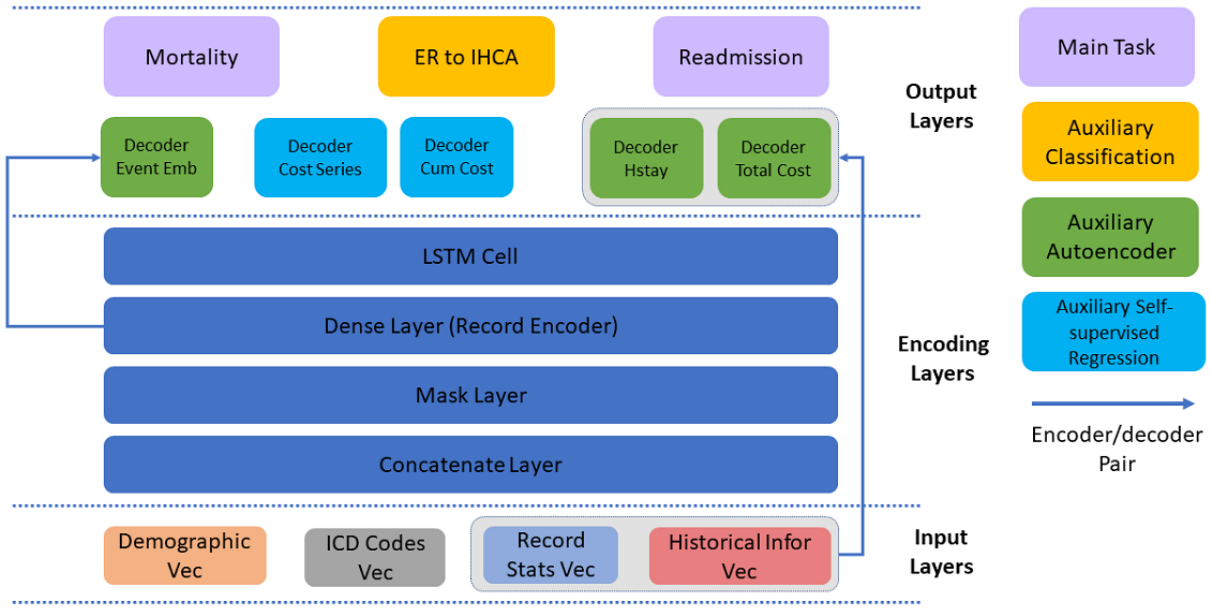
Theorem 1 given any gradient vector field $G(\theta^{(t)}) = \nabla_{\theta}$ (θ) (1) to denote the main task and an arbitrary vector field $V(\theta^{(t)})$ to denote the gradient from another auxiliary task, the update strategy using:

$$\theta^{(t+1)} := \theta^{(t)} - \alpha^{(t)} (G(\theta^{(t)}) + V(\theta^{(t)}) + \max (0, \cos(G(\theta^{(t)}), V(\theta^{(t)}))) \textbf{ (2)}$$

with a proper can coverage to a local minimum.

Following this theorem, HVec can learn and converge with a large data set.

Figure 6. Multiple outputs for multitask learning. Cum: cumulative; Emb: embedding; ER: emergency room; Hstay: hospital staying; ICD: International Classification of Disease; IHCA: in-hospital cardiac arrest; Infor: information; LSTM: long short-term memory; Stats: statistics; Vec: vector.



Person Upsampling and CA Record Weighting

Although the gradients from different tasks can be monitored in multitask learning, with a heavily imbalanced data set (eg, CA-mortality), the auxiliary tasks may fail to regularize the main task. When we trained the HVec model, all records of a person were treated as a single sequential record and fed together into the model. Therefore, in each batch, the batch size was equal to the number of people in the batch. Compared with the number of all records, the number of CA records for each person was relatively rare. CA-mortality and CA-readmission were rare compared with ALL-mortality and ALL-readmission.

The weighting strategy [24] was proposed to solve this problem from two perspectives: at the person level, patients with CA-mortality/CA-readmission records were upsampled per batch (see Figure 7); at the record level, a higher weight was assigned to CA records to make the objective function more sensitive to CA-mortality/CA-readmission records. The upsampling of patients with CA-mortality/CA-readmission records can guarantee that at the person level, there are more CA-mortality/CA-readmission records [25].

Assigning a higher weight to CA records could also emphasize the CA records during training by modifying the loss functions accordingly. Considering N-loss functions $\{L_1,\ldots,L_N\}$ corresponding to auxiliary tasks, the loss function can be written as (3).

Where (4) is a function that equals 1 if the statement in the bracket is true and otherwise 0; w_{pos} and w_{neg} are the positive and negative class weight, respectively. Combined with the gradient update strategy in equation 1, the HVec can learn from the extremely imbalanced data set effectively.

Figure 7. Person upsampling.

Input: Total iteration T, positive class people S_{pos}, negative class people S_{neg}, batch size n, desired imbalance ratio δ.

Output: Model M

$$S_{pos} \longrightarrow it_{pos}, S_{neg} \longrightarrow it_{neg}$$

While Training **do**

In each SGD batch, $B_{all} \leftarrow \emptyset$

$$n * \delta \longrightarrow n_{pos}, n * (1-\delta) \longrightarrow n_{neg}$$

IF $|it_{pos}| < n_{pos}$:

$$it_{pos} \bigcup B_{all} \longrightarrow B_{all}$$

$|it_{pos}| \longrightarrow r$, $S_{pos} \longrightarrow it_{pos}$

random sample $|b_{pos}| = (1-r)$ from it_{pos}, $it_{pos} - S_{remain} \longrightarrow it_{pos}$

ELSE:

random sample $|b_{pos}| = r$ from it_{pos}, $it_{pos} - b_{pos} \longrightarrow it_{pos}$

END

$$b_{pos} \bigcup B_{all} \longrightarrow B_{all}$$

IF $|it_{neg}| < n_{neg}$:

$it_{neg} \cup B_{all} \longrightarrow B_{all}$,

$$|it_{neg}| \longrightarrow r, S_{neg} \longrightarrow it_{neg}$$

random sample $|b_{neg}| = (1-r)$ from it_{neg}, $it_{neg} - S_{remain} \longrightarrow it_{neg}$

ELSE:

random sample $|b_{neg}| = r$ from it_{neg}, $it_{neg} - b_{neg} \longrightarrow it_{neg}$

END

$$b_{neg} \bigcup B_{all} \longrightarrow B_{all}$$

Train model M with B_{all} and *Equation 1*

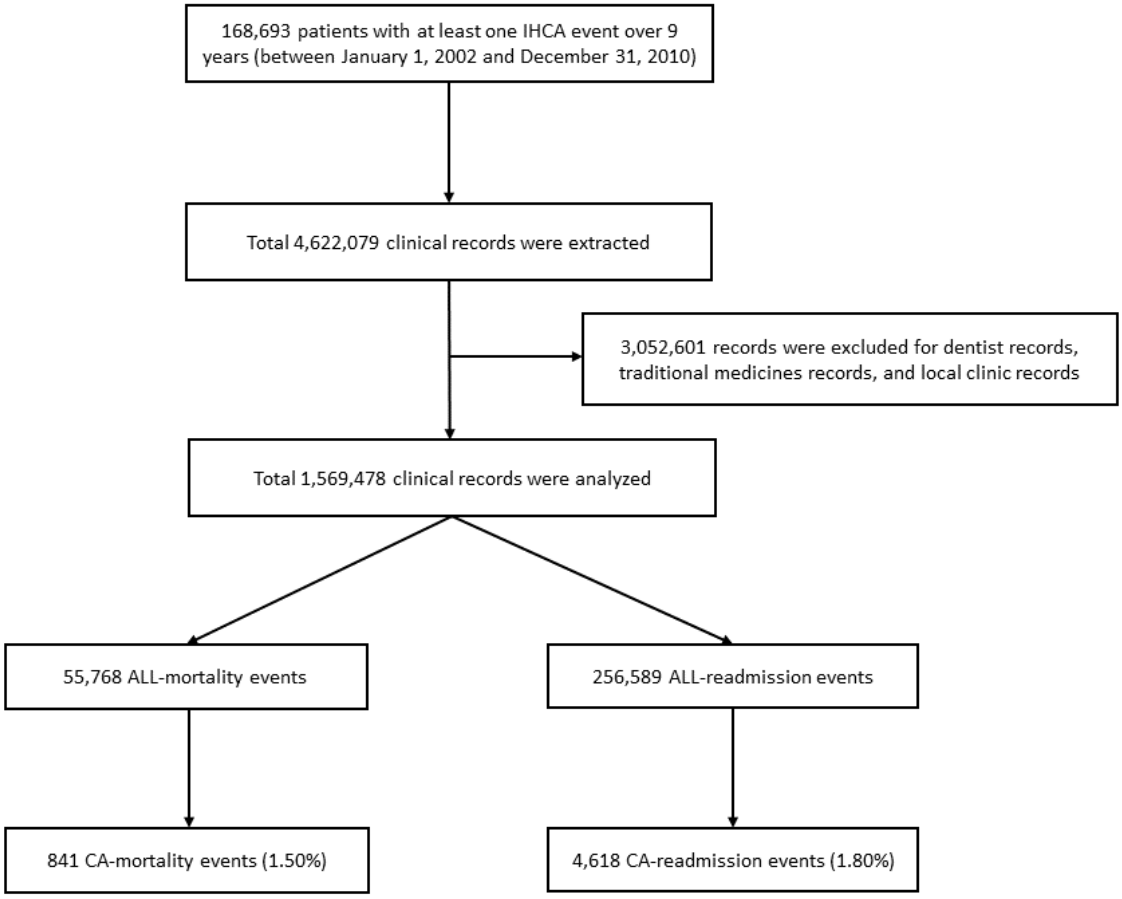
END

Results

Overview

A total of 168,693 patients and 4,622,079 clinical records were extracted from the NHIRD over 9 years, and 3,052,601 records (dentist records, traditional medicine records, or local clinic records) were excluded because these records were concentrated with repetitive conditions and mainly added noise to the machine learning models (Figure 8).

Figure 8. CONSORT diagram of the study cohort. ALL-mortality: 30-day mortality after all records; ALL-readmission: 30-day readmission after all records; CA: cardiac arrest; CA-mortality: 30-day mortality after CA records; CA-readmission: 30-day readmission after CA records; IHCA: in-hospital cardiac arrest.



There were 1,569,478 clinical records in the cleaned database, including both inpatient and outpatient records, from 168,693 patients (mean number of records per person: 9.30, SD 10.90) who have experienced at least one IHCA event. The results indicate an imbalanced data set, where the most imbalanced task was CA-mortality with a ratio of 0.53%. The characteristics of the study population are summarized in Table 4. There were 173,345 IHCA records (11.04% of the total clinical records), and on average, there were 1.02 IHCA records for each person.

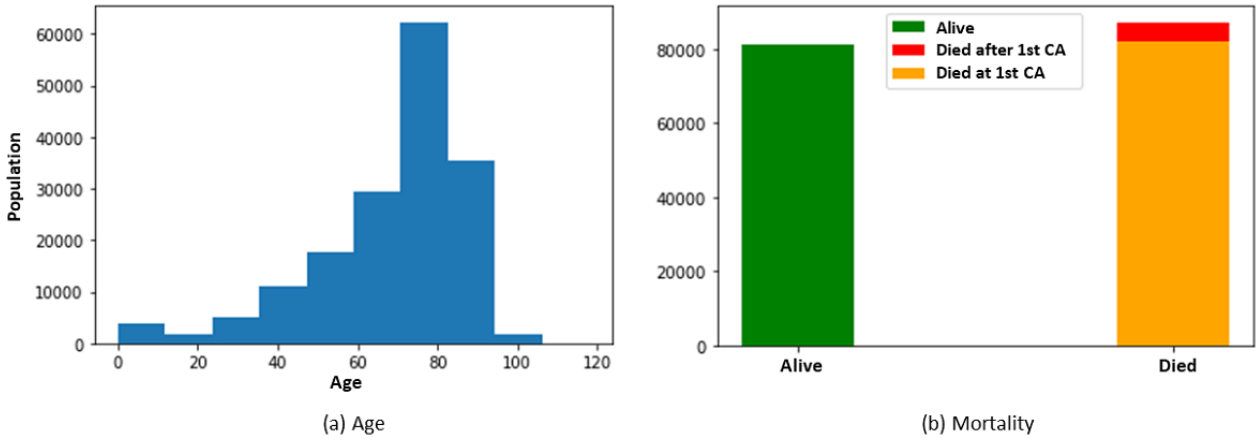
The age of the patients in the data set ranged from 0 (newborn) to 118 years (mean age 68.66, SD 18.96 years), including 104,691 females and 64,002 males. Overall, 164,322 patients (97.4%) had CA only once, 4,174 patients (2.4%) had CA twice, and only 197 patients (0.2%) had CA more than twice. Death was recorded for 87,311 patients (51.75% mortality rate). Of these 87,311 patients, 82,225 patients died during their first hospitalization for CA (94.17%; Figure 9).

Table 4. Characteristics of the study population.[a]

Characteristics	Study population (N=168,693)
Age (years), mean (SD)	68.66 (18.96)
Gender (male), n (%)	64,002 (37.9)
Record number per person, mean (SD)	9.30 (10.90)
Cardiac arrest frequency, n (%)	
1	164,322 (97.4)
2	4174 (2.4)
≥3	197 (0.2)
Mortality, n (%)	87,311 (51.75)

[a]Continuous variables are presented as the mean (SD), and categorical variables are presented as the number (percentage of the study population).

Figure 9. Age and mortality statistics of the data set. CA: cardiac arrest.



(a) Age (b) Mortality

Experiment 1: Single-Task Learning Versus Multitask Learning

In this experiment, person upsampling and event weighting were not applied. The model performance is summarized in Tables 5 and 6.

As shown in Table 5, multitask learning could improve the model performance for ALL-mortality and CA-mortality in terms of the AUROC and F1 scores. Based on single-task and multitask results, there was a relatively high improvement in performance for the extremely imbalanced CA-mortality task compared with the ALL-mortality task. However, the precision was relatively low due to the imbalance ratio, which also affected the F1 score. Moreover, the F1 score of CA-mortality was too low for real-life applications.

Table 5. Single-task and multitask learning performance for CA-mortality and ALL-mortality.[a]

Mortality	CA-mortality				ALL-mortality			
	AUROC[b]	F1	Precision	Recall	AUROC	F1	Precision	Recall
Single-task learning	0.658	0.014	0.010	0.024	0.663	0.130	0.101	0.180
Multitask learning	0.752	0.049	0.041	0.060	0.711	0.147	0.093	0.349

[a]CA-mortality: cardiac arrest mortality (30-day mortality after CA records); ALL-mortality: 30-day mortality after all records.

[b]AUROC: area under the receiver operating characteristic.

As shown in Table 6, the improvement in multitask learning for readmission prediction was not as significant as that for mortality prediction. Furthermore, in CA-readmission prediction, the F1 score and precision were decreased.

Table 6. Single-task and multitask learning performance for CA-readmission and ALL-readmission.[a]

Readmission	CA-readmission				ALL-readmission			
	AUROC[b]	F1	Precision	Recall	AUROC	F1	Precision	Recall
Single-task learning	0.847	0.214	0.162	0.315	0.872	0.554	0.424	0.801
Multitask learning	0.852	0.209	0.152	0.335	0.889	0.562	0.430	0.811

[a]CA-readmission: cardiac arrest readmission (30-day readmission after CA records); ALL-readmission: 30-day readmission after all records.

[b]AUROC: area under the receiver operating characteristic.

In this experiment, compared with single-task learning, multitask learning could achieve a better performance for ALL-mortality and ALL-readmission. However, multitask learning could not solve the extremely imbalanced data set.

Experiment 2: Improving CA Prediction Performance

Experiment 1 showed that models had difficulties making good predictions based on the CA-mortality and CA-readmission data due to the extremely imbalanced data set. In this experiment, we demonstrated that by applying person upsampling and event weighting, we could further improve the performance for CA-mortality and CA-readmission without considerably affecting ALL-mortality and ALL-readmission. In our experiment, the upsampling rate indicates how many times upsampling was performed for the positive class (patients with CA-mortality or CA-readmission) in a batch, and an upsampling rate of 1 means we did not perform upsampling. The CA event weight indicates the loss weight w_{pos} in equation 3, and we always set w_{neg}=1. We used different upsampling rates and event weights in our experiments. Our results are

summarized in Tables 7 and 8, including the previous results on single-task model performance (first row of each table).

Overall, compared with the 30-day mortality task, the imbalanced 30-day mortality task showed a larger increase in performance when applying balancing techniques. As shown in Table 7, after upsampling and event weighting, the models demonstrated improved performance for both ALL-mortality and CA-mortality. For CA-mortality, the F1 score was increased by 36.7% (from 0.049 to 0.067).

Table 7. Mortality models with their respective hyperparameter configuration and their performance in predicting both CA-mortality and ALL-mortality.[a]

Mortality			CA-mortality		ALL-mortality	
Upsampling rate	Event weight	Multitask	AUROC[b]	F1	AUROC	F1
1	1	No	0.658	0.014	0.663	0.130
1	1	Yes	0.752	0.049	0.711	0.147
10	1	Yes	0.808	0.064	0.728	0.155
10	5	Yes	0.802	0.067	0.726	0.158

[a]CA-mortality: cardiac arrest mortality (30-day mortality after CA records); ALL-mortality: 30-day mortality after all records.

[b]AUROC: area under the receiver operating characteristic.

Similar to ALL-mortality and CA-mortality, the AUROC and F1 score were increased for both ALL-readmission and CA-readmission by applying the two techniques. Based on the results in Table 8, a minor but consistent increase was achieved in CA-readmission prediction. However, there was no significant improvement in ALL-readmission prediction. This is because upsampling and event weighting would not greatly affect the learning process for a more balanced data set.

Table 8. Readmission models with their respective hyperparameter configuration and their performance in predicting CA-readmission and ALL-readmission.[a]

Readmission			CA-readmission		ALL-readmission	
Upsampling rate	Event weight	Multitask	AUROC[b]	F1	AUROC	F1
1	1	No	0.847	0.214	0.872	0.554
1	1	Yes	0.852	0.209	0.889	0.562
5	1	Yes	0.861	0.230	0.884	0.555
5	5	Yes	0.862	0.237	0.884	0.555

[a]CA-readmission: cardiac arrest readmission (30-day readmission after CA records); ALL-readmission: 30-day readmission after all records.

[b]AUROC: area under the receiver operating characteristic.

In summary, the results indicated that multitask learning, upsampling, and event weighting could effectively improve the model prediction performance for an imbalanced data set. We also showed that these techniques could be collectively used to achieve better results for an extremely imbalanced data set.

Discussion

In this study, we constructed a large patient database that includes 9 years of EHRs for over 168,000 IHCA patients, which can be used for future IHCA-related research. In addition, we developed an HVec model (LSTM model) that uses a multitask learning strategy to predict the 30-day mortality and readmission.

The results showed that our model could successfully predict future mortality and readmission using EHR data for IHCA patients. We proposed the person upsampling and record weighting strategies to handle the extremely imbalanced data problem in this study. After applying these techniques, some improvements were achieved in CA-mortality and CA-admission prediction.

In contrast to other studies using deep learning models to predict another IHCA event after CA [26], our study focused on predicting the future outcomes of IHCA patients after discharge. To the best of our knowledge, this is the first study to predict mortality and readmission after IHCA events by machine learning. The model may serve as a surveillance system for those who experienced IHCA. Patients with a high risk of mortality or readmission in the near future could be identified and re-evaluated before discharge. This study also demonstrated the potential of another model for predicting future mortality and readmission after each record using previous EHRs (ALL-mortality/ALL-readmission). The model might help identify those with a high risk in inpatient and outpatient situations. However, mortality and readmission rates are different in the general population compared with the patients selected in this study. Using hospital EHRs with patients' information in the NHIRD, we can construct a real-time alert system based on machine learning methods to predict the adverse events of IHCA survivors and improve their outcomes. Further prospective studies are needed to verify the utility of this system in the general population.

Several studies have reported models for predicting the outcomes of CA patients [8]. In a systematic review of current prediction models, the median AUROC value was 0.84 with an IQR of 0.80 to 0.89 [8]. For IHCA survivors, Chan et al. reported that the cardiac arrest survival post-resuscitation in-hospital (CASPRI) score could be used to predict favorable neurologic outcomes after discharge [3]. The AUROC of the CASPRI score was 0.80. Nanayakkara et al used deep learning models to predict the IHCA events of CA survivors, and the AUROC was 0.87 [26]. In our study, we encountered the difficulty of imbalanced data. Using proper learning strategies, we achieved comparable AUROC values (0.808 for CA-mortality and 0.862 for CA-readmission). However, the recall and precision rates were low in our study due to imbalanced data. This is a challenge we aim to resolve. When using historical medical records to predict outcomes, in many cases, the records contain information that may indicate the outcomes. For example, critical diagnosis and rescue medication are often associated with mortality. Including these types of information can facilitate training and give a high AUROC and F1; however, the model itself is of limited use. Mortality records were set as the negative class in our study to avoid overfitting. In further studies, the model may be improved by adjusting the threshold to optimize the trade-off between specificity and sensitivity.

In our HVec framework, we encoded each person's EHRs in two levels of latent vectors (record-level and person-level) and ensured that the model learns both simultaneously using the unsupervised autoencoder strategy. The predictive results were promising with these latent vectors. The latent vectors in these two levels may be further explored to facilitate clinicians' decision-making and provide better clinical interventions. In addition, the person vector may be used as a biomarker to evaluate the overall health status of a person beyond the health care setting. Along with some recently developed models such as Deep Patient [22] and MixEHR [27], we showed that the use of deep neural networks to extract information from EHRs might solve complex clinical research problems.

An imbalanced data distribution is common in clinical research, especially for disease-related predictions. In comparison with common diseases, many important diseases lack positive cases, making it difficult to train a good model. In a previous study [19], the authors attempted to address this problem by using a balanced, distributed data set and train a deep learning model with the balanced data set. Similarly, in this study involving CA-mortality and CA-readmission, we demonstrated that by carefully designing model learning strategies (eg, multitask learning and upsampling), some common problems in clinical research could be solved effectively with machine learning models.

This study has some limitations. First, the IHCA cohort was retrospectively extracted from the NHIRD. Further studies are needed to evaluate the efficacy of this model as an early warning system and determine how this system affects patients' outcomes. Second, as our model was developed based on the NHIRD, the generalization of this model to other health insurance data sets is not proven. Third, each patient's vital signs and laboratory data were not included in the analysis due to the study design. A combination of EHRs and patients' clinical data may further improve model performance. Fourth, traditional machine learning methods have the limitation of interpretability. Specific risk factors for 30-day mortality/readmission were unknown in this study. In future studies, we plan to develop an explainable model and investigate specific predictive factors in the model.

In summary, our model showed good performance in predicting 30-day mortality and readmission after IHCA, which can help clinicians monitor CA patients' status better. We aim to provide more insights to clinicians with proactive intervention recommendations. Nevertheless, a challenge remains in the interpretative ability of the deep learning model. Our future work will mainly focus on the interpretative power of the model trained using EHRs.

Acknowledgments

This study was supported by the Ministry of Science and Technology, Taiwan (project: 109-2634-F-002-031). CYC, YLH, and CHH were involved in data collection. SA, AW, KCF, and JX were involved in constructing the deep learning model. CYC and SA contributed equally to this paper and drafted the article. Finally, YLH, CHH, and RS made the final revision. YLH, CHH, and RS contributed equally as corresponding authors. All authors approved the final version of the article.

Conflicts of Interest

None declared.

References

1. Writing GM, Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, et al. Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association. Circulation 2016 Jan 26;133(4):e38-e360. [doi: 10.1161/CIR.0000000000000350] [Medline: 26673558]
2. Wang C, Huang C, Chang W, Tsai M, Yu P, Wu Y, et al. Outcomes of Adult In-Hospital Cardiac Arrest Treated with Targeted Temperature Management: A Retrospective Cohort Study. PLoS One 2016;11(11):e0166148 [FREE Full text] [doi: 10.1371/journal.pone.0166148] [Medline: 27820847]

3. Chan PS, Spertus JA, Krumholz HM, Berg RA, Li Y, Sasson C, Get With the Guidelines-Resuscitation Registry Investigators. A validated prediction tool for initial survivors of in-hospital cardiac arrest. Arch Intern Med 2012 Jun 25;172(12):947-953 [FREE Full text] [doi: 10.1001/archinternmed.2012.2050] [Medline: 22641228]
4. Holmberg MJ, Moskowitz A, Raymond TT, Berg RA, Nadkarni VM, Topjian AA, American Heart Association's Get With The Guidelines-Resuscitation Investigators. Derivation and Internal Validation of a Mortality Prediction Tool for Initial Survivors of Pediatric In-Hospital Cardiac Arrest. Pediatr Crit Care Med 2018 Mar;19(3):186-195 [FREE Full text] [doi: 10.1097/PCC.0000000000001416] [Medline: 29239980]
5. Chan PS, Nallamothu BK, Krumholz HM, Spertus JA, Li Y, Hammill BG, American Heart Association Get with the Guidelines–Resuscitation Investigators. Long-term outcomes in elderly survivors of in-hospital cardiac arrest. N Engl J Med 2013 Mar 14;368(11):1019-1026 [FREE Full text] [doi: 10.1056/NEJMoa1200657] [Medline: 23484828]
6. Chelly J, Mpela A, Jochmans S, Brunet J, Legriel S, Guerin L, et al. OHCA (Out-of-Hospital Cardiac Arrest) and CAHP (Cardiac Arrest Hospital Prognosis) scores to predict outcome after in-hospital cardiac arrest: Insight from a multicentric registry. Resuscitation 2020 Nov;156:167-173. [doi: 10.1016/j.resuscitation.2020.09.021] [Medline: 32976962]
7. Chung C, Chiu W, Huang Y, Chan L, Hong C, Chiu H. Identifying prognostic factors and developing accurate outcome predictions for in-hospital cardiac arrest by using artificial neural networks. J Neurol Sci 2021 Jun 15;425:117445. [doi: 10.1016/j.jns.2021.117445] [Medline: 33878655]
8. Carrick RT, Park JG, McGinnes HL, Lundquist C, Brown KD, Janes WA, et al. Clinical Predictive Models of Sudden Cardiac Arrest: A Survey of the Current Science and Analysis of Model Performances. JAHA 2020 Aug 18;9(16):e017625. [doi: 10.1161/jaha.119.017625]
9. Jansen JO, Cuthbertson BH. Detecting critical illness outside the ICU: the role of track and trigger systems. Curr Opin Crit Care 2010 Jun;16(3):184-190. [doi: 10.1097/MCC.0b013e328338844e] [Medline: 20305556]
10. Smith GB, Prytherch DR, Schmidt PE, Featherstone PI. Review and performance evaluation of aggregate weighted 'track and trigger' systems. Resuscitation 2008 May;77(2):170-179. [doi: 10.1016/j.resuscitation.2007.12.004] [Medline: 18249483]
11. Gao H, McDonnell A, Harrison DA, Moore T, Adam S, Daly K, et al. Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. Intensive Care Med 2007 Apr;33(4):667-679. [doi: 10.1007/s00134-007-0532-3] [Medline: 17318499]
12. Sharpley JT, Holden JC. Introducing an early warning scoring system in a district general hospital. Nurs Crit Care 2004;9(3):98-103. [doi: 10.1111/j.1362-1017.2004.0063.x] [Medline: 15152751]
13. Priestley G, Watson W, Rashidian A, Mozley C, Russell D, Wilson J, et al. Introducing Critical Care Outreach: a ward-randomised trial of phased introduction in a general hospital. Intensive Care Med 2004 Jul;30(7):1398-1404. [doi: 10.1007/s00134-004-2268-7] [Medline: 15112033]
14. Chan PS, Nallamothu BK, Krumholz HM, Curtis LH, Li Y, Hammill BG, American Heart Association's Get With The Guidelines-Resuscitation Investigators. Readmission rates and long-term hospital costs among survivors of an in-hospital cardiac arrest. Circ Cardiovasc Qual Outcomes 2014 Nov;7(6):889-895 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.114.000925] [Medline: 25351479]
15. Morita Y, Haruna T, Haruna Y, Nakane E, Yamaji Y, Hayashi H, et al. Incidence, predictors, causes, and costs of 30-day readmission after in-hospital cardiopulmonary resuscitation in the United States. Resuscitation 2019 Jan;134:19-25. [doi: 10.1016/j.resuscitation.2018.12.001] [Medline: 30566891]
16. Hsieh C, Su C, Shao S, Sung S, Lin S, Kao Yang Y, et al. Taiwan's National Health Insurance Research Database: past and future. Clin Epidemiol 2019;11:349-358 [FREE Full text] [doi: 10.2147/CLEP.S196293] [Medline: 31118821]
17. DeZorzi C, Boyle B, Qazi A, Luthra K, Khera R, Chan PS, et al. Administrative Billing Codes for Identifying Patients With Cardiac Arrest. J Am Coll Cardiol 2019 Apr 02;73(12):1598-1600 [FREE Full text] [doi: 10.1016/j.jacc.2019.01.030] [Medline: 30922482]
18. Soltani R, Tomberg A. Text2node: a cross-domain system for mapping arbitrary phrases to a taxonomy. arXiv. 2019 Apr 11. URL: https://arxiv.org/abs/1905.01958v1 [accessed 2019-04-11]
19. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. JMLR Workshop Conf Proc 2016 Aug;56:301-318 [FREE Full text] [Medline: 28286600]
20. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med 2018;1:18 [FREE Full text] [doi: 10.1038/s41746-018-0029-1] [Medline: 31304302]
21. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. IEEE J Biomed Health Inform 2018 Dec;22(5):1589-1604. [doi: 10.1109/JBHI.2017.2767063] [Medline: 29989977]
22. Miotto R, Li L, Kidd BA, Dudley JT. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. Sci Rep 2016 Dec 17;6:26094 [FREE Full text] [doi: 10.1038/srep26094] [Medline: 27185194]
23. Du Y, Czarnecki W, Jayakumar S, Pascanu R, Lakshminarayanan B. Adapting auxiliary losses using gradientsimilarity. arXiv 2018 Dec 5:1812 [FREE Full text]
24. Haibo He, Garcia E. Learning from Imbalanced Data. IEEE Trans Knowl Data Eng 2009 Sep;21(9):1263-1284. [doi: 10.1109/TKDE.2008.239]

25. Ertekin Ş. Adaptive Oversampling for Imbalanced Data Classification. In: Gelenbe E, Lent R, editors. Information Sciences and Systems 2013. Lecture Notes in Electrical Engineering, vol 264. Cham: Springer; Sep 24, 2013:261-269.
26. Nanayakkara S, Fogarty S, Tremeer M, Ross K, Richards B, Bergmeir C, et al. Characterising risk of in-hospital mortality following cardiac arrest using machine learning: A retrospective international registry study. PLoS Med 2018 Nov;15(11):e1002709 [FREE Full text] [doi: 10.1371/journal.pmed.1002709] [Medline: 30500816]
27. Li Y, Nair P, Lu XH, Wen Z, Wang Y, Dehaghi AAK, et al. Inferring multimodal latent topics from electronic health records. Nat Commun 2020 May 21;11(1):2536 [FREE Full text] [doi: 10.1038/s41467-020-16378-3] [Medline: 32439869]

Abbreviations

AUROC: area under the receiver operating characteristic
CA: cardiac arrest
CASPRI: cardiac arrest survival post-resuscitation in-hospital
CPR: cardiopulmonary resuscitation
EHR: electronic health record
ETL: extract, transform, and load
HVec: hierarchical vectorizer
ICD: International Classification of Disease
IHCA: in-hospital cardiac arrest
LSTM: long short-term memory
NHIRD: National Health Insurance Research Database
OHCA: out-of-hospital cardiac arrest
RNN: recurrent neural network
TTS: track-and-trigger system

Edited by R Kukafka; submitted 08.02.21; peer-reviewed by A Ćirković, W Yang, SF Sung, B Richards; comments to author 15.03.21; revised version received 09.05.21; accepted 21.07.21; published 13.09.21.

Please cite as:
Chi CY, Ao S, Winkler A, Fu KC, Xu J, Ho YL, Huang CH, Soltani R
Predicting the Mortality and Readmission of In-Hospital Cardiac Arrest Patients With Electronic Health Records: A Machine Learning Approach
J Med Internet Res 2021;23(9):e27798
URL: https://www.jmir.org/2021/9/e27798
doi:10.2196/27798
PMID:34515639

Original Paper

The Classification of Six Common Skin Diseases Based on Xiangya-Derm: Development of a Chinese Database for Artificial Intelligence

Kai Huang[1,2,3,4], MD; Zixi Jiang[1,2,3,5], BA; Yixin Li[1,2,3,5], BA; Zhe Wu[6], MD; Xian Wu[6], PhD; Wu Zhu[1,2,3,4], MD, PhD; Mingliang Chen[1,2,3,4], MD, PhD; Yu Zhang[7], MD, PhD; Ke Zuo[8], MD, PhD; Yi Li[9], MD, PhD; Nianzhou Yu[1,2,3,4], BA; Siliang Liu[1,2,3,4], BA; Xing Huang[1,2,3,4], BA; Juan Su[1,2,3,4], MD, PhD; Mingzhu Yin[1,2,3,4], MD, PhD; Buyue Qian[10], MD, PhD; Xianggui Wang[2,3,4,11*], MD, PhD; Xiang Chen[1,2,3,4*], MD, PhD; Shuang Zhao[1,2,3,4*], MD, PhD

[1]Department of Dermatology, Xiangya Hospital, Central South University, Changsha, China

[2]Hunan Engineering Research Center of Skin Health and Disease, Xiangya Hospital, Central South University, Changsha, China

[3]Hunan Key Laboratory of Skin Cancer and Psoriasis, Xiangya Hospital, Central South University, Changsha, China

[4]National Clinical Research Center of Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China

[5]Xiangya School of Medicine, Central South University, Changsha, China

[6]Tencent Medical AI Lab, Shenzhen, China

[7]Day Surgery Center, Xiangya Hospital, Central South University, Changsha, China

[8]Department of Computer Science, National University of Defense Technology, Changsha, China

[9]School of Automation, Central South University, Changsha, China

[10]Department of Electronic Information Engineering, Xi'an Jiaotong University, Xi'an, China

[11]Department of Ophthalmology, Xiangya Hospital, Central South University, Changsha, China

[*]these authors contributed equally

Corresponding Author:
Shuang Zhao, MD, PhD
Department of Dermatology
Xiangya Hospital
Central South University
87 Xiangya Road, Kaifu District
Changsha
China
Phone: 86 13808485224
Email: shuangxy@csu.edu.cn

Abstract

Background: Skin and subcutaneous disease is the fourth-leading cause of the nonfatal disease burden worldwide and constitutes one of the most common burdens in primary care. However, there is a severe lack of dermatologists, particularly in rural Chinese areas. Furthermore, although artificial intelligence (AI) tools can assist in diagnosing skin disorders from images, the database for the Chinese population is limited.

Objective: This study aims to establish a database for AI based on the Chinese population and presents an initial study on six common skin diseases.

Methods: Each image was captured with either a digital camera or a smartphone, verified by at least three experienced dermatologists and corresponding pathology information, and finally added to the Xiangya-Derm database. Based on this database, we conducted AI-assisted classification research on six common skin diseases and then proposed a network called Xy-SkinNet. Xy-SkinNet applies a two-step strategy to identify skin diseases. First, given an input image, we segmented the regions of the skin lesion. Second, we introduced an information fusion block to combine the output of all segmented regions. We compared the performance with 31 dermatologists of varied experiences.

Results: Xiangya-Derm, as a new database that consists of over 150,000 clinical images of 571 different skin diseases in the Chinese population, is the largest and most diverse dermatological data set of the Chinese population. The AI-based six-category classification achieved a top 3 accuracy of 84.77%, which exceeded the average accuracy of dermatologists (78.15%).

Conclusions: Xiangya-Derm, the largest database for the Chinese population, was created. The classification of six common skin conditions was conducted based on Xiangya-Derm to lay a foundation for product research.

(J Med Internet Res 2021;23(9):e26025) doi:10.2196/26025

KEYWORDS

artificial intelligence; skin disease; convolutional neural network; medical image processing; automatic auxiliary diagnoses; dermatology; skin; classification; China

Introduction

Skin conditions affect 1.9 billion people [1] and are the fourth-leading cause of the nonfatal disease burden across 188 low- and high-income countries [2-4]. Patients with skin diseases usually account for 10% of the number of outpatients in Chinese general hospitals and are also among the most common patients in primary care. However, there is a severe lack of dermatologists, particularly in rural Chinese areas. Based on incomplete statistics from the Chinese Society of Dermatology and Sexually Transmitted Diseases, there are approximately 30,000 dermatologists in China, accounting for 0.0016% of the total number of people with skin diseases.

The total number of dermatologists in China is less than 30,000, which results in a large gap between the supply and demand of dermatological resources, especially in primary care hospitals. In addition, some long-standing problems in the Chinese medical industry still exist, such as the uneven distribution of high-quality doctor resources, primary care hospital doctors prone to misdiagnosing patients, missed diagnoses, and insufficient knowledge of rare diseases [5,6]. Addressing these problems necessitates the use of artificial intelligence (AI) in Chinese dermatology. AI has transformed health care [7] and has become one of the most trending research topics. Deep learning algorithms, powered by advances in computation and the creation of large databases, have recently been shown to outperform humans in visual tasks [8]. Recent studies in the medical field, such as those on automatic lesion segmentation and intelligent disease recognition, have successfully used AI to identify skin diseases based on image data [9]. The use of AI in Chinese dermatology started later than in some high-income countries but has been developing rapidly; it could be a solution to the problems that exist in dermatology.

Using deep learning methods for skin disease identification requires a large amount of data. Skin images from the Chinese population are limited. Most public databases on skin diseases are from the US or European Union (EU) population, with a relatively small size and without corresponding pathology information to support their validity [9,10]. The images in the ISIC database, for example, come mainly from light-skinned populations in the United States, Europe, and Australia [11]. Other researchers prefer to use data sets established based on their own patients. These data sets contain information about limited types of skin diseases, include a smaller sample size, or have decreased reliability [12,13]. These problems hinder the use of Chinese AI products, because the systems developed based on these data sets are unreliable for automatically diagnosing members of the Chinese population. Therefore, it is necessary to build a data set with skin images from the Chinese population that would enable successful diagnosis of skin diseases for Chinese patients.

In this paper, we aim to build a database consisting of skin images based on the Chinese population and present an AI-based method for classifying six common skin diseases to lay a foundation for further product research.

Methods

Construction of Xiangya-Derm

As approved by the ethics committee of Central South University, we captured each image with either a digital camera or a smartphone in the Dermatology Outpatient Department, which relies on the Xiangya Big Data Collection Platform. The images were collected according to uniform standards from many cooperating hospitals in Nanjing, Guangzhou, Wuhan, Changsha (cities in China), etc. Most of these images have corresponding pathology information to ensure the accuracy of diagnosis information. For the other common and easy-to-diagnose skin diseases, images were verified by at least three experienced dermatologists. Meanwhile, the dermatologists were asked to annotate the specific location of the skin lesions in each image using an open source software called LabelImg [14]. Finally, the images with a clear diagnosis were added to Xiangya-Derm, a Chinese database. We classified the existing data into six types of skin diseases (malignant tumor, benign tumor, erythema papule scale, bullous, allergic, and connective tissue) based on the classification of common skin diseases and the amount of existing data.

Study of Six Common Skin Diseases

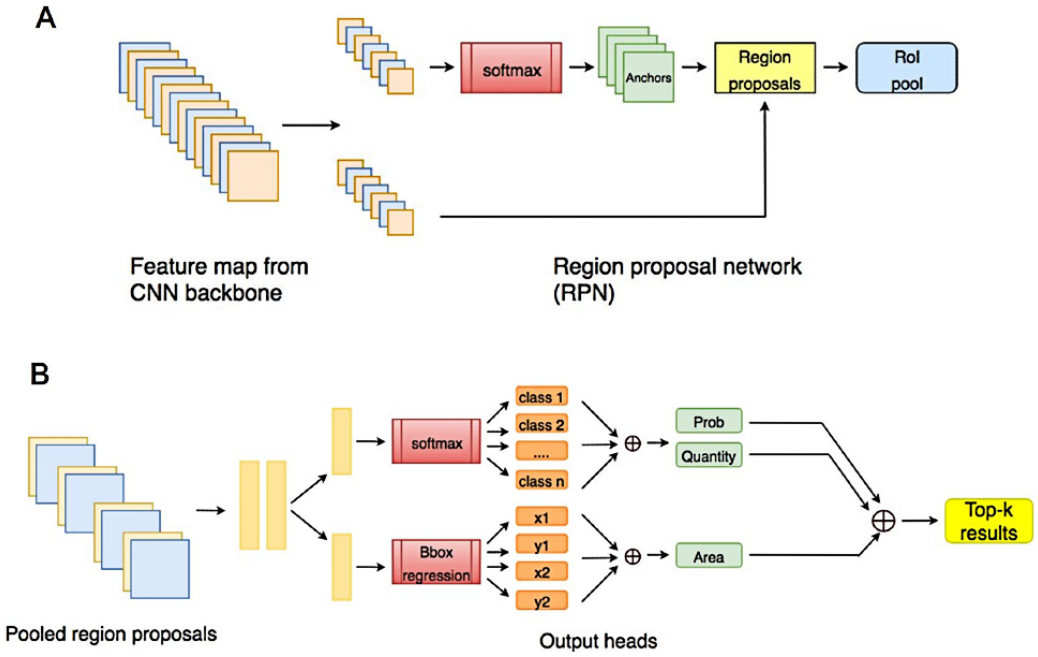
After the construction of Xiangya-Derm, we conducted research on the classification of some common skin diseases from the Xiangya-Derm database using a convolutional neural network (CNN). First, six different subtypes of common dermatoses with additional data magnitudes and common outpatients (psoriasis, basal cell carcinoma, seborrheic keratosis, pemphigus, eczema, and lupus erythematosus) were selected, and a data set consisting of 5660 clinical images was constructed. These six subtypes of skin diseases were selected because they are the most common and representative skin diseases in primary care hospitals in China and because more data were available for these diseases than for any other. This data set contained common skin disease types, such as skin tumors, erythema, and scaling disease. In this data set, the image numbers of psoriasis and other diseases were unbalanced. If the data set had been used directly to classify each type of disease separately, it would have created a severe bias in the network against psoriasis. Therefore, we downsampled psoriasis samples to balance the number of samples among different categories. Moreover, we

re-established a database containing six types of diseases, totaling approximately 3000 images, which ensured that the number of samples in different categories was within a reasonable range. In this range, the disease with the largest number of samples had over 1000 images, while the disease with the fewest samples had fewer than 200 images.

Based on a balanced number of samples in different categories, we proposed a network, called Xy-SkinNet, that was trained to be a six-category model for these six common types of diseases. Before we constructed the model, images in the database for modeling were checked. We found that the background of the clinical images in our database was complex and the size of lesions was relatively small. To overcome these problems and improve the classification performance, the model was designed to first detect the regions of interest that contained lesions. A specially designed information fusion block was then used to output the result by considering all the regions. The network used modified ResNet (M-ResNET) as the backbone network and Faster R-CNN [15], a two-stage object detection method, as the basic structure. Regarding the training and testing data set size, we trained our model with 2400 images and tested it with the remaining 600 images (training size:testing size=4:1) in the database.

Faster R-CNN is an improved version of R-CNN [16] and Fast R-CNN [17]. As a two-stage method, Faster R-CNN comprises two parts. The first part is a region proposal network (RPN), which generates several candidate regions on the entire image according to some preset box (anchor) and keeps only the foreground regions. The second part classifies and locates each remaining region in a fine-grained manner. These two parts share some network parameters. The structure of the model is shown in Figure 1. Based on the basic structure of the method, we improved it by adding an information fusion block for dermatosis characteristics.

Figure 1. Structure of Xy-SkinNet. CNN: convolutional neural network.



Human Versus Computer Competition

To evaluate the performance of Xy-SkinNet and validate its effectiveness in a clinical situation, we conducted a contrast experiment with 31 dermatologists of different seniority levels, including professors, senior attending doctors, young attending doctors, and medical students of dermatology. In our contrast experiment, Xy-SkinNet and all dermatologists had to identify the type of disease from a set of 100 clinical images without a time limit. Their performances were evaluated using accuracy as the metric. To ensure the fairness and validity of the comparison, all the images used contained only one type of skin disease and had been previously verified using corresponding histopathological information. Histopathological examination was the golden standard for diagnoses of these six types of skin diseases. Compared with other diagnostic technologies, histopathological examination was more accurate and objective. If there were some cases whose histopathological features were atypical, we let another three specialists (who did not take part in human vs. computer competition) determine a final diagnosis. Images for this contrast experiment were selected by dermatologists who did not participate in the experiment, and the images were not included in the training or testing data sets used in the process of building of our model.

Results

Xiangya-Derm, a Chinese Dermatological Database

Xiangya-Derm, as a new Chinese database, contains nearly 20 years of data from 15 hospitals, including Xiangya Hospital Central South University, the Third Hospital of Central South University, and Henan Provincial People's Hospital. Most images were matched with pathology information to ensure the accuracy of the image diagnostic information. Other common skin diseases that are easily recognized by the naked eye were independently confirmed by three dermatologists. Each patient's clinical history was recorded on the Xiangya Big Data Collection Platform. It is worth mentioning that most of the data in the

Xiangya-Derm database have been labeled with bounding boxes by skin experts (Figure 2).

There are over 150,000 images in the Xiangya-Derm database, including 571 different skin diseases. Approximately, 60,000 images have been assigned annotations of skin type and lesion location. Common skin diseases, such as benign and malignant skin tumors, erythema papule scales, allergic skin diseases, and bullous skin diseases, are included (Figure 3). Of the total images in the database, 14,063 images are of skin tumors, including 5356 images of malignant tumors and 8707 images of benign tumors.

For common skin diseases, the diagnosis was independently made by three dermatologists based on history, clinical images, and dermoscopic features (Figure 2A). For rare skin diseases and skin tumors, the diagnosis was made based on history, clinical images, dermoscopic features, and histopathological features (Figure 2B). All the information about the patients and their diseases was collected in our database.

The database mainly consists of six types of skin diseases: malignant tumor, benign tumor, erythema papule scale, connective tissues, allergic, and bullous (Figure 3A). Representative examples in the Xiangya-Derm database are presented in Figure 3B, including psoriasis, melanoma, and dermatofibrosarcoma protuberans (an extremely rare disease). In addition to clinical images, dermascopic images, medical imaging information (eg, B-scan ultrasonography, positron emission tomography–computed tomography [PET-CT]) and history were included.

Figure 2. Pattern of the Xiangya Big Data Collection Platform.

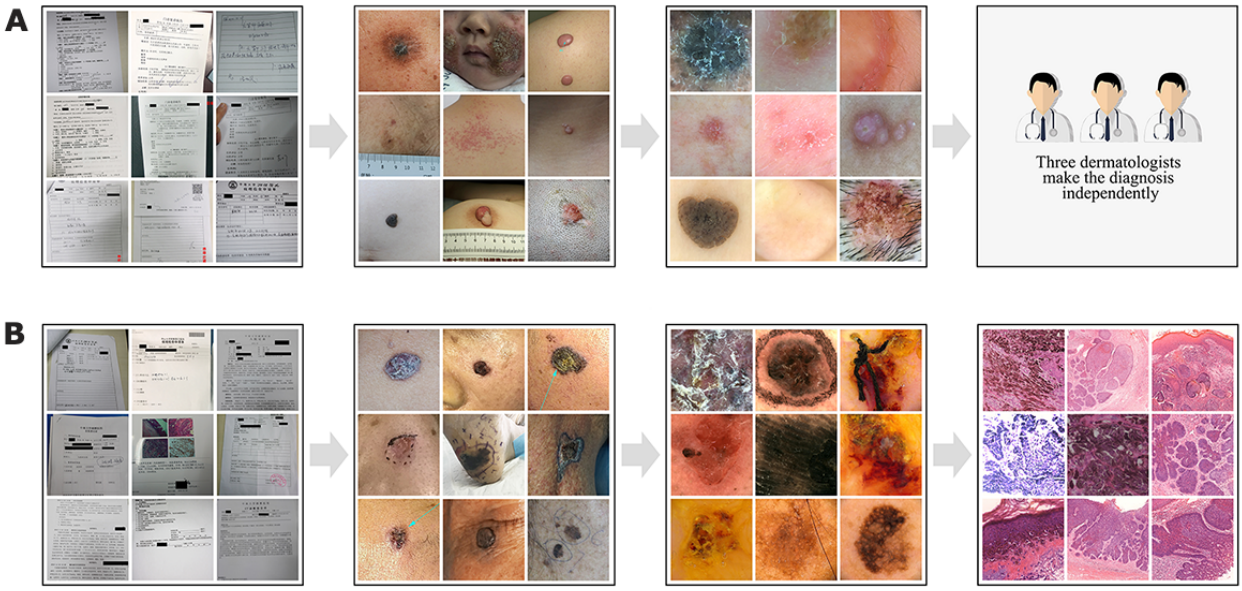
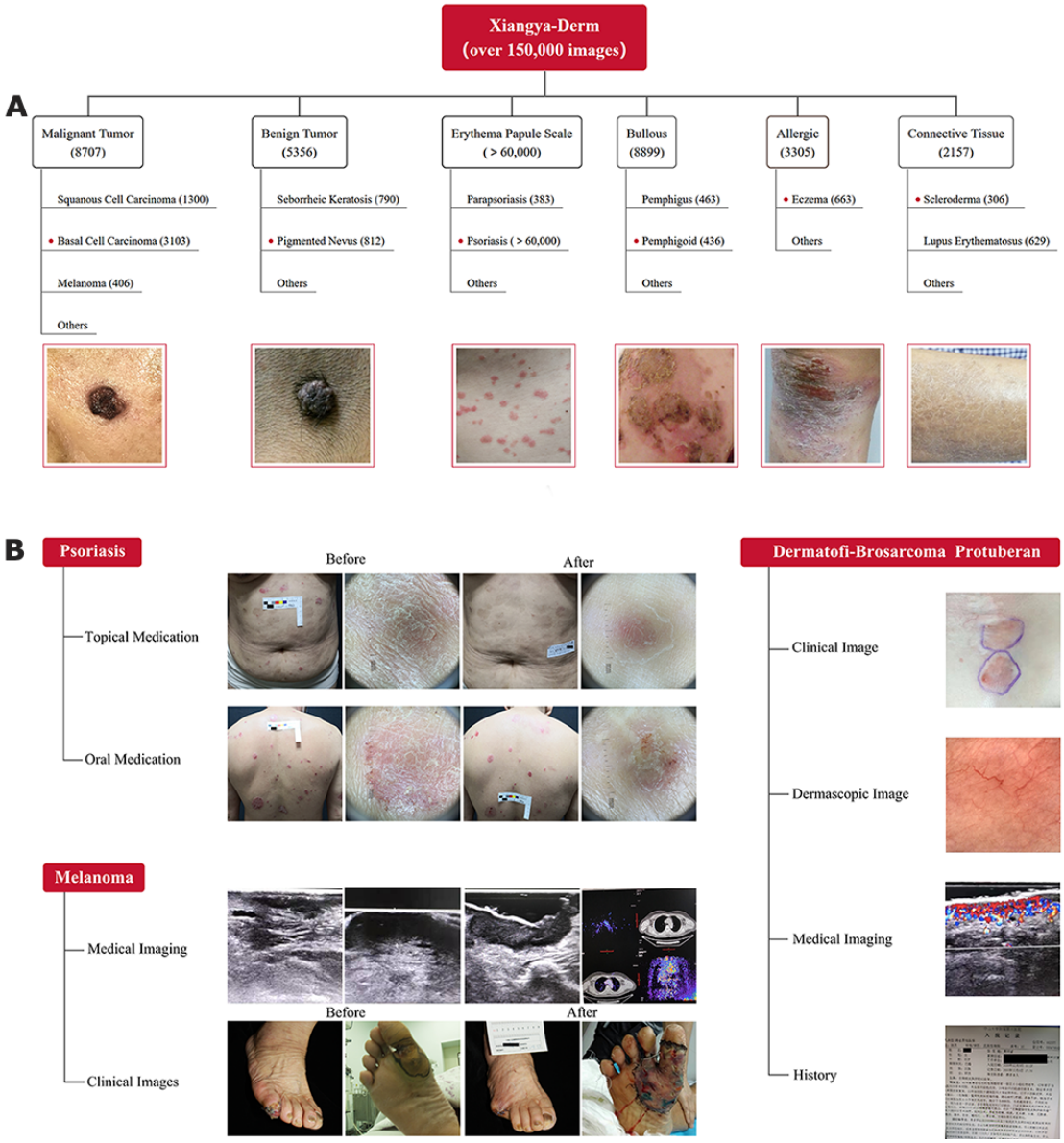


Figure 3. Branching tree diagram of the Xiangya-Derm database.



Xy-SkinNet Performance on Six Common Skin Diseases

The performance when using different CNN structures as the backbone model on the testing data set is shown in Table 1. The ResNet-152 model showed better performance than others in terms of top 1 and top 3 mean accuracies (61.48% and 77.38%, respectively). Xy-SkinNet's top differential diagnosis was for pemphigus and had a top 1 accuracy of 80.0%. When Xy-SkinNet was allowed three diagnoses, its top 3 accuracy rose to 90.0%. We also found that the performance of the ResNet-101 model was similar to that of the ResNet-152 model for certain diseases, such as seborrheic keratosis, pemphigus, and lupus erythematosus. However, the ResNet-101 model was better than the ResNet-152 model in terms of the top 1 accuracy on eczema (59.1% vs 54.5%) and the top 3 accuracies on lupus erythematosus (70.0% vs 60.0%).

The comparison between the experimental results of Xy-SkinNet and the average performance of dermatologists is shown in Table 2. The receiver operating characteristic (ROC) curve of the model for six types of diseases is shown in Figure 4, in which the performance of each dermatologist is also plotted. As the results showed (Table 2), Xy-SkinNet achieved a top 1 accuracy of 64.75%, while the average performance of the dermatologists was 62.13%. Moreover, Xy-SkinNet outperformed the dermatologists in terms of the top 3 accuracy, where Xy-SkinNet achieved an accuracy of 84.77%, while the dermatologists achieved an accuracy of 78.15%. As shown in Figure 4, for Xy-SkinNet, the area under the curve (AUC) values for the six types of diseases were 0.81, 0.85, 0.90, 0.96, 0.79, and 0.76, respectively, which were higher than the corresponding values of the average performance level of the dermatologists.

Moreover, we recorded the time costs for Xy-SkinNet and each dermatologist in our experiments. The results showed that Xy-SkinNet takes only 10 seconds to identify all 100 clinical images, while some dermatologists took over 2500 seconds or over 25 seconds for a single image. Even dermatologists with minimal time costs took up to 1216 seconds to classify the images. These results showed that the method adopted in this paper can achieve comparable performance in terms of accuracy

to that of dermatologists and outperform dermatologists in terms of time costs.

The comparison of accuracies (%) on six different skin diseases between Xy-SkinNet and the average of dermatologists with different levels in the external data set provided by dermatologists are shown in Table 2. Xy-SkinNet achieved a top 1 accuracy of 64.75%, while the average performance of the dermatologists was 62.13%. Moreover, Xy-SkinNet outperformed the dermatologists in terms of the top 3 accuracy, where Xy-SkinNet obtained an accuracy of 84.77%, while the average performance of the dermatologists was 78.15%.

Table 1. Classification accuracies (%) of Xy-SkinNet.[a]

Backbone	PSO[b] (%)	BCC[c] (%)	SEK[d] (%)	PEM[e] (%)	ECZ[f] (%)	LUE[g] (%)	Mean (%)
Top 1							
ResNet-50	61.9	20.0	60.0	70.0	45.5	40.0	49.57
ResNet-101	54.8	20.0	70.0	80.0	59.1	50.0	55.65
ResNet-152	69.4	40.0	75.0	80.0	54.5	50.0	61.48
Xy-SkinNet	64.1	50.0	70.0	75.0	83.9	45.5	64.75
Top 3							
ResNet-50	66.7	70.0	75.0	90.0	68.2	60.0	71.65
ResNet-101	61.9	40.0	80.0	90.0	72.7	70.0	69.10
ResNet-152	71.6	80.0	90.0	90.0	72.7	60.0	77.38
Xy-SkinNet	67.7	70.0	90.0	90.0	100.0	90.9	84.77

[a]The classification accuracies (%) of Xy-SkinNet on six different skin diseases using different backbone networks as the feature extractor are shown. The ResNet-152 model showed better performance than the ResNet-50 and ResNet-101 models in terms of top 1 and top 3 mean accuracies. However, our original algorithm performed better than the others, especially in terms of the top 3 accuracies.

[b]PSO: psoriasis.

[c]BCC: basal cell carcinoma.

[d]SEK: seborrheic keratosis.

[e]PEM: pemphigus.

[f]ECZ: eczema.

[g]LUE: lupus erythematosus.

Table 2. Comparison of accuracy (%) between Xy-SkinNet and the average of dermatologists.

Class	Sample size	Top 1 (computer; %)	Top 1 (dermatologists; %)	Top 3 (computer; %)	Top 3 (dermatologists; %)
LUE[a]	11	45.5	60.6	90.9	70.0
ECZ[b]	31	83.9	47.5	100.0	76.9
SEK[c]	10	70.0	60.8	90.0	77.9
PSO[d]	93	64.1	75.7	67.7	87.2
BCC[e]	10	50.0	67.4	70.0	84.9
PEM[f]	20	75.0	60.8	90.0	72.0
All	175	64.75	62.13	84.77	78.15

[a]LUE: lupus erythematosus.

[b]ECZ: eczema.

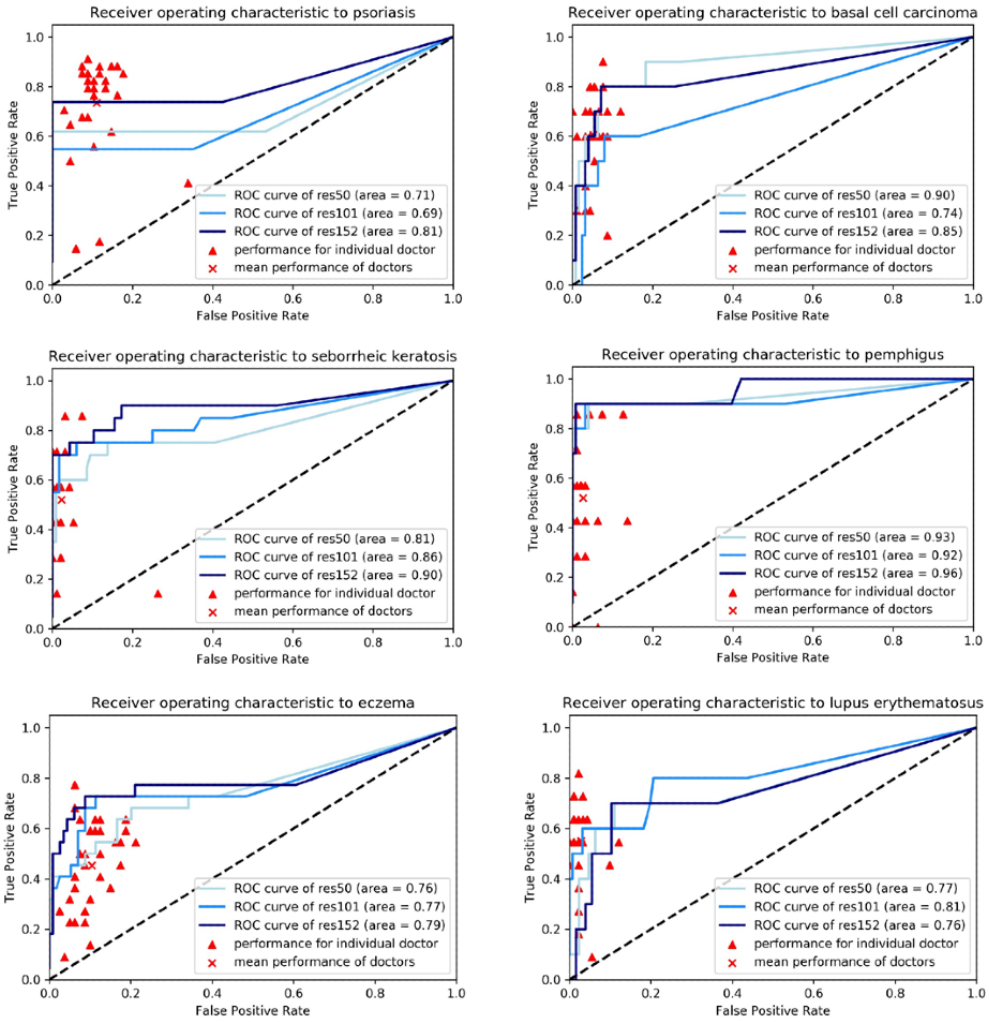
[c]SEK: seborrheic keratosis.

[d]PSO: psoriasis.

[e]BCC: basal cell carcinoma.

[f]PEM: pemphigus.

Figure 4. ROC curve of the six-category classification model. The three blue lines represent the mean ROC curves over the ResNet50, ResNet101, and ResNet152 runs. Each triangular point represents the results of one dermatologist. Results suggested that the mean performance of the dermatologists was lower than that of the models for most diseases. ROC: receiver operating characteristic.



Discussion

Principal Results

In this study, we established a new database, Xiangya-Derm, which consists of over 150,000 clinical images of 571 different skin diseases in the Chinese population. Xiang-Derm is the first integrated, normative database based on skin conditions in the Chinese population. Based on this database, we selected six common skin diseases and proposed an AI network, Xy-SkinNet. The top 1 and top 3 diagnostic accuracies of Xy-SkinNet were higher than those of dermatologists from the Department of Dermatology. This study was an attempt at exploring AI products and services and has successfully set the stage for future development. An increasing number of studies are incorporating clinical images [18-21].

There are already some open databases, such as AtlasDerm, Derm101, and Dermnet. Considering differences in skin color, Xiangya-Derm can provide data for realizing AI diagnosis of skin diseases among the Chinese population. Many existing databases lack medical history information, especially information about pathological diagnosis, and, potentially, contain some misdiagnosed photos. Notably, one of the greatest advantages of Xiangya-Derm is that most images contain corresponding skin pathology results, providing the category annotation of a gold standard, which can be most effectively applied to various research studies and in the development of AI. This feature ensures that the diagnostic information about pictures used for deep learning is accurate and reduces the diagnostic errors caused by misdiagnosis. Of course, there are also a small number of unmatched pictures in our database, which is correlated with the lack of corresponding dermoscopic images.

In addition, XiangyaDerm provides image data with the location for all skin lesions, thus enabling researchers to apply object detection algorithms in computer vision for the automatic diagnosis of skin diseases. Moreover, each image has a full set of clinical information about the patient, including demographic information, complaints, current medical history, past medical history, and family history. Given the complete set of big data, conducting further research on AI diagnosis using multimodal data, which is more coincident with the real-world diagnosis process and more intuitive for both doctors and patients, is achievable.

Comparison With Prior Work

Many attempts have been made in the field of automatic diagnosis of dermatosis. For example, Alćonet et al [22] proposed a system that consisted of a feature extractor and a classifier. This system is capable of adapting its decision-making process based on patient information. Blum et al [23] performed digital image analysis using 64 different analytical parameters based on 837 cases of melanocyte lesions to establish a computer algorithm for diagnosing melanocyte lesions. The diagnostic accuracies of the algorithm for complete and partial imaging were 82% and 84%, respectively. Another study used multiple databases to perform dermoscopic and digital dermoscopic examinations of cutaneous melanoma and compared the diagnostic accuracies of different dermal algorithms and digital dermatoscopy joint intelligence [24].

CNNs have presented varying results in medical fields when using an improved algorithm [8,25,26]. Pal et al [10] designed and produced a psoriasis skin biopsy image data set of 90 images and attempted to segment the images using several different neural network models. Their results were superior to the manual extraction features. Sun et al [9] attempted to classify dermatological images using CNNs as feature extraction methods. They used a data set that included 198 types of diseases and 6584 images. However, despite using a large data set, the final result was inferior to that obtained by the use of AI extraction features. Han et al [27] used a deep learning algorithm to classify clinical images of 12 dermatoses, with performance comparable to that of 16 dermatologists. Tschandl et al [28] showed that computers can achieve an accuracy similar to that of medical students by using a CNN-based method to diagnose skin diseases on an image. In our study, the AI-based six-category classification outperformed the dermatologists. Patients' multimodal data were used to create a simulated clinical setting for AI. Thus, the comparison between AI and the dermatologists is fair as the diagnosis made by dermatologists in real clinical practice partially depends on history and other nonimage information.

Relevant research reports also increasingly support that deep learning systems can help improve the accuracy of nondermatologists or clinicians with a lack of experience, especially in primary care [1]. AI can also foster and improve the confidence of doctors. In China, 95% of dermatologists are concentrated in tertiary hospitals, and the number of dermatologists at county hospitals, especially township hospitals, is less than 5% of the total number. In addition, numerous county and township hospitals have a limited number of dermatologists. The proportion of doctors and patients is unbalanced, and the level of doctors is uneven, causing delayed diagnosis for many patients, poor prognosis, aggravation of the family economic burden, etc [29]. In our research, we selected six skin diseases: psoriasis, basal cell carcinoma, seborrheic keratosis, pemphigus, eczema, and lupus erythematosus. Our research may satisfy the needs of primary doctors, as these six conditions are common in primary care and outpatient work. In the future, we intend to explore an auxiliary diagnosis system that can provide dermatologists with three possible diagnoses, ranked by accuracy, for one lesion. The risk of misdiagnosis is likely to decline for primary diagnoses when combining AI and manual correction.

Because there are few dermatologists in China's rural areas, we plan to build a multidimensional AI platform that assists Chinese primary dermatologists in diagnosing skin diseases through deep learning and the construction of Xiangya-Derm. In light of the lack of criteria for skin disease data sets, we would like to encourage the establishment of standards in the future. It would be beneficial to create a new data set or improve the existing data sets with multimodal clinical information.

Limitations

Although we included six types of skin diseases in our data set, a large number of skin diseases were not considered. One of the reasons for this is that these six diseases cover most of the skin conditions in clinical practice. With this database and the development of AI products, new data for rare skin diseases can be collected at the same time as the diagnosis of common skin diseases. In addition, the quality of health services has improved in China, and some skin diseases, such as bacterial and parasitic skin diseases, are rarely seen. Thus, it is difficult to collect data on these diseases. Moreover, the clinical images in our database were collected from the Chinese population, making the database ineffective for training an intelligent global skin disease diagnosis system.

Conclusions

Xiangya-Derm, the largest database for skin diseases in the Chinese population, was created, and six common skin conditions were classified based on Xiangya-Derm to lay a foundation for product research.

Acknowledgments

This work is supported by the National Key R&D Program of China (2018YFC0117004); the Outstanding Youth Project of Hunan, China (2020RC3059); the Hunan Provincial Innovation Foundation for Postgraduate; and the Fundamental Research Funds for the Central University of Central South University.

Conflicts of Interest

None declared.

References

1. Liu Y, Jain A, Eng C, Way DH, Lee K, Bui P, et al. A deep learning system for differential diagnosis of skin diseases. Nat Med 2020 May 18;26(6):900-908. [doi: 10.1038/s41591-020-0842-3]
2. Pangti R, Mathur J, Chouhan V, Kumar S, Rajput L, Shah S, et al. A machine learning-based, decision support, mobile phone application for diagnosis of common dermatological diseases. J Eur Acad Dermatol Venereol 2020 Nov 12;35(2):536-545. [doi: 10.1111/jdv.16967]
3. Hollestein LM, Nijsten T. An insight into the global burden of skin diseases. Journal of Investigative Dermatology 2014 Jun;134(6):1499-1501. [doi: 10.1038/jid.2013.513]
4. Karimkhani C, Dellavalle RP, Coffeng LE, Flohr C, Hay RJ, Langan SM, et al. Global skin disease morbidity and mortality: an update from the Global Burden of Disease study 2013. JAMA Dermatol 2017 May 01;153(5):406-412 [FREE Full text] [doi: 10.1001/jamadermatol.2016.5538] [Medline: 28249066]
5. Li CX, Shen CB, Xue K, Shen X, Jing Y, Wang ZY, et al. Artificial intelligence in dermatology: past, present, and future. Chin Med J (Engl) 2019 Sep 05;132(17):2017-2020 [FREE Full text] [doi: 10.1097/CM9.0000000000000372] [Medline: 31425274]
6. Hashmi S. 'Coming of Age' of artificial intelligence: evolution of survivorship care through information technology. Bone Marrow Transplant 2015 Nov 9;51(1):41-42. [doi: 10.1038/bmt.2015.271]
7. Naylor CD. On the prospects for a (deep) learning health care system. JAMA 2018 Sep 18;320(11):1099. [doi: 10.1001/jama.2018.11103]
8. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Jan 25;542(7639):115-118. [doi: 10.1038/nature21056]
9. Sun X, Yang J, Sun M, Wang K. A benchmark for automatic visual classification of clinical skin disease images, in European Conference on Computer Vision. 2016 Sep 17 Presented at: European Conference on Computer Vision; October 8-16, 2016; Amsterdam p. 2016. [doi: 10.1007/978-3-319-46466-4_13]
10. Pal A, Garain U, Chandra A, Chatterjee R, Senapati S. Psoriasis skin biopsy image segmentation using deep convolutional neural network. Computer Methods and Programs in Biomedicine 2018 Jun;159:59-69. [doi: 10.1016/j.cmpb.2018.01.027]
11. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 2017 Jun 1;39(6):1137-1149. [doi: 10.1109/TPAMI.2016.2577031] [Medline: 27295650]
12. Brinker TJ, Hekler A, Utikal JS, Grabe N, Schadendorf D, Klode J, et al. Skin cancer classification using convolutional neural networks: systematic review. J Med Internet Res 2018 Oct 17;20(10):e11936. [doi: 10.2196/11936]
13. Sies K, Winkler JK, Fink C, Bardehle F, Toberer F, Buhl T, et al. Past and present of computer-assisted dermoscopic diagnosis: performance of a conventional image analyser versus a convolutional neural network in a prospective data set of 1,981 skin lesions. European Journal of Cancer 2020 Aug;135:39-46. [doi: 10.1016/j.ejca.2020.04.043]
14. LabelImg. URL: https://tzutalin.github.io/labelImg/ [accessed 2021-08-08]
15. Fujisawa Y, Otomo Y, Ogata Y, Nakamura Y, Fujita R, Ishitsuka Y, et al. Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis. Br J Dermatol 2019 Feb;180(2):373-381. [doi: 10.1111/bjd.16924] [Medline: 29953582]
16. Ross G, Jeff D, Trevor D, Jitendra M. Rich feature hierachies for accurate object detection and semantic segmentation. : IEEE; 2014 Sep 25 Presented at: 2014 IEEE Conference on Computer Vision and Pattern Recognition; Columbus, OH, USA; 23-28 June 2014 p. 2014. [doi: 10.1109/CVPR.2014.81]
17. Ross G. Fast R-CNN, in Proceedings of the IEEE International Conference on Computer Vision. 2015 Sep 07 Presented at: 2015 IEEE International Conference on Computer Vision (ICCV); Dec. 7-13, 2015; Santiago p. 1440-1448. [doi: 10.1109/ICCV.2015.169]
18. Xiaoxiao S, Jufeng Y, Ming S, Kai W. A benchmark for automatic visual classification of clinical skin disease images. 2016 Sep 17 Presented at: European Conference on Computer Vision; October 11-14, 2016,; Amsterdam, The Netherlands. [doi: 10.1007/978-3-319-46466-4_13]
19. Haofu L, Yuncheng L, Jiebo L. Skin disease classification versus skin lesion characterization: achieving robust diagnosis using multi-label deep neural networks. 2017 Apr 24 Presented at: 2016 23rd International Conference on Pattern Recognition (ICPR); 4-8 Dec. 2016; Cancun, Mexico. [doi: 10.1109/icpr.2016.7899659]
20. Haofu L, Jiebo L. A deep learning approach to universal skin disease classification. AI Magazine 2018 Sep 18:2015.
21. Jufeng Y, Xiaoxiao S, Jie L, Paul L. R. Clinical skin lesion diagnosis using representations inspired by dermatologist criteria. 2018 Dec 17 Presented at: J.L., Clinical Skin Lesion Diagnosis using Representations Inspired by Dermatologist Criteria, in IEEE Conference on Computer Vision and Pattern Recognition; 2018; Salt Lake City, UT, USA. [doi: 10.1109/cvpr.2018.00137]
22. Fernandez Alcon J, Ciuhu C, ten Kate W, Heinrich A, Uzunbajakava N, Krekels G, et al. Automatic imaging system with decision support for inspection of pigmented skin lesions and melanoma diagnosis. IEEE J. Sel. Top. Signal Process 2009 Feb;3(1):14-25. [doi: 10.1109/jstsp.2008.2011156]
23. Blum A, Luedtke H, Ellwanger U, Schwabe R, Rassner G, Garbe C. Digital image analysis for diagnosis of cutaneous melanoma. Development of a highly effective computer algorithm based on analysis of 837 melanocytic lesions. Br J Dermatol 2004 Nov;151(5):1029-1038. [doi: 10.1111/j.1365-2133.2004.06210.x]

24. Rajpara S, Botello A, Townend J, Ormerod A. Systematic review of dermoscopy and digital dermoscopy/artificial intelligence for the diagnosis of melanoma. Br J Dermatol 2009 Aug 20;161(3):591-604. [doi: 10.1111/j.1365-2133.2009.09093.x]
25. Alper A, Mesut G, Elif Mine K. Deep learning based skin cancer diagnosis. 2017 Jun 29 Presented at: 2017 25th Signal Processing and Communications Applications Conference (SIU); 15-18 May 2017; Antalya, Turkey. [doi: 10.1109/SIU.2017.7960452]
26. Alper B, Mehmet Emin Y, Hasan B, Abdullah Ç. Deep neural network based diagnosis system for melanoma skin cancer. 2017 Jun 29 Presented at: 2017 25th Signal Processing and Communications Applications Conference (SIU); 15-18 May 2017; Antalya, Turkey p. 2017. [doi: 10.1186/s40537-019-0276-2]
27. Han SS, Kim MS, Lim W, Park GH, Park I, Chang SE. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. Journal of Investigative Dermatology 2018 Jul;138(7):1529-1538. [doi: 10.1016/j.jid.2018.01.028]
28. Tschandl P, Kittler H, Argenziano G. A pretrained neural network shows similar diagnostic accuracy to medical students in categorizing dermatoscopic images after comparable training conditions. Br J Dermatol 2017 Jul 19;177(3):867-869. [doi: 10.1111/bjd.15695]
29. Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, et al. Human–computer collaboration for skin cancer recognition. Nat Med 2020 Jun 22;26(8):1229-1234. [doi: 10.1038/s41591-020-0942-0]

Abbreviations

AI: artificial intelligence
AUC: area under the curve
CNN: convolutional neural network
EU: European Union
M-ResNet: modified ResNet
PET-CT: positron emission tomography–computed tomography
ROC: receiver operating characteristic
RPN: region proposal network

Edited by R Kukafka; submitted 25.01.21; peer-reviewed by NB Prakash, B Li; comments to author 17.03.21; revised version received 24.03.21; accepted 27.07.21; published 21.09.21.

Please cite as:
Huang K, Jiang Z, Li Y, Wu Z, Wu X, Zhu W, Chen M, Zhang Y, Zuo K, Li Y, Yu N, Liu S, Huang X, Su J, Yin M, Qian B, Wang X, Chen X, Zhao S
The Classification of Six Common Skin Diseases Based on Xiangya-Derm: Development of a Chinese Database for Artificial Intelligence
J Med Internet Res 2021;23(9):e26025
URL: https://www.jmir.org/2021/9/e26025
doi:10.2196/26025
PMID:34546174

Original Paper

A Fully Automated Analytic System for Measuring Endolymphatic Hydrops Ratios in Patients With Ménière Disease via Magnetic Resonance Imaging: Deep Learning Model Development Study

Chae Jung Park[1*], MSc; Young Sang Cho[2*], MD; Myung Jin Chung[3], MD, PhD; Yi-Kyung Kim[3], MD; Hyung-Jin Kim[3], MD, PhD; Kyunga Kim[1,4], PhD; Jae-Wook Ko[5], MD, PhD; Won-Ho Chung[2*], MD, PhD; Baek Hwan Cho[6], PhD

[1]Department of Digital Health, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul, Republic of Korea

[2]Department of Otorhinolaryngology-Head and Neck Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

[3]Department of Radiology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

[4]Biomedical Statistics Center, Data Science Research Institute, Research Institute for Future Medicine, Samsung Medical Center, Seoul, Republic of Korea

[5]Department of Clinical Pharmacology and Therapeutics, Samsung Medical Center, Seoul, Republic of Korea

[6]Department of Medical Device Management and Research, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Baek Hwan Cho, PhD
Department of Medical Device Management and Research
Samsung Advanced Institute for Health Sciences & Technology
Sungkyunkwan University
81 Irwon-ro, Gangnam-gu
Seoul, 06355
Republic of Korea
Phone: 82 234100885
Email: baekhwan.cho@samsung.com

Abstract

Background: Recently, the analysis of endolymphatic hydropses (EHs) via inner ear magnetic resonance imaging (MRI) for patients with Ménière disease has been attempted in various studies. In addition, artificial intelligence has rapidly been incorporated into the medical field. In our previous studies, an automated algorithm for EH analysis was developed by using a convolutional neural network. However, several limitations existed, and further studies were conducted to compensate for these limitations.

Objective: The aim of this study is to develop a fully automated analytic system for measuring EH ratios that enhances EH analysis accuracy and clinical usability when studying Ménière disease via MRI.

Methods: We proposed the 3into3Inception and 3intoUNet networks. Their network architectures were based on those of the Inception-v3 and U-Net networks, respectively. The developed networks were trained for inner ear segmentation by using the magnetic resonance images of 124 people and were embedded in a new, automated EH analysis system—inner-ear hydrops estimation via artificial intelligence (INHEARIT)-version 2 (INHEARIT-v2). After fivefold cross-validation, an additional test was performed by using 60 new, unseen magnetic resonance images to evaluate the performance of our system. The INHEARIT-v2 system has a new function that automatically selects representative images from a full MRI stack.

Results: The average segmentation performance of the fivefold cross-validation was measured via the intersection of union method, resulting in performance values of 0.743 (SD 0.030) for the 3into3Inception network and 0.811 (SD 0.032) for the 3intoUNet network. The representative magnetic resonance slices (ie, from a data set of unseen magnetic resonance images) that were automatically selected by the INHEARIT-v2 system only differed from a maximum of 2 expert-selected slices. After comparing the ratios calculated by experienced physicians and those calculated by the INHEARIT-v2 system, we found that the average intraclass correlation coefficient for all cases was 0.941; the average intraclass correlation coefficient of the vestibules

was 0.968, and that of the cochleae was 0.914. The time required for the fully automated system to accurately analyze EH ratios based on a patient's MRI stack was approximately 3.5 seconds.

Conclusions: In this study, a fully automated full-stack magnetic resonance analysis system for measuring EH ratios was developed (named INHEARIT-v2), and the results showed that there was a high correlation between the expert-calculated EH ratio values and those calculated by the INHEARIT-v2 system. The system is an upgraded version of the INHEARIT system; it has higher segmentation performance and automatically selects representative images from an MRI stack. The new model can help clinicians by providing objective analysis results and reducing the workload for interpreting magnetic resonance images.

(J Med Internet Res 2021;23(9):e29678) doi:10.2196/29678

KEYWORDS

deep learning; magnetic resonance imaging; medical image segmentation; Ménière disease; inner ear; endolymphatic hydrops; artificial intelligence; machine learning; multi-class segmentation; convolutional neural network; end-to-end system; clinician support; clinical decision support system; image selection; clinical usability; automation

Introduction

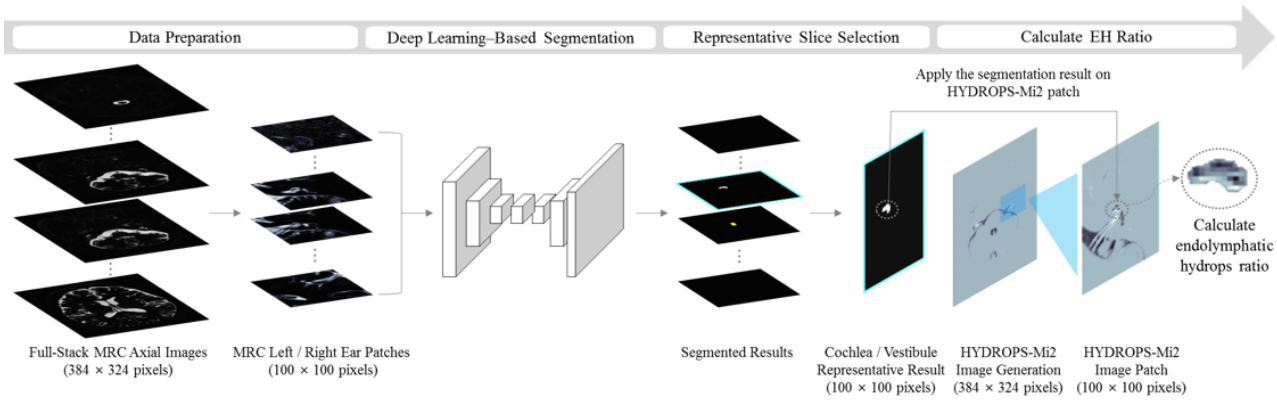
Although many medical fields have been developed over the past few decades, medical imaging techniques, such as computed tomography and magnetic resonance imaging (MRI), have advanced greatly [1]. Experienced radiologists and physicians typically interpret such images in the clinical field. However, in recent years, due to the remarkable development of machine learning, the situation is changing [2]. Artificial intelligence, including machine learning, is widely used in various fields of medical science, and image analysis via a convolutional neural network is developing rapidly [3].

Ménière disease is a multifactorial disorder with typical symptoms, such as recurrent vertigo attacks, fluctuating hearing loss, tinnitus, and sensations of ear fullness. The prevalence of Ménière disease varies by region and study, but the estimated prevalence ranges from 30 to 150 patients per 100,000 people [4]. In particular, it is known that this prevalence is higher within White and female populations and increases with age [5]. Endolymphatic hydrops (EH) is a histologic hallmark of Ménière disease in which the endolymphatic spaces in the cochlea and the inner ear vestibule are distended [6]. According to current diagnostic criteria, pure tone audiometry is the only objective test for the diagnosis of definite or probable Ménière disease. Further, electrocochleography is a common test for estimating EH ratios [7]. However, electrocochleography is used only as a reference examination for diagnosing Ménière disease because it does not directly show the endolymphatic space. Similarly, Ménière disease is challenging to diagnose objectively, and efforts have been made in recent years to directly measure EH ratios by using MRI [8-10].

A protocol for image-based EH analysis was suggested in a previous study [11]; it required specific image viewer software to generate a hybrid image of signals (ie, a hybrid of the reversed image of the positive endolymph signal and native image of the positive perilymph signal [HYDROPS] or a HYDROPS image multiplied by T_2-weighted magnetic resonance cisternography [HYDROPS-Mi2]). This process involves the manual contouring of inner ear organs for boundary segmentation, which is mostly performed by medical experts.

The need for automated analyses has emerged because conventional quantitative analyses require more time and effort than typical image interpretation processes. The automatic measurement of EH ratios via MRI was proposed based on the deep learning approach in our previous study [12]. Our research showed that the convolutional neural network–based deep learning model—the inner-ear hydrops estimation via artificial intelligence (INHEARIT) system—could efficiently segment cochleae and vestibules in magnetic resonance images and calculate the EH ratios of the segmented regions [12]. However, our study had a few limitations. First, full-stack image validation was not conducted in our previous study. Thus, medical experts were needed to manually select representative image slices of cochleae and vestibules from the full magnetic resonance image stack and load them into the system, and this human user process was time consuming. Second, validation with an isolated data set was not performed, which made it difficult to verify the robustness of the system. Lastly, various deep learning models were not used, except for the Visual Geometry Group (VGG) network architecture—a VGG-19–based network [13]. To compensate for these limitations, we developed a fully automated analytic system for calculating EH ratios by using deep learning and MRI—the INHEARIT-version 2 (INHEARIT-v2) system. The entire framework for this analytic system is depicted in Figure 1.

Figure 1. The proposed inner-ear hydrops estimation via artificial intelligence-version 2 (INHEARIT-v2) framework. EH: endolymphatic hydrops; HYDROPS-Mi2: hybrid of the reversed image of the positive endolymph signal and native image of the positive perilymph signal image multiplied by T_2-weighted magnetic resonance cisternography; MRC: magnetic resonance cisternography.



Methods

Study Participants

Two data sets were used in this study—one for the deep learning–based training models and another for validation. For the training models, the magnetic resonance images of 124 patients (males: n=57; females: n=67; age: mean 49.3 years, SD 16.02 years; range 17-76 years) who participated in our previous INHEARIT study were used [12]. For the additional test, 60 new participants were recruited from Samsung Medical Center from February 2018 to September 2019. Written informed consent was obtained from all participants. This study was approved by the Institutional Review Board of Samsung Medical Center (approval number: 2020-06-046).

Data Set for Analysis

We scanned all patients' inner ears with a 3.0-T MRI device (MAGNETOM Skyra; Siemens Medical Solutions). Gadobutrol (gadolinium-DO3A-butriol; Gadovist 1.0), which was used as the contrast agent, was intravenously administered to patients before MRI. Different magnetic resonance scans were made by using heavily T_2-weighted magnetic resonance cisternography (MRC) to distinguish the outlines of organs. Positive perilymph images (PPIs) and positive endolymph images (PEIs) were used to evaluate EHs. The HYDROPS images were obtained by subtracting the number of PEIs from the number of PPIs. MRC images, PPIs, and PEIs had identical fields of view, matrix sizes, and slice thicknesses [8,14].

Data Annotation by Physicians

One neuroradiologist and one neuro-otologist independently evaluated the MRC images. Representative slices of 4 classes—the left cochlea and vestibule and the right cochlea and vestibule—were selected, and their contours were annotated manually. The main slices of a cochlea or vestibule and their regions of interest (ROIs) were chosen based on the following criteria. For the cochlear ROI, among all MRI slices in which the basal, middle, and apical views of organs were visible, the slice with the largest modiolus height was chosen. For the vestibular ROI, the lowest slice, in which the lateral semicircular canal ring was obvious in images that were rotated by more than 240°, was chosen, and images of the ampulla were excluded [12]. Each physician drew contours of the ROIs on MRC images, and the contours were compared to determine a point of agreement.

Deep Learning Training Models for Inner Ear Segmentation

To segment the ROIs, 2 deep learning models were developed based on the architectures of the Inception-v3 [15] and U-Net [16] networks. The models were named 3into3Inception and 3intoUNet. For segmentation via the Inception-v3 network, deconvolutional layers were added after the conventional feature extraction of convolutional layers, and skip connections were included between convolutional and corresponding deconvolutional layers, as we did in our previous study [12]. The purpose of the U-Net network is to conduct segmentation; this network consists of feature contraction and feature expansion parts that function as encoders and decoders, respectively. In this network, a portion of contracted information is copied and concatenated to the corresponding expansion part, thereby reducing the amount of information that is lost during the segmentation process.

The physician-selected main slice image and the previous and next images were loaded simultaneously from full-stack MRC images for model training. The cross-sectional area of target inner ear structures is relatively small in a whole-brain image. Since this can cause a class imbalance problem, image patches of 100×100 pixels were acquired from each left and right reference point. This process is shown with the 3intoUNet network in Multimedia Appendix 1; 3 sequential MRC image patches were independently fed into each of the networks, and features were summated before the addition of deconvolutional layers. The green dotted boxes represent parts of the network that perform the same function (ie, the contraction of features in each input image). This sequential 3-input approach has been shown to yield high performance in terms of medical image segmentation [12]. The implementation and analyses of these models were performed via the Python 3.5 (Python Software Foundation) environment. The NumPy library was used for arithmetic calculation, the sklearn and ImageIO libraries were used for image preprocessing, and the TensorFlow library was used for model training.

The developed models were trained with a selectively annotated data set of 124 subjects. Afterward, the models were fine-tuned with both the selectively annotated and fully annotated data sets, per the optimum curriculum learning strategy for the segmentation of weakly annotated data [12]. Although we analyzed 4 target organs (the left cochlea, left vestibule, right cochlea, and right vestibule), the number of representative image slices per subject varied between 2 and 4 because a cochlea and a vestibule from each side could have been in the same image slice depending on the anatomical structure of the person. First, we trained a model with the selectively annotated data set and conducted moderate augmentation (1584 times). Afterward, we fine-tuned the model with both the fully annotated and selectively annotated data sets and conducted high augmentation (14,544 times). Therefore, the original 262 slices from the selectively annotated data set were augmented to 412,008 slices, and the initial 372 slices from the fully annotated and selectively annotated data sets were augmented to 5,410,368 slices for training. For augmentation, image patches were randomly flipped along the horizontal direction, cropped via random shifting (ie, from a reference point), and had their pixel intensity changed. There was a wide range of variation between the augmented images and the original image (eg, differences in intensity, shifted cropping areas, etc), and the physicians agreed to use the augmented images as training inputs.

Fivefold cross-validation was conducted, and segmentation performance was evaluated by measuring the intersection over union (IoU) between the ground truth areas (clinician-annotated region) and prediction areas (model-based, automatically determined region). The IoU was calculated as follows:

$$\text{IoU} = \text{area of overlap/area of union} = A_{overlap}/(A_{GT} + A_{pred} - A_{overlap}) \quad \textbf{(1)}$$

In equation 1, A_{GT} is the ground truth area, A_{pred} is the prediction area, and $A_{overlap}$ is the intersection between A_{GT} and A_{pred}. The model was trained on graphical processing units (NVIDIA GTX 1080Ti; Nvidia Corporation). The parameters were determined via grid searching and optimized to a learning rate of $1e^{-5}$ with the Adam optimizer [17]; a dropout rate [18] of 0.4 and batch size of 4 were used. Batch normalization and mean subtractions were performed to prevent internal covariate shifts.

Full-Stack Image Segmentation of an Additional Test Data Set

Once model training was completed, the model was tested with an additional data set of unseen magnetic resonance images. As shown in Figure 1, the full-stack MRC images of a patient were fed into a deep learning network. Slice indices that represented cochleae and vestibules were selected from a stack of segmentation results. The selected results were applied to HYDROPS-Mi2 patches as masks, and EH ratios were calculated for segmented regions.

Figure 2 shows the process of conducting a full-stack magnetic resonance image segmentation analysis by using the test data set. A subject's full-stack images were given an index number that ranged from 1 to the total number of slices (N). Figure 2 shows a subject's full-stack MRC image patches for the left and right ears and their corresponding segmentation results. The INHEARIT-v2 platform–selected representative slice for each class was indicated by a cyan-colored slice boundary. Figure 2 also shows the segmentation results and the MRC patches for a selected index number and the ground truths of each class. In this example, the selected index number was identical to the ground truths of all classes. The system chose the representative slices for each class—the left cochlea, left vestibule, right cochlea, and right vestibule—based on the size of the segmented area in each class. The representative key slice selection process was formulated as follows:

In equation 2, i is a slice index, N is the total number of segmented images of a subject, s^i_c represents a segmented area in the *ith* slice in class C, I_C is the key slice index of class C, and S_c is the segmented region in the key slice in class C. Thus, the key slices with the largest segmented areas for each class were selected. The slices chosen by the INHEARIT-v2 system and human experts were compared, and the index distance between the slices was calculated. The chosen segmentation results were used as masks for EH ratio calculation in the next step.

The EH ratio was estimated by using a HYDROPS-Mi2 image, which is an image that is generated via the pixel-wise multiplication of HYDROPS and MRC image signals [11] for a given index slice. The EH ratio is defined as follows [12]:

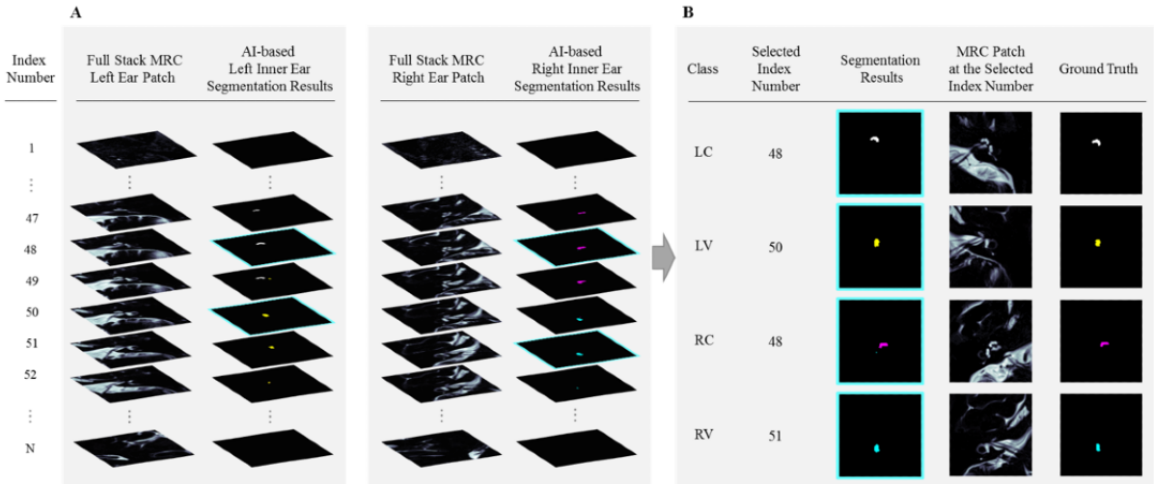
$$\text{EH ratio} = \text{Total number of pixels with a negative value in the segmentation area/total number of pixels in the segmentation area} \quad \textbf{(3)}$$

The above equation (equation 3) can be restated as follows:

$$\text{EH ratio} = (P_C^{Seg} \cap P_C^{Neg})/P_C^{Seg} \quad \textbf{(4)}$$

In equation 4, P_C^{Seg} denotes the total number of pixels in the segmented area in class C, and P_C^{Neg} denotes the total number of pixels with negative values in class C.

Figure 2. The full-stack magnetic resonance image segmentation analysis process for the test data set. A: A full stack of MRC images and their segmentation results. B: Representative slices and the segmentation results of each class. AI: artificial intelligence; MRC: magnetic resonance cisternography.



Statistical Analysis

The agreement between the values calculated by experts (the neuroradiologist and the neuro-otologist) and the INHEARIT-v2–predicted values was measured by using a single-score intraclass correlation coefficient (ICC), which was based on a two-way model and Pearson correlation coefficient. The analyses were performed by using R software (The R Foundation) [19].

Results

Model Experiments

Table 1 shows the segmentation performance of the developed artificial intelligence–based models—the 3into3VGG, 3into3Inception, and 3intoUNet networks. Average performance values were represented as the average IoU values from the fivefold cross-validation with SDs, which were 0.761 (SD 0.036) for 3into3VGG, 0.784 (0.025) for 3into3Inception, and 0.811 (0.032) for 3intoUNet. The U-Net–based model, 3intoUNet, had the highest mean area under the receiver operating characteristic curve. Therefore, the 3intoUNet model was used for the additional test.

Table 1. The segmentation performance of the proposed networks—3into3VGG, 3into3Inception, and 3intoUNet. These networks were based on the Visual Geometry Group-19, Inception-v3, and U-Net networks, respectively. The results are represented as the average intersection over union values from the fivefold cross-validation with SDs.

Networks	IoU,[a] mean (SD)
3into3VGG	0.761 (0.036)
3into3Inception	0.784 (0.025)
3intoUNet	0.811 (0.032)

[a]IoU: intersection over union.

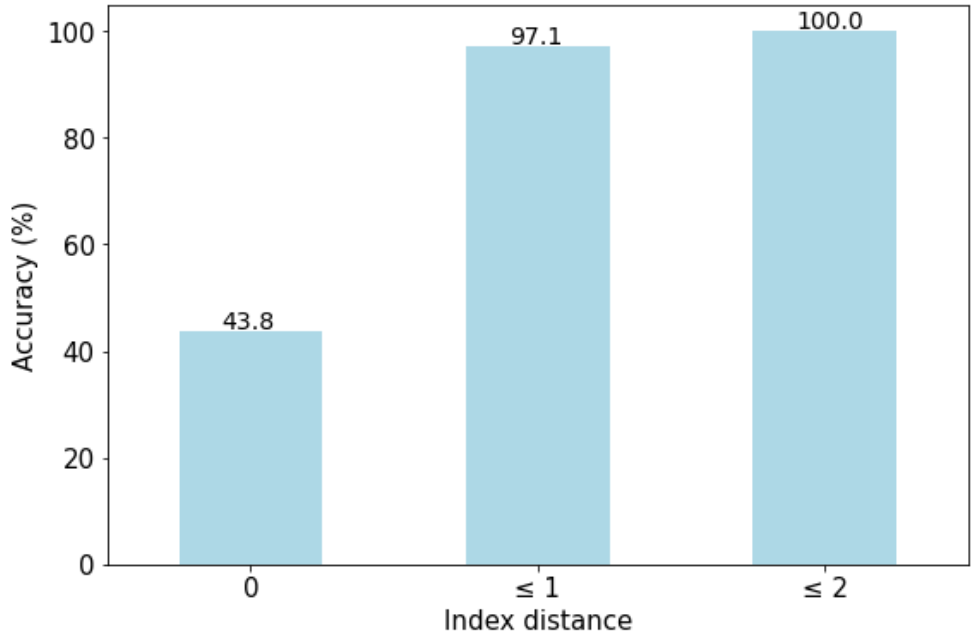
Additional Test With the Full-Stack Image Data Set

The data set of magnetic resonance images that were collected from 60 new participants (males: n=19; females: n=41; age: mean, 47.1 years, SD 15.27 years; range 21-68 years) for an additional test consisted of 33 subjects with definite Ménière disease (unilateral or bilateral), 17 subjects with sensorineural hearing loss and vertigo, and 10 normal subjects without any symptoms.

System validation was performed on the full-stack images of the subjects. Each image consisted of 104 MRC axial-view image slices, and 3 sequential slices—the main axial-view image slices and the previous and next slices—were fed into the trained model as an input. The model automatically segmented the organs by simultaneously analyzing the three input images and generated a segmentation result as an output. A total of 102 segmented output images were acquired from a subject's stack, and 1 image from each class was selected as a representative result.

With regard to the total 240 image slices of the target organs of 60 subjects, the distance (ie, the number of slice indices) between the manually selected image slices and system-selected image slices was 0 for 105 cases (43.8%), 1 in 233 cases (97.1%), and 2 in 240 cases (100%), as shown by the graph in Figure 3. This means that the representative magnetic resonance slices (ie, from a data set of unseen magnetic resonance images) that were automatically selected by the INHEARIT-v2 system only differed from a maximum of 2 expert-selected slices.

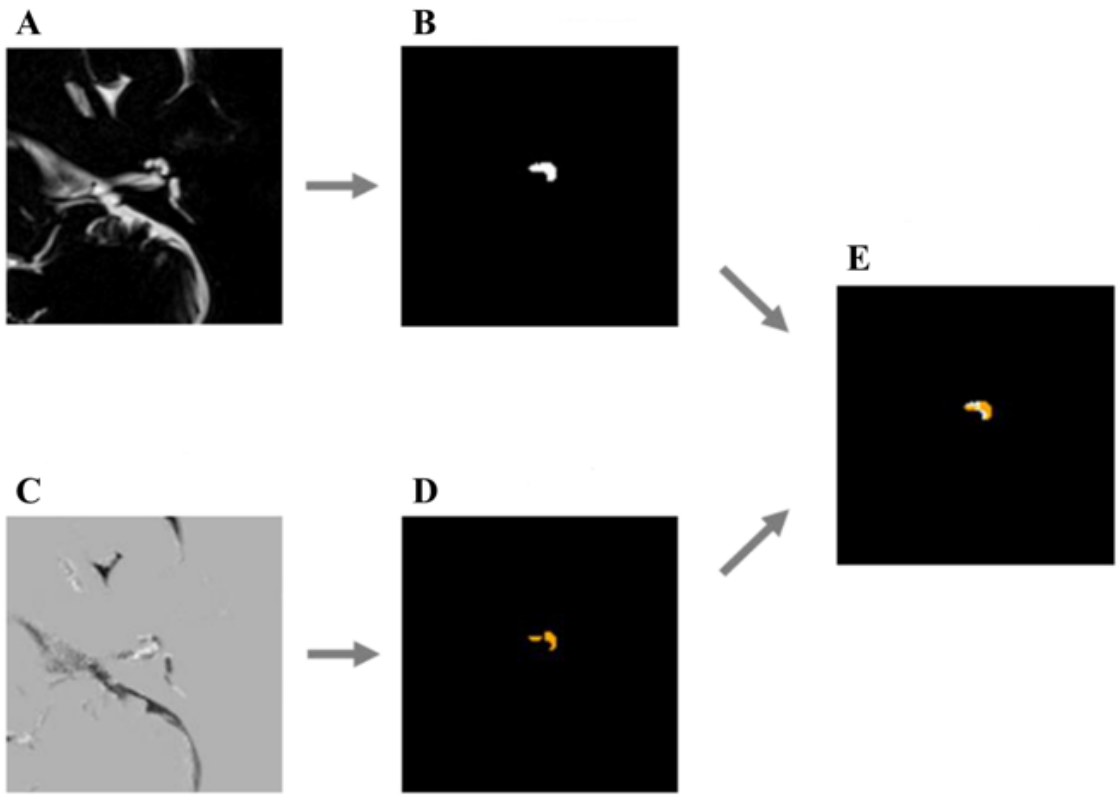
Figure 3. Performance of the inner-ear hydrops estimation via artificial intelligence-version 2 (INHEARIT-v2) system in finding the representative slice index in the magnetic resonance image stacks of 240 target organs of 60 subjects.



Agreement Analysis of EH Ratios That Were Calculated by Using the Test Data Set

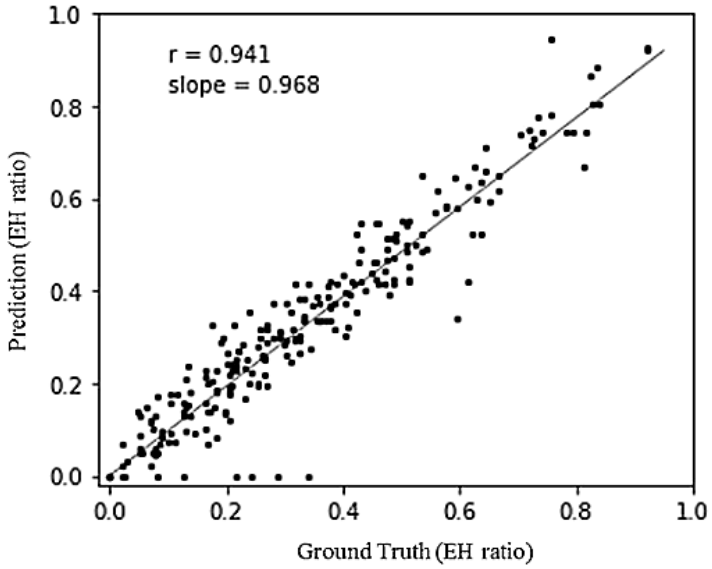
HYDROPS-Mi2 images were generated at the selected slice index, and EH ratios were calculated for all additional test cases. A case example of segmentation and EH visualization is represented in Figure 4; an MRC patch for a left ear was segmented, and the EH of the left cochlea (as seen in the HYDROPS-Mi2 patch) is shown in orange. The segmented result was used as a mask. A hydrops was overlaid on the segmented result, and the calculated EH ratio was 0.633.

Figure 4. Visualization of an endolymphatic hydrops in a left cochlea. A: Magnetic resonance cisternography patch (left ear). B: Left cochlea segmentation result. C: HYDROPS-Mi2 patch (left ear). D: Endolymphatic hydrops. E: Endolymphatic hydrops in the left cochlea. HYDROPS-Mi2: hybrid of the reversed image of the positive endolymph signal and native image of the positive perilymph signal image multiplied by T_2-weighted magnetic resonance cisternography.



The agreement between the physician-calculated and INHEARIT-v2–estimated EH ratios was calculated. The average ICC value for all cases was 0.941; the average ICC of the vestibules was 0.968, and that of cochleae was 0.914 (Figure 5). The average INHEARIT-v2–based calculation time was 3.585 seconds (SD 0.642 seconds) per subject.

Figure 5. The correlation between physician-calculated (ground truth) EH ratios and those predicted by the proposed platform (based on the segmentation area). EH: endolymphatic hydrops.



Discussion

Principal Findings

Due to the development of MRI machines, data and image processing technology, big data, and cloud systems, the role of artificial intelligence in the medical field will likely increase with time. Recently, algorithms have begun to help clinicians in real-world clinics and have been used to predict clinical outcomes that are useful for health care systems [3]. We developed a fully automated analytic system—the INHEARIT-v2 system—for calculating EH ratios by using deep learning and MRI. The INHEARIT-v2 system automatically segmented cochleae and vestibules in brain MRC images and detected representative slices for these organs. The system estimated the EH ratios of segmented regions in key image slices, and the results had a high correlation with those that were manually calculated by experts during an additional test.

Since the mid-2000s, the analysis of EHs by using 3T MRI has been widely accepted as a useful method for the diagnosis of Ménière disease among various groups [10,20]. In particular, injecting contrast media through an intravenous route is a widely used MRI method for patients with Ménière disease because it is less invasive, results in shorter waiting times for patients, and allows physicians to observe both ears simultaneously [11]. Recently, hydropses have been evaluated in several studies either by precisely dividing each area of the cochlea and vestibule or by using a 3D model to measure the whole volume of the endolymphatic space based on magnetic resonance images [21,22].

Recent medical image studies have reported the development of fully automated analytic systems. The implemented tools were designed to provide end-to-end workflows to minimize the amount of human intervention during analysis [23,24]. Several recent studies used deep learning techniques to segment inner ears and related organs in medical images. Some studies were conducted on computed tomography images to segment other organs, such as the sigmoid sinus, facial nerves, or temporal bones [25,26]. Another study analyzed magnetic resonance images for labyrinth segmentation [27]. These studies analyzed various ear-related organs in 3D space, but the main purpose of segmentation was primarily anatomical visualization for surgical planning. However, our study is different because the objective of segmentation was to calculate EH ratios in patients with Ménière disease by conjoining different magnetic resonance image modalities.

The results of our study substantially complemented the limitations of our previous study [12]. Although our former results laid the foundation for conducting EH analyses with deep learning technology, in this study, the improvements allowed for immediate clinical diagnosis and follow-up. Consequently, the accuracy of the EH analysis increased, and even if a human expert did not choose a representative image slice, the representative section was automatically extracted from the magnetic resonance image stack and analyzed successfully.

Another technical improvement over our previous study is that in this study, various deep learning models, such as the U-Net and Inception-v3 networks, were used as a base architecture. We developed a better performing model that effectively segmented inner ears. Furthermore, an additional test was performed to prove the robustness of the system. By developing the INHEARIT-v2 system, expert interpretation became much easier and faster, and more objective analyses were possible.

The best performance, which was based on the mean area under the receiver operating characteristic curve from the fivefold cross-validation, was achieved by the 3intoUNet network, which was based on the U-Net architecture. The U-Net network is designed for medical image segmentation [16]. As such, it is distinguishable from other deep learning models, such as VGG-19 and Inception-v3. Further, it was initially developed for the classification of raw images, such as those of a car, building, or human [13,15]. The 3intoUNet network allows for the analysis of consecutive images, similar to how medical experts browse an MRI stack to identify the location of a target organ. This serial-image training approach was suggested in our previous study on using a VGG-19–based model, which proved to be effective in a medical image segmentation task [12]. The same approach was successful for both the U-Net–based and Inception-v3–based networks. Since inner ears occupy only a small portion of the area (<60 pixels) in whole-brain images (384×324 pixels), the 3intoUNet network has a generalizable architecture for analyzing full-stack magnetic resonance images.

With regard to the performance of the system in automatic representative slice selection, when compared to physicians' choices, the system's choices had a gap distance of ≤2 for 100% (240/240) of the test data set. The physicians' goal was to select a key slice of an organ based on the organ's anatomical relationship with other organs [12], whereas the system's goal was to locate the slice with the largest segmented area for each organ. Despite the possible minimal misalignment between the automatically selected and manually selected key slices, the correlation between the INHEARIT-v2–calculated and physician-calculated EH ratio values was high, indicating that the amount of misalignment did not substantially affect the EH ratio calculation. Based on our results, measuring the entire hydrops volume of the cochlea and vestibule for diagnosis is unnecessary.

Several concerns might arise when artificial intelligence systems are used in real clinical settings. In this study, we found that using such artificial intelligence systems could be an alternative to manually measuring hydrops ratios in real settings. However, this does not mean that Ménière disease can only be diagnosed by this system. The cutoff value for the EH ratio was not clearly defined during the diagnosis of definite Ménière disease. Future studies for estimating the EH ratio cutoff value in the diagnosis of definite Ménière disease are still needed. In addition, HYDROPS-Mi2 images were used to analyze EH ratios. However, these images were acquired via software modification, and hydropses in modified images can be more exaggerated compared to those in original images. The final concern is that Ménière disease is a multifactorial disease, which means that making an accurate diagnosis is typically a very complex process. Other diseases that mimic the symptoms of Ménière disease should be ruled out in clinical settings. The main purpose of hydrops measurement via artificial intelligence is to provide clinical support, which can be helpful for medical professionals when making a final clinical diagnosis.

Future studies can include additional normal control subjects to determine the optimal EH ratio threshold by comparing individuals with Ménière disease and healthy individuals without the disease. However, the association between clinical symptoms and EHs is not uniform from patient to patient; thus, such analyses require a clinician's comprehensive judgment [28]. To improve these analyses, in addition to magnetic resonance images, a model for analyzing heterogeneous data, such as the clinical variables used for diagnosis, can be applied to the deep learning algorithm. In addition, because the cause and mechanism of Ménière disease have not been fully elucidated, such technology can be widely used for the differential diagnosis of other conditions that are thought to be associated with EHs [29].

Notably, we fully automated the calculation of EH ratios by developing an analytic system—the INHEARIT-v2 system—by using MRI and deep learning, which have important clinical implications. Although several aspects should be further investigated, this framework will be a helpful tool for clinicians who adopt an MRI analysis approach for diagnosing patients with Ménière disease.

Conclusion

We developed a fully automated system—the INHEARIT-v2 system—for calculating EH ratios by using deep learning and MRI. The proposed system can quickly and accurately analyze EHs without the intervention of an expert of various inner ear diseases, including Ménière disease experts. This automatic system can perform objective and time-saving analyses for assessing the EH ratios in the inner ear magnetic resonance images of patients with Ménière disease.

Acknowledgments

This work was supported by National Research Foundation of Korea grants, which are funded by the Korean government (the Ministry of Science and ICT; grants NRF-2020R1F1A1073904 and NRF-2021M3E5D2A01022399), and the Future Medicine 20*30 Project of the Samsung Medical Center (grant SMX1210791).

Authors' Contributions

BHC, WHC, and JWK designed the research. WHC and YSC collected the data. CJP, YKK, and KK analyzed the data. CJP and YSC wrote the main paper. BHC, WHC, and HJK provided critical revisions, discussed the results and implications, and commented on the manuscript at all stages.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The concept of the 3intoUNet network.
[PNG File , 326 KB - jmir_v23i9e29678_app1.png]

References

1. Brody H. Medical imaging. Nature 2013 Oct 31;502(7473):S81. [doi: 10.1038/502s81a] [Medline: 24187698]
2. Shen D, Wu G, Suk H. Deep learning in medical image analysis. Annu Rev Biomed Eng 2017 Jun 21;19:221-248 [FREE Full text] [doi: 10.1146/annurev-bioeng-071516-044442] [Medline: 28301734]
3. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
4. Kotimäki J, Sorri M, Aantaa E, Nuutinen J. Prevalence of Meniere disease in Finland. Laryngoscope 1999 May;109(5):748-753. [doi: 10.1097/00005537-199905000-00013] [Medline: 10334225]
5. Tyrrell JS, Whinney DJD, Ukoumunne OC, Fleming LE, Osborne NJ. Prevalence, associated factors, and comorbid conditions for Ménière's disease. Ear Hear 2014;35(4):e162-e169. [doi: 10.1097/AUD.0000000000000041] [Medline: 24732693]
6. Hallpike CS, Cairns H. Observations on the pathology of Ménière's syndrome: (Section of Otology). Proc R Soc Med 1938 Sep;31(11):1317-1336 [FREE Full text] [Medline: 19991672]
7. Durrant JD, Wang J, Ding DL, Salvi RJ. Are inner or outer hair cells the source of summating potentials recorded from the round window? J Acoust Soc Am 1998 Jul;104(1):370-377. [doi: 10.1121/1.423293]
8. Naganawa S, Suzuki K, Nakamichi R, Bokura K, Yoshida T, Sone M, et al. Semi-quantification of endolymphatic size on MR imaging after intravenous injection of single-dose gadodiamide: comparison between two types of processing strategies. Magn Reson Med Sci 2013 Dec 25;12(4):261-269 [FREE Full text] [doi: 10.2463/mrms.2013-0019] [Medline: 24172793]
9. Gürkov R. Menière and friends: Imaging and classification of hydropic ear disease. Otol Neurotol 2017 Dec;38(10):e539-e544. [doi: 10.1097/MAO.0000000000001479] [Medline: 29135874]
10. Cho YS, Ahn JM, Choi JE, Park HW, Kim YK, Kim HJ, et al. Usefulness of intravenous gadolinium inner ear MR imaging in diagnosis of Ménière's disease. Sci Rep 2018 Dec 03;8(1):17562 [FREE Full text] [doi: 10.1038/s41598-018-35709-5] [Medline: 30510158]
11. Iida T, Teranishi M, Yoshida T, Otake H, Sone M, Kato M, et al. Magnetic resonance imaging of the inner ear after both intratympanic and intravenous gadolinium injections. Acta Otolaryngol 2013 May;133(5):434-438. [doi: 10.3109/00016489.2012.753640] [Medline: 23294239]
12. Cho YS, Cho K, Park CJ, Chung MJ, Kim JH, Kim K, et al. Automated measurement of hydrops ratio from MRI in patients with Ménière's disease using CNN-based segmentation. Sci Rep 2020 Apr 24;10(1):7003 [FREE Full text] [doi: 10.1038/s41598-020-63887-8] [Medline: 32332804]
13. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv. Preprint posted online on September 4, 2014 [FREE Full text]
14. Naganawa S, Yamazaki M, Kawai H, Bokura K, Sone M, Nakashima T. Imaging of Ménière's disease after intravenous administration of single-dose gadodiamide: utility of subtraction images with different inversion time. Magn Reson Med Sci 2012;11(3):213-219 [FREE Full text] [doi: 10.2463/mrms.11.213] [Medline: 23037568]
15. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. 2016 Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27-30, 2016; Las Vegas, Nevada, USA. [doi: 10.1109/cvpr.2016.308]
16. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. 2015 Presented at: 18th International Conference on Medical Image Computing and Computer-Assisted Intervention; October 5-9, 2015; Munich, Germany p. 234-241. [doi: 10.1007/978-3-319-24574-4_28]
17. Kingma DP, Ba J. Adam: a method for stochastic optimization. 2015 Presented at: The 3rd International Conference on Learning Representations (ICLR); May 7-9, 2015; San Diego, California, USA.
18. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15:1929-1958 [FREE Full text]
19. The R Project for Statistical Computing. The R Foundation. URL: https://www.R-project.org/ [accessed 2021-08-06]
20. Quatre R, Attyé A, Karkas A, Job A, Dumas G, Schmerber S. Relationship between audio-vestibular functional tests and inner ear MRI in Meniere's disease. Ear Hear 2019;40(1):168-176. [doi: 10.1097/AUD.0000000000000584] [Medline: 29698363]
21. Attyé A, Eliezer M, Boudiaf N, Tropres I, Chechin D, Schmerber S, et al. MRI of endolymphatic hydrops in patients with Meniere's disease: a case-controlled study with a simplified classification based on saccular morphology. Eur Radiol 2017 Aug;27(8):3138-3146. [doi: 10.1007/s00330-016-4701-z] [Medline: 27999985]
22. Connor SEJ, Pai I. Endolymphatic hydrops magnetic resonance imaging in Ménière's disease. Clin Radiol 2021 Jan;76(1):76.e1-76.e19. [doi: 10.1016/j.crad.2020.07.021] [Medline: 32892985]

23. Nalepa J, Lorenzo PR, Marcinkiewicz M, Bobek-Billewicz B, Wawrzyniak P, Walczak M, et al. Fully-automated deep learning-powered system for DCE-MRI analysis of brain tumors. Artif Intell Med 2020 Jan;102:101769. [doi: 10.1016/j.artmed.2019.101769] [Medline: 31980106]
24. Thomas N, Perumalla A, Rao S, Thangaraj V, Ravi K, Geethanath S, et al. Fully automated end-to-end neuroimaging workflow for mental health screening. 2020 Presented at: 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE); October 26-28, 2020; Cincinnati, Ohio, USA. [doi: 10.1109/bibe50027.2020.00109]
25. Nikan S, Van Osch K, Bartling M, Allen DG, Rohani SA, Connors B, et al. PWD-3DNet: A deep learning-based fully-automated segmentation of multiple structures on temporal bone CT scans. IEEE Trans Image Process 2021;30:739-753. [doi: 10.1109/TIP.2020.3038363] [Medline: 33226942]
26. Fauser J, Stenin I, Bauer M, Hsu WH, Kristin J, Klenzner T, et al. Toward an automatic preoperative pipeline for image-guided temporal bone surgery. Int J Comput Assist Radiol Surg 2019 Jun;14(6):967-976. [doi: 10.1007/s11548-019-01937-x] [Medline: 30888596]
27. Vaidyanathan A, van der Lubbe MFJA, Leijenaar RTH, van Hoof M, Zerka F, Miraglio B, et al. Deep learning for the fully automated segmentation of the inner ear on MRI. Sci Rep 2021 Feb 03;11(1):2885 [FREE Full text] [doi: 10.1038/s41598-021-82289-y] [Medline: 33536451]
28. Gürkov R, Pyykö I, Zou J, Kentala E. What is Menière's disease? A contemporary re-evaluation of endolymphatic hydrops. J Neurol 2016 Apr;263 Suppl 1:S71-S81 [FREE Full text] [doi: 10.1007/s00415-015-7930-1] [Medline: 27083887]
29. Nakashima T, Pyykkö I, Arroll MA, Casselbrant ML, Foster CA, Manzoor NF, et al. Meniere's disease. Nat Rev Dis Primers 2016 May 12;2:16028. [doi: 10.1038/nrdp.2016.28] [Medline: 27170253]

Abbreviations

EH: endolymphatic hydrops
HYDROPS: hybrid of the reversed image of the positive endolymph signal and native image of the positive perilymph signal
HYDROPS-Mi2: hybrid of the reversed image of the positive endolymph signal and native image of the positive perilymph signal image multiplied by T_2-weighted magnetic resonance cisternography
ICC: intraclass correlation coefficient
INHEARIT: inner-ear hydrops estimation via artificial intelligence
INHEARIT-v2: inner-ear hydrops estimation via artificial intelligence-version 2
IOU: intersection over union
MRC: magnetic resonance cisternography
MRI: magnetic resonance imaging
PEI: positive endolymph image
PPI: positive perilymph image
ROI: region of interest
VGG: Visual Geometry Group

Edited by R Kukafka; submitted 17.04.21; peer-reviewed by M Elbattah, A Staffini; comments to author 26.05.21; revised version received 15.07.21; accepted 27.07.21; published 21.09.21.

Please cite as:
Park CJ, Cho YS, Chung MJ, Kim YK, Kim HJ, Kim K, Ko JW, Chung WH, Cho BH
A Fully Automated Analytic System for Measuring Endolymphatic Hydrops Ratios in Patients With Ménière Disease via Magnetic Resonance Imaging: Deep Learning Model Development Study
J Med Internet Res 2021;23(9):e29678
URL: https://www.jmir.org/2021/9/e29678
doi:10.2196/29678
PMID:34546181

Original Paper

Prevention of Suicidal Relapses in Adolescents With a Smartphone Application: Bayesian Network Analysis of a Preclinical Trial Using In Silico Patient Simulations

Stephane Mouchabac[1,2], MD, MSc; Philippe Leray[3], PhD; Vladimir Adrien[1,2], MD, PhD; Fanny Gollier-Briant[4], MD, PhD; Olivier Bonnot[4,5], MD, PhD

[1]Department of Psychiatry, Sorbonne Université, Hôpital Saint Antoine- APHP, Paris, France

[2]Infrastructure of Clinical Research In Neurosciences- Psychiatry, Brain and Spine Institute (ICM), Inserm UMRS 1127, Centre national de la recherche scientifique, Sorbonne Université, Paris, France

[3]Laboratoire des Sciences du Numérique de Nantes, Centre national de la recherche scientifique, University of Nantes, Nantes, France

[4]Department of Child and Adolescent Psychiatry, Centre hospitalier universitaire de Nantes, Nantes, France

[5]Pays de la Loire Psychology Laboratory EA4638, Nantes, France

Corresponding Author:
Olivier Bonnot, MD, PhD
Department of Child and Adolescent Psychiatry
Centre hospitalier universitaire de Nantes
30 boulevard Jean Monnet
Nantes, 44000
France
Phone: 33 4323232
Email: olivier.bonnot@chu-nantes.fr

Abstract

Background: Recently, artificial intelligence technologies and machine learning methods have offered attractive prospects to design and manage crisis response processes, especially in suicide crisis management. In other domains, most algorithms are based on big data to help diagnose and suggest rational treatment options in medicine. But data in psychiatry are related to behavior and clinical evaluation. They are more heterogeneous, less objective, and incomplete compared to other fields of medicine. Consequently, the use of psychiatric clinical data may lead to less accurate and sometimes impossible-to-build algorithms and provide inefficient digital tools. In this case, the Bayesian network (BN) might be helpful and accurate when constructed from expert knowledge. Medical Companion is a government-funded smartphone application based on repeated questions posed to the subject and algorithm-matched advice to prevent relapse of suicide attempts within several months.

Objective: Our paper aims to present our development of a BN algorithm as a medical device in accordance with the American Psychiatric Association digital healthcare guidelines and to provide results from a preclinical phase.

Methods: The experts are psychiatrists working in university hospitals who are experienced and trained in managing suicidal crises. As recommended when building a BN, we divided the process into 2 tasks. Task 1 is structure determination, representing the qualitative part of the BN. The factors were chosen for their known and demonstrated link with suicidal risk in the literature (clinical, behavioral, and psychometrics) and therapeutic accuracy (advice). Task 2 is parameter elicitation, with the conditional probabilities corresponding to the quantitative part. The 4-step simulation (use case) process allowed us to ensure that the advice was adapted to the clinical states of patients and the context.

Results: For task 1, in this formative part, we defined clinical questions related to the mental state of the patients, and we proposed specific factors related to the questions. Subsequently, we suggested specific advice related to the patient's state. We obtained a structure for the BN with a graphical representation of causal relations between variables. For task 2, several runs of simulations confirmed the a priori model of experts regarding mental state, refining the precision of our model. Moreover, we noticed that the advice had the same distribution as the previous state and was clinically relevant. After 2 rounds of simulation, the experts found the exact match.

Conclusions: BN is an efficient methodology to build an algorithm for a digital assistant dedicated to suicidal crisis management. Digital psychiatry is an emerging field, but it needs validation and testing before being used with patients. Similar to psychotropics,

any medical device requires a phase II (preclinical) trial. With this method, we propose another step to respond to the American Psychiatric Association guidelines.

Trial Registration: ClinicalTrials.gov NCT03975881; https://clinicaltrials.gov/ct2/show/NCT03975881

(J Med Internet Res 2021;23(9):e24560) doi:10.2196/24560

KEYWORDS

suicide; bayesian network; smartphone application; digital psychiatry; artificial intelligence

Introduction

The recent adoption of smartphone health applications that collect data (weight, exercises, etc) and allow the consumer to see graphs and diagrams illustrates the sociological and psychological power of digital self-care and self-management. This phenomenon is of particular importance for adolescents and young adults. However, asking patients to fill in scales or questionnaires on their smartphone instead of on a computer or a paper is not new. This is called ecological momentary assessment (EMA), a naturalistic method to access clinical data [1]. However, to be useful for users, there is a need for advice, therapeutic care, or at least personalized feedback related to the answers provided, which could be the field of ecological momentary intervention (EMI).

With the development of machine learning, it is obvious that EMI will have significant implications in the future. There are only a few experiences in psychiatry with EMI, but recent reviews suggest that results are very promising in mood disorders, anxiety, and schizophrenia [2-5]. However, those apps have different approaches. Some of them focus on providing in-app therapy, such as iBobbly [6], SuperBetter [7], Get Happy [8], or ACT Smart [9]. Few of them provide a more personalized therapy according to EMA answers from a medical team (connected device) such as MONARCA (monitoring treatment and prediction of bipolar disorder episodes) [10] or more automated like PRISM (the person real-time intervention for stabilizing mood) [11]. Except for Mindstrong [12], which uses digital phenotype (ie, no psychological data but only digital use skills such as typing or screen scrolling), none of them are using in-app artificial intelligence (AI) based on EMA data. Interestingly, and somehow counterintuitively, the efficacy of these apps for depression and anxiety is better when they are not part of a program with human interactions [2,13]. This could be an argument for more intuitive and personalized apps capable of being independent of direct care. We also know that the fill rate is higher when patients have more severe symptoms [14].

Suicide attempts (SAs) are a major health care issue and could be an interesting focus for add-on smartphone app care. Suicide is a highly challenging multifactor process encompassing genetics, psychological, social, cultural, and life-experienced trauma influences [15]. More than 25 million people worldwide attempt suicide each year, and around 800,000 people die from suicide [16]. Repeated epidemiological studies show that SA history is one of the major risks of suicide. Within a year of the SA, the repetition rate in adolescent and young adult populations ranges from 15% to 28% [17]. Rates of repetition are higher within 6 months of the previous SA [18]. To date, no suicide prevention program is better than others, and all of them require the active participation of health care professionals (eg, training, calls, interviews, etc) [19]. However, studies have shown that personalized brief contact interventions reduce recurrence after an SA [20]. A specific and personalized app could optimize these strategies.

Therefore, to prevent a patient's SA and suicide, we have created an EMA plus EMI new approach based on a mobile health care application. Our application, which is not connected to the medical team via the internet, will collect data from the patient twice a day regarding anxiety, mood, and sleep disorders, with decreasing frequency during 1 year. The data entries by the patient will be made via analogic visual scales and drop-down lists. Then, algorithm-based feedback enhanced with AI will proactively inform the patient with comments and advice based on the World Health Organization's recommendations or associated self-coping or mindfulness practices.

The aim is to position our smartphone application like a health care partner, which is why we named it "Medical Companion." However, to be relevant, our app needs to encompass AI.

Many specialists are convinced that AI and machine learning will be major breakthroughs in medicine and psychiatry [21-23] and could disrupt the practices [24]. Machine learning is a computational strategy that automatically determines (learns) methods and parameters to find the best solution for a problem rather than provide a previously set solution programmed a priori by a human. The strength of machine learning is its ability to explore multiple patterns in complex data. The algorithms used in machine learning can categorize, cluster, and predict by using supervised and unsupervised techniques. The most used algorithm is the support vector machine, a multivariate supervised learning technique that classifies subjects into groups within a margin-based statistical framework.

Promising achievements have been reached by this kind of bottom-up analysis with machine learning. For example, in medical imaging, researchers used more than 1000 anonymous patient X-rays to train a naïve Bayesian network (BN) to detect tuberculosis [25]. The algorithm had close to a 100% accuracy rate. There are several recent reviews and perspectives in this field [22,26-28].

The same approach has been used in psychiatry, with a research focus on diagnosis, treatment effect prediction, and outcome prediction. Authors used physiological markers [29], neuroimaging data [30] or clinical specific scales [31] with up to 75% accuracy [30,32].

However, for optimal accuracy, machine learning needs a substantial amount of data. Contrary to biological or imaging data, clinical psychiatry produces heterogenic data often

associated with poor availability and management [21,33]. Moreover, this prediction bias increases with large samples when calculated with probabilistic statistics [21,33,34]. Therefore, some authors suggest the use of more expert-based algorithms using BN [35,36]. Our goal is to create an algorithm that selects different types of pre-established advice based on each individual's limited data.

Our algorithm, developed with this expert BN technique, will achieve a high level of personalized advice and comments based on individual answers to our in-app questions.

This paper describes the method to build the algorithm. It is part of a large trial, oriented on clinical efficiency and designed per recent American Psychiatric Association (APA) guidelines for digital devices [37]. We are currently working on the smartphone application by itself, and we will run a clinical trial to evaluate its feasibility, efficacy, and use (clinical trial NCT03975881) in real life.

Methods

Theoretical Background of BN

Pearl [38] defines a BN as a probabilistic graphical model that allows the representation of a set of variables and their probabilistic relationships using a directed acyclic graph (DAG). By connecting the cause to the effect with an arrow, we obtain the most intuitive graphical representation of the influence of one event, one act, or one variable on another. It models the causal relationships well. Interestingly, BNs are knowledge representation models that can be built from expert experiences. Furthermore, a BN can learn from data, updating knowledge of the status of a subset of variables when other variables (the evidence variables) are observed and perform inferences by an incremental process.

A BN is an annotated DAG that encodes a joint distribution for a random set of variables X. Formally, a BN for X is a pair: $B = [G,\Theta]$. For a summary of notations used in the text, see Textbox 1.

The first component $G = (X, U)$ is determined by a set of variables $X = \{X_1, X_2...X_n\}$ whose elements are called vertices or nodes and a set $U = \{u_1, u_2,..., u_m\}$ of the Cartesian product $X \times X$ whose elements are called arcs or edges, which represent the direct dependencies between the variables. For an arc $u = (X_i, X_j)$, x_i is the initial end (origin), X_j is the final end (destination). An arc u is directed, starting from X_i and arriving at X_j. Above all, the graph G encodes independence assumptions to satisfy the local Markov property—each variable X_i is independent of its nondescendants given its parents in G.

The second component of the pair, namely Θ, represents the set of parameters that quantify the network. It contains a set of conditional probability distributions (CPD) $\Theta_i = P_B (X_i|In\Pi_i)$ for each variable X_i, where Π_i denotes the set of parents of X_i in G.

Textbox 1. Notations used in the text and their meaning.

- *B:* Bayesian network
- *G:* Graph
- Θ*:* Set of parameters that quantifies the network
- *X:* Node variable in the model
- Π_i: Set of parents of X_i
- *U:* Set of edges
- *u:* An arc

The BN defines a unique joint probability distribution over X given by:

In BNs, conditional probability tables (CPTs) associated with each node should be defined to measure the relationships between variables. However, it has been pointed out that it is usually difficult to quantify the CPTs due to the complexity of the BN, which is defined by its dimension (ie, the number of independent parameters used to describe its CPD).

where r_i is the cardinality of X_i and  is the number of configurations of the parents of X_i. If X_i has no parents, q_i=1.

So, if a binary variable has n binary parents, the corresponding contribution q_i of its CPD to the model complexity is exponential and equals 2^n, since each r_j=2. Specifically, the elicitation of all BN probabilities is a complex and time-consuming task. To simplify a CPT, we can use a noisy-or model (NOM) where the number of probability values to estimate is now proportional to the number n of parents. Thus, it allows for simplifying the CPD [38].

Building a BN From Expertise

The knowledge bases are built to formalize our knowledge in specific domains and support our reasoning on events and decisions in a structured way. A BN is a declarative ("knowing what") knowledge-representation formalism constructed from expertise. Due to its probabilistic content, it allows for exploiting more efficiently the structure of the knowledge bases.

Constructing a BN from expertise is based on several steps. For this purpose, Kjaerulff and Madsen [39] divide the process into 2 tasks: (1) structure determination, representing the qualitative part of the BN, and (2) parameter elicitation, with the conditional probabilities, corresponding to the quantitative part.

Determining the structure requires skills, inventiveness, close communication with experts, and a high level of expertise. This process needs to address 2 main tasks: (1) identification of the relevant variables and (2) identification of the links between the variables.

The parameter elicitation makes the tacit knowledge as explicit as possible (and therefore easier to convey) and formalizes the expert's reasoning in an inference engine. It aims at eliciting subjective conditional probabilities from expertise to artificially reproduce the analysis of the situation and the decision-making of the expert. Finally, expert domain knowledge can be handled as prior distribution, but it is not certain that the expert has the experience to formulate valid probability judgments naturally. Authors point out that the experts must be assisted by an experienced probabilistic facilitator who will provide feedback to the experts, for example, by using simulations [40]. Inconsistencies will be discussed and will allow the model to be adjusted. Inversely, the facilitator needs a clear understanding of the decision problem for which an elicited probability distribution is required.

Many elicitation methods exist, but the most popular is the roulette method (or the "chips and bins method" or histogram method) [41]. It is prized by the experts because it provides visual feedback in the form of histograms.

Building Our BN

Our project (clinical trial NCT03975881) is an innovative approach to prevent the relapse of SAs and suicides in patients with a previous SA. It is the core of our healthcare smartphone application. The program is an add-on to the usual care process and is built to work autonomously. The BN will help to:

1. Estimate actual psychiatric and behavioral state based on users' answers to in-app questions.
2. Match the estimated state with appropriate and accurate preexisting advice and care provided by the smartphone app.

Variables Determination for Estimating Behavioral and Psychiatric State

At first, we have identified contextual variables that are mandatory data in any clinical study (ie, age, gender, type of SA, date, and localization). Only age and gender will be exploited in the algorithm to reduce the required amount of expert estimation. Then we defined clinical questions related to the mental state of the patients. All questions (Q) have been provided by experts in the form of a single choice and are related to specifics factors. The elicitation of each clinical dimension corresponds to a set of questions $Q = \{Q_1 \ldots Q_m\}$.

The clinical dimensions were chosen for their known and demonstrated link with the suicidal risk in the literature [42,43]. Some are clinical markers related to depression (mood, cognitive retardation, sleep or appetite disorders, and physical pain) or anxiety, as well as a measure of the presence and the severity of suicidal ideation. The intensity measures of suicidal thoughts were adapted from specific psychometric scales (ie, the Columbia-Suicide Severity Rating Scale) [44]. Others are behavioral markers that can be disturbed if the patient's clinical condition changes (eg, the use of video games) or may promote suicidal ideation (eg, substance use disorder).

We had 9 clinical dimensions, 23 questions, and 10 pieces of advice. All questions were linked according to clinical knowledge and translated into the BN model. Categories were anxiety, sleep, appetite, physical complaints (eg, headache, stomach pain, etc), cognitive impairment (concentration and memory), mood, suicidal thought and behavior, and addictive behavior (eg, alcohol, tobacco, screens, and games, etc).

To simplify and fit a four-level Likert scale, an expert group created the questions to approach the mental state category. For example, "How anxious are you?", "How much were you able to cope with your anxiety?" and "How long were you anxious during the last day?" For the last question, instead of "absent," "low," "medium," or "high," we opted for "none," "less than 1 hour," "from 1 to 3 hours," and "almost all day." All questions were a priori scored by an expert according to the BN structure.

All patient answers were made of visual analog scales (VASs). The formulated questions were slightly modified every day (and twice a day during the first weeks) to avoid a perception of a process that seems too "automatic."

More precisely, the discussions between the experts led us to identify 3 variables for each factor, described as follows:

1. Immediate value of factor (IF) corresponds to the measure in percentage extracted from the questions for the VAS at instant *t*.
2. Cumulated value of factor (CF) is the mean of time-repeated immediate value. If we consider that the weight given to each datum is equivalent, the more the data increases, the lower the weight of each datum, which will decrease the responsiveness of the system (ie, the anxiety experienced several weeks ago might not have the same impact for calculating an estimated state than a few hours ago). To address this issue, we have introduced a forgetting factor α to reduce the weight of older data as follows:
   - $CF_{new} = (1-\alpha) \times IF + \alpha \times CFold$
   - If $\alpha=0$, no aggregation, CF=IF; if $\alpha=1$, no update, the CF remains unchanged.
3. Contextual severity (CS) corresponds to a discrete variable with a domain (high, medium, low, absent) that describes the seriousness of one given dimension with respect to one given context (which is the CF describing the user characteristic). For instance, for some dimensions, a high IF is "severe" when the CF for this dimension is also high, but it is less serious when the CF is low (ie, if a patient anxiety state is "none," a single high anxiety score stays relatively low compared to a patient who is repeatedly anxious, independent of their level of immediate anxiety).

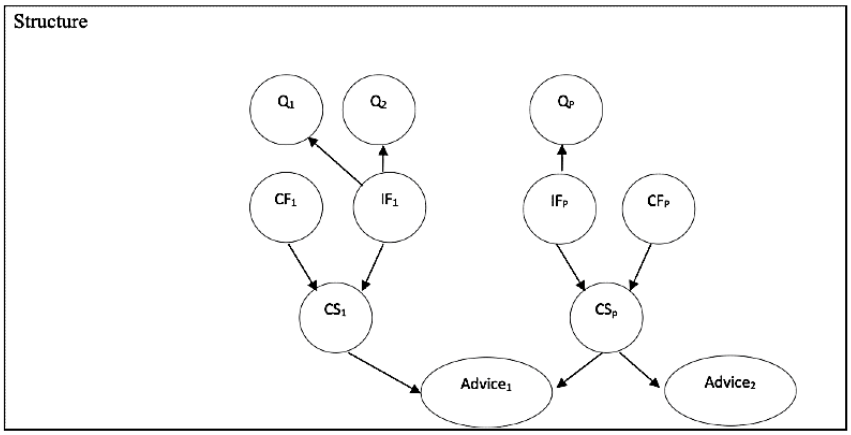
Structure Determination of the BN

The BN structure (Figure 1) models the expert way such as:

1. The state regarding one dimension can vary according to the patient context, but as contextual variables, only age and gender are included in the model.
2. The patient will not answer a question regarding his state similarly for the corresponding dimension (eg, answers concerning anxiety will be different if he is usually very anxious vs not anxious).
3. The severity of one specific dimension depends on a "comparison" between its IF and CF, confronting temporal aspects of states.
4. Each advice depends on the severity of potentially several dimensions because it could be triggered by one or many dimensions.

The CS of a specific dimension depends on both IF and CF. The combination of different CS from a selected set of dimensions gives the most efficient advice related to those dimensions. Nodes are represented as circles and edges as arrows. Responses to each question (Q) associated to each dimension depend on the IF of the dimension, and they become independent of each other if we know the value of IF.

Figure 1. Bayesian Network Structure. CF: cumulated value of factor; CS: contextual severity; IF: immediate value of factor; Q: question.



Parameters Elicitation

According to the 2-task process described by Kjaerulff and Madsen [45], we identified from the previous structure the following parameters (ie, CPD) to validate the work of our experts. Due to the high number of dimensions used in this work and the large volume of data, we chose to illustrate this part only with a selected factor, anxiety, because it is a dimension whose severity can change very acutely. Therefore, we can illustrate the model's reactivity to an acute change of clinical state. The same has been applied to each dimension.

For each dimension *I*, we defined an a priori probability as follows: $P(IF_i=high)$, $P(IF_i=medium$, $P(IF_i=low)$, and $P(IF_i=absent)$ so that a "random" patient will be associated with this dimension to the corresponding level (high, medium, low, and absent) regarding his context (age and gender). $P(CF_i)$ will be initialized to $P(IF_i)$.

For each question related to its dimension, $P(Q = j/IF_i = k)$ is the probability that the answer of this question Q will be the *j*th *one*, given that the patient IF is in its *k*th level.

The main issue is to estimate how a specific individual with specific clinical characteristics may answer a question about his specific state with sufficient accuracy. However, there are no epidemiological or clinical works available with such a level of detail regarding the studied population. Thus, it is impossible to refer to scientific data for the percentage of patients with a medium level of anxiety answering low anxiety to a question asked. Through their knowledge and experience, experts may be able to provide accurate answers. To estimate the value in the domains (absent, low, medium, and high), we use probability calculations as follows:

- $P(CS_i/IF_i,CF_i)$: Defined the severity given the dimensions' IF and CF.
- $P(advice/CS_1...CS_p)$: The scoring depends on all the dimensions' severity, specifically, CS, and as previously described in the theoretical background section of BN, a NOM distribution is used to simplify this dependency.

Experts also offered advice addressing many efficient actions for suicidal crisis management such as lifestyle, sleep, relaxation or mindfulness exercises, management of social rhythms, physical activity, and emergency solutions (ie, calling person in the patient's pre-established circle of confidence, exercising with abdominal breathing, practicing mindfulness or relaxation, and calling emergency services) [44].

Each piece of advice might be useful for every factor but at different levels of accuracy. Therefore, the advice must have a good sensitivity (ie, if it is essential, the system must not omit it). Similarly, if it is not essential given the patient's state, it should not be triggered in excess, so it must have good specificity. For this purpose, experts have estimated the a priori probability $P(advice/ CS_1...CS_p)$ of advice efficiency due to the intensity of each factor.

Algorithm Accuracy Testing by Simulation

Step 1

Building the algorithm is a multistep process. First, we perform a simulation of the first data set for mental state provided by the expert (Table 1) using the BN simulation model. Our model generates three distributions of factors: (a) a priori, (b) when all answers associated with the factor are the highest and have very severe scoring P(++), and (c) when all answers associated

with the factor are the lowest and have less severe scoring P(--). Simulations are made for each category of mental state.

Table 1. Expert-based Bayesian network construct regarding anxiety.

What is your level of anxiety?	Distribution for 100 subjects with no (absent) anxiety	Distribution for 100 subjects with low level of anxiety	Distribution for 100 subjects with medium level of anxiety	Distribution for 100 subjects with high level of anxiety
No anxiety	100	5	0	0
Low anxiety	0	90	15	5
Medium anxiety	0	5	80	15
High anxiety	0	0	5	80
Total	100	100	100	100

Step 2

Return to the expert with simulation results. If simulations are accurate regarding medical expert knowledge, step 3 is possible. If not, experts provide a new data set, and we rerun step 1.

Step 3

Next, we perform a simulation of the first data set for the advice provided by experts using the BN simulation model (Table 2). Our model generates incremental updates (1 to 15 successive scorings by patients) of (a) IF in a specific pattern of the first 9 successive scorings at the lowest score, (b) CS values (based on CF) in a specific pattern of the first 9 successive scorings at lowest score then, IF in a progressive increasing severity pattern from scoring 10 to 15, (c) CS values (based on CF) in a progressive increasing severity pattern from scoring 10 to 15 then, IF for an immediate highest severity scoring starting at answer 10 lasting until answer 15, (d) CS values (based on cumulated severity) for an immediate highest severity scoring starting at answer 10 lasting until answer 15, and (d) for each of the 1 to 15 scorings in both patterns (ie, immediate or progressive), our algorithm will generate the advice and recommendations.

Table 2. Conditional probability of contextual severity as a function of the immediate value and cumulated value of a given factor.

CF[a] and IF[b]	Contextual severity function of CF and IF			
	High	Medium	Low	Absent
High				
Medium	0.98	0.02	0	0
High	0.99	0.01	0	0
Low	0.9	0.1	0	0
Absent	0.85	0.15	0	0
Medium				
High	0.19	0.8	0.01	0
Medium	0	0.99	0.01	0
Low	0	0.71	0.28	0.01
Absent	0	0.66	0.3	0.04
Low				
High	0.11	0.23	0.65	0.01
Medium	0	0.19	0.8	0.01
Low	0	0.01	0.98	0.01
Absent	0	0.01	0.7	0.29
Absent				
High	0.05	0.15	0.3	0.5
Medium	0	0.12	0.23	0.65
Low	0	0.01	0.19	0.8
Absent	0	0	0.01	0.99

[a]CF: cumulated value of factor.

[b]IF: immediate value of factor.

Step 4

Different values and their distribution along the train of different pre-established scoring are presented to experts for review. They are asked to examine the accuracy and consistency of the advice proposed compared to their clinical expertise and knowledge. Then, if necessary, changes are made in the model, and the simulation is rerun beginning with step 3. Reviews and reruns are made as many times as necessary to achieve our goal. Expert opinion is a qualitative variable (yes or no).

Results

Simulations were done according to our 4-step methodology. For each run, experts provided estimations regarding the distribution of answers in specific populations (Table 1). Experts are asked to estimate for 100 subjects with a specific state regarding anxiety (absence, low, medium, or high) how they should answer, and what would be the distribution of answers (no, low, medium, or high) because not all highly anxious patients will submit a high anxiety rating. This corresponds to $P(answer(Q)/IF_i)$.

At each run, we calculated the conditional probability of CS as a function of the IF and CF of a given factor (Table 2). For example, if the IF and CF are high, the probability that CS will be high is 0.99 (expressing few changes). But if IF is low, with a high CF, the probability that the CF remains high is 0.9 and zero for low and absent anxiety, respectively.

The distribution of dimensions can be seen in Figure 2. After two simulation runs, the experts found that the advice proposals were accurate. The a priori status is a distribution of all 4 severity levels (absent, low, medium, and high) as proposed by the experts. In the two extreme conditions (lowest and highest severity), the experts are far less heterogeneous and become more accurate regarding clinical knowledge. Priori denotes the expert estimated distribution of answers by severity in our population of patients with previous SAs (absent in blue, low in green, moderate in orange, and severe in red). A P(++) rating of users' scores illustrates a severe rating for all questions related to anxiety, and we calculated the algorithm-generated score for each related category of mental state. A P(--) rating of users' score demonstrates the lowest rating for all questions related to anxiety, and we calculated the algorithm-generated score for each related category of mental state.

In the second phase, the main goal was to achieve an accurate match between all scoring by categories and the advice. The experts did provide a priori distribution of answers according to the BN structures (Table 3). For example, the first line means "what is the a priori probability that the immediate state is high, medium, low, and absent knowing that the person is a man aged above 18 years.

The final simulations of matching answers and advice for anxiety are provided in Figure 3 and Figure 4, and an example of an expert-based BN construct is shown in Table 4. Figure 3 represents the final run simulation of the probability of advice proposal during incremental updates of 15 successive scorings by virtual patients according to an immediate worsening after 9 successive lowest scores. In Figure 3a, for each dimension, ordinate scores above zero correspond to IF and below zero CS values (based on CF). Moreover, in Figure 3b, the normalized score of the 10 pieces of advice related to anxiety are represented over the 15 steps of the simulation. Figures 4a and b show the final run simulation of the probability of advice proposal during incremental updates according to a progressive worsening after 9 successive lowest scores. Table 4 demonstrates an expert's BN construct regarding the "call friends" advice and their accuracy in categorizing patients with various anxiety levels. For example, this specific advice might be relevant for 50 out of 100 patients with low anxiety level and 99 out of 100 with high anxiety level. This corresponds to P(advice = call friend/$CS_{anxiety}$). This construction has been made for each piece of advice.

Figure 2. Distribution of probabilities for each dimension, for a priori status (P), high scoring P(++), and absent scoring status P(--).

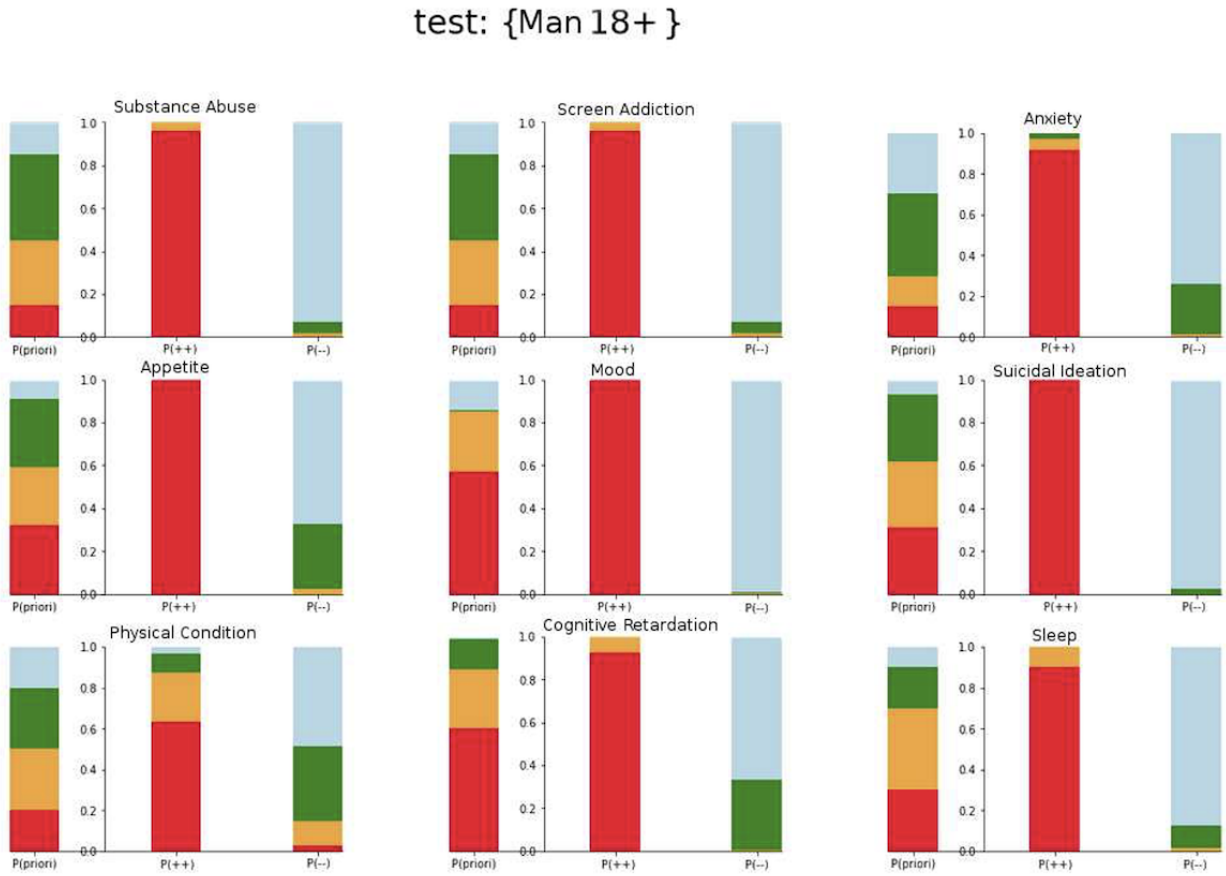


Table 3. Probability distribution for anxiety of the immediate value and cumulated value a priori based on age and gender.

	High	Medium	Low	Absent
Men, age (years)				
>18	0.15	0.15	0.4	0.3
<18	0.2	0.2	0.35	0.25
Women, age (years)				
>18	0.28	0.28	0.22	0.22
<18	0.3	0.3	0.2	0.2

In the first part of each figure (Figure 3a and Figure 4a), IF and CS for each dimension are represented. Above zero is the a priori distribution of IF; below zero reflects the progressive increase of knowledge (learning ability) represented as CS becoming progressively the mirror of IF over time.

In the second part of each figure (Figure 3b and Figure 4b), advice suggestion probabilities are initially fixed in the proportions of a general population of patients that committed SA independent of their clinical state. By increasing the knowledge of the specific clinical state of a given patient (ie, with a low and constant level of anxiety), the advice is adapting and stabilizing. When anxiety IF scoring becomes suddenly (3) or progressively (4) severe, some of them are rising (eg, emergencies and calls to referent psychiatrist or the person of trust), and others are decreasing, finally stabilizing again in another configuration.

The simulation stops when the advice has the same distribution as the a priori state and the advice is clinically relevant. After two simulation runs, experts found accurate matching.

Figure 3. Final run simulation of the probability of advice proposal during incremental updates (immediate worsening after 9 successive lowest scores). IPSRT: interpersonal and social rhythm therapy.

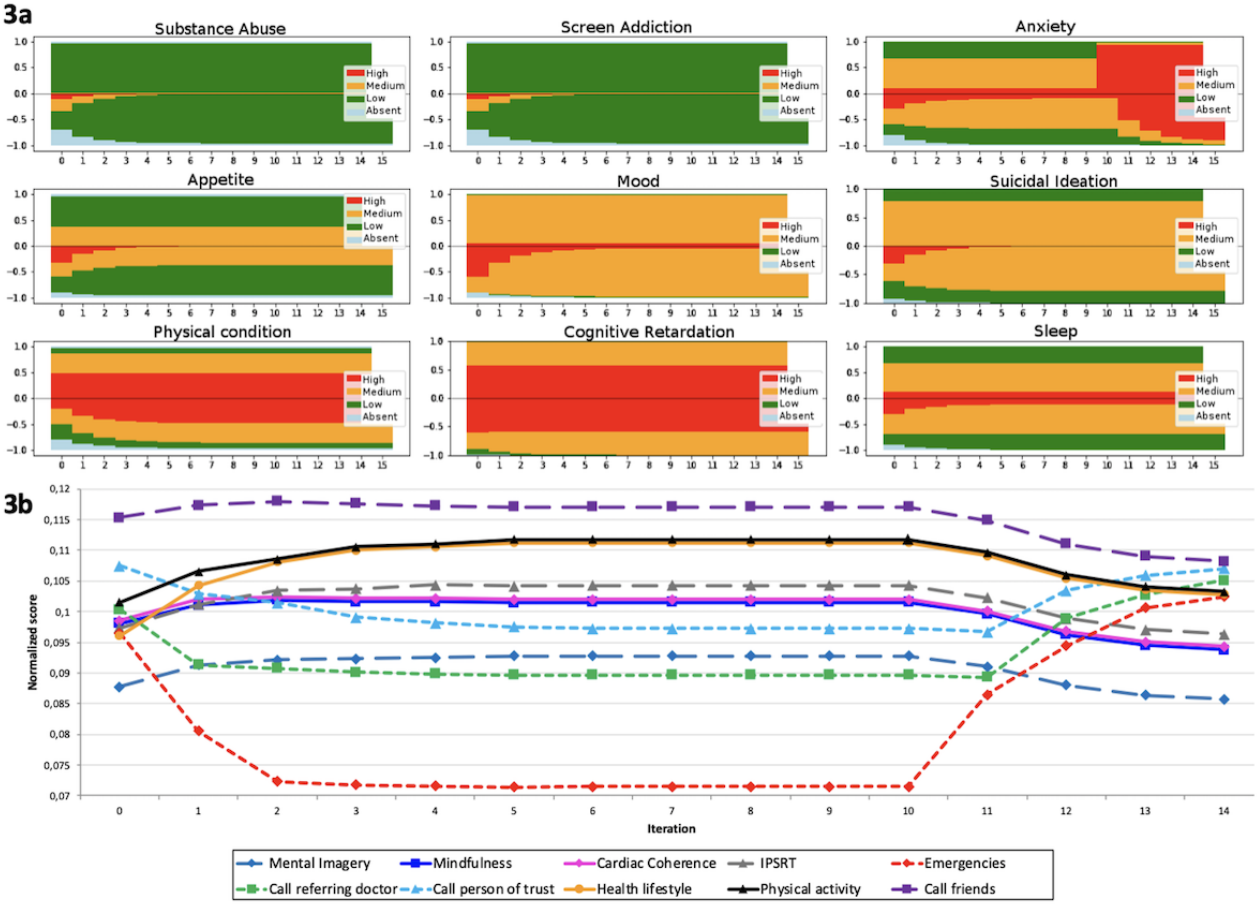


Figure 4. Final run simulation of the probability of advice proposal during incremental updates (immediate worsening after 9 successive lowest scores). IPSRT: interpersonal and social rhythm therapy.

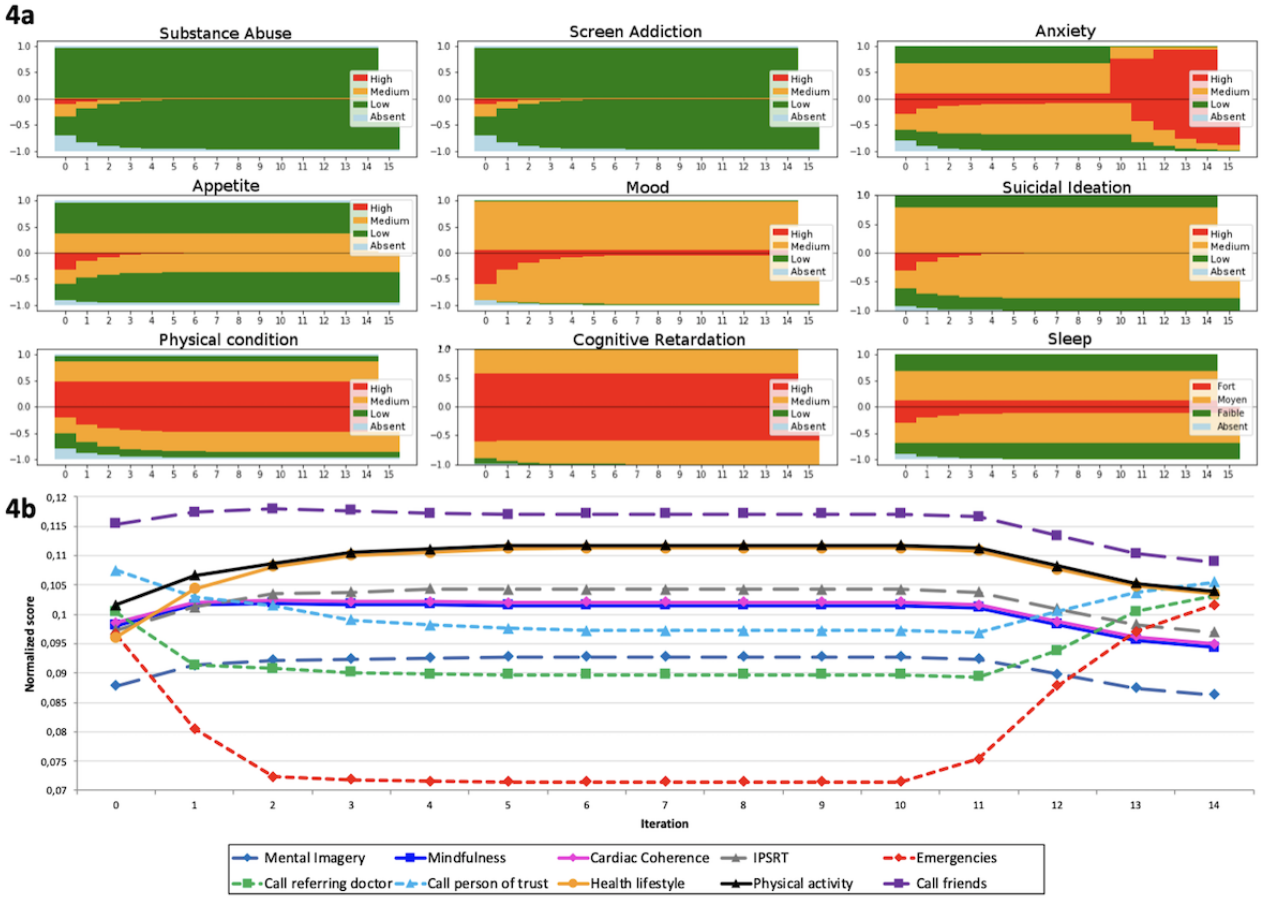


Table 4. Expert-based Bayesian network construct regarding "call friends" advice.

Contextual severity: call friends, social contacts (circle of proximity)	Anxiety		
	High	Medium	Low
Yes	99	80	50
No	1	20	50

Discussion

Principal Findings

Our results are consistent with our hypothesis and strongly suggest that BNs are an interesting model for developing algorithms in highly expert professional fields, especially when data are unavailable.

Our model showed that depending on the clinical dimensions of a patient with a stable clinical state, it will rapidly converge (in three iterations) to a pattern of pertinent advice. In the example described in Figure 3 and Figure 4 for a patient with a constant and low level of anxiety, the pattern of advice converges to a solution that inputs priority advice without caregivers' interventions. Indeed, going to the emergency and calling the referent psychiatrist or a person of trust are not pertinent advice for this level of severity. When the clinical state worsens, the model still demonstrates a rapid adaptation and stabilization to this event. The advice pattern reorders itself pertinently by prioritizing emergencies and caregiver intervention, and other forms of advice such as mindfulness or therapy become less pertinent. These less pertinent types of advice decrease over time when the clinical state worsens (Figure 3 and Figure 4). They are not directly dependent on the severity of anxiety, and their absolute probability stays constant. However, their relative probability (ie, the scores after normalization across every form of advice) decreases over time.

We also see that the weight of some advice, such as mindfulness and cardiac coherence, are the same. This might be explained by the fact that only the weight linked to cognitive retardation is different. Cognitive retardation does not play a role in a scenario on anxiety (Figure 3 and Figure 4), showing that its level of contextual severity stays the same over time and is independent of the level of anxiety.

As we expect, such an a priori expert model is valuable to test preclinical situations, and the results of our simulation with various patient profiles allow us to build a valid BN. The APA points out that application developers often make many claims, while the level of scientific evidence that accompanies these

apps is often quite low [37,46]. In a recent study [47], the authors analyzed the claims of 73 mental health applications (the top-ranked applications from the two largest app stores) for acceptability and efficiency. There was a difference between positive statements about their effectiveness and acceptability. They found that less than 53% of these claims were associated with evidence in the scientific literature, and 33% referred to techniques for which no evidence could be found.

For this purpose, the APA proposes a hierarchical framework of 5 levels to evaluate apps [39]. The first level of the model aims to evaluate the quality of the information referenced by the application to decide if we can consider using the app. Scientific data is very important in this step. Levels 2 and 3 constitute the basic medical decision-making process centered on nonmaleficence. As in any therapeutic intervention or evaluation, apps induce risks that must be scientifically evaluated but are often overlooked; that is the goal of the second level. Level 3 aims to evaluate the necessary scientific evidence, providing us reasons for potential use. Level 4 evaluates the usability of the apps and level 5 tests interoperability.

This framework highlights the need for the scientific development of applications that correspond to the standards of therapeutic research. For this reason, we consider that a preclinical phase is essential before continuing to the clinical phase.

During the design phase of the device, the preclinical step makes it possible to check its performance and safety and will help us evaluate its later acceptability for the patient. By comparison, when it comes to the development of a drug, this phase makes it possible to evaluate a molecule in cultured cells (in vitro) and in animals (in vivo), but in our case, we can talk about evaluation in silico that is often used in recent pharmacological studies [48].

With our Bayesian approach, we can make simulations of patients, considering these to be preclinical tests as recommended by good practices in clinical therapeutic research, and we can increase its scientific validity. Nevertheless, it still is a simple model with only 9 clinical dimensions and 10 types of advice driven by 23 questions. In developing preclinical tests in the future, models will have to be improved with more variables.

There are numerous healthcare apps in mobile app stores. At the moment, very few provide academic validation with state-of-the-art clinical trials. To date, there are only 2 FDA (Food and Drug Administration)-approved or CE (Conformitè Europëenne)-approved apps (reSET@ and Flow) in psychiatry [49]. One provides digitalized depression prevention treatment [50], and the other is for opioid addiction management [51].

Our project is one of the first attempts to build autonomous care support for suicide prevention in a specific at-risk population of patients with previous SA with a methodology in accordance with APA guidelines.

From a larger point of view, the main idea driving this project is that it is possible to build an app that mimics a doctor's care enough to be a valuable (and efficient) add-on. The empowerment of patients is a vital topic, and economic issues are also important.

But most importantly, we are showing that algorithms for clinical and behavioral disturbance in psychiatry might be a crucial field for BN, particularly when there is not enough available data to build such algorithms. For example, there is important scientific literature for anxiety in patients with depression, suicidal ideation, or SA; however, the data are not specific enough to answer very specific questions regarding sleep disturbance according to the level of anxiety in a patient 3 months after SAs. Therefore, expert opinion is crucial.

Such an algorithm, embedded in smartphone apps, could be used by many patients and consequently produce a large amount of new specific data. This data will eventually help us improve our model by incrementally updating the parameters (ie, all the probability distributions) to recognize one given factor better or associate one advice type better to some situations.

Since we will have a lot of usage data, it is important to determine if the recommendations are followed by action and if this type of algorithm increases their relevance. This could be of particular interest for therapeutic compliance or even access to emergency care. The clinical stage of our research, associated with our application's usage data, will allow us to study this question.

Conclusion

We are convinced that using digital devices with efficient algorithms is crucial for medical treatment in terms of reliability and safety. However, to date, very few devices meet accurate methodological requirements. Our work is a proof of concept that emphasizes the need for preclinical trials by algorithm development. Additionally, it shows that BN is an accurate and very efficient branch of AI in psychiatry and clinical psychology.

However, building the application and randomized controlled clinical trials are necessary to confirm our choices and the overall efficacy of our device.

Acknowledgments

We would like to thank Jean Christophe Guilbaud for additional editing.

Conflicts of Interest

None declared.

References

1. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [Medline: 18509902]
2. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
3. Myin-Germeys I, Klippel A, Steinhart H, Reininghaus U. Ecological momentary interventions in psychiatry. Curr Opin Psychiatry 2016 Jul;29(4):258-263. [doi: 10.1097/YCO.0000000000000255] [Medline: 27153125]
4. Torous J, Firth J, Mueller N, Onnela JP, Baker JT. Methodology and Reporting of Mobile Heath and Smartphone Application Studies for Schizophrenia. Harv Rev Psychiatry 2017;25(3):146-154 [FREE Full text] [doi: 10.1097/HRP.0000000000000133] [Medline: 28234658]
5. Versluis A, Verkuil B, Spinhoven P, van DPMM, Brosschot JF. Changing Mental Health and Positive Psychological Well-Being Using Ecological Momentary Interventions: A Systematic Review and Meta-analysis. J Med Internet Res 2016;18(6):e152 [FREE Full text] [doi: 10.2196/jmir.5642] [Medline: 27349305]
6. Tighe J, Shand F, Ridani R, Mackinnon A, De LMN, Christensen H. Ibobbly mobile health intervention for suicide prevention in Australian Indigenous youth: a pilot randomised controlled trial. BMJ Open 2017 Dec 27;7(1):e013518 [FREE Full text] [doi: 10.1136/bmjopen-2016-013518] [Medline: 28132007]
7. Roepke AM, Jaffee SR, Riffle OM, McGonigal J, Broome R, Maxwell B. Randomized Controlled Trial of SuperBetter, a Smartphone-Based/Internet-Based Self-Help Tool to Reduce Depressive Symptoms. Games Health J 2015 Jun;4(3):235-246. [doi: 10.1089/g4h.2014.0046] [Medline: 26182069]
8. Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, et al. CBT for depression: a pilot RCT comparing mobile phone vs. computer. BMC Psychiatry 2013;13:49 [FREE Full text] [doi: 10.1186/1471-244X-13-49] [Medline: 23391304]
9. Ivanova E, Lindner P, Ly KH, Dahlin M, Vernmark K, Andersson G, et al. Guided and unguided Acceptance and Commitment Therapy for social anxiety disorder and/or panic disorder provided via the Internet and a smartphone application: A randomized controlled trial. J Anxiety Disord 2016 Dec;44:27-35. [doi: 10.1016/j.janxdis.2016.09.012] [Medline: 27721123]
10. Faurholt-Jepsen M, Frost M, Ritz C, Christensen EM, Jacoby AS, Mikkelsen RL, et al. Daily electronic self-monitoring in bipolar disorder using smartphones - the MONARCA I trial: a randomized, placebo-controlled, single-blind, parallel group trial. Psychol Med 2015 Oct;45(13):2691-2704. [doi: 10.1017/S0033291715000410] [Medline: 26220802]
11. Depp CA, Ceglowski J, Wang VC, Yaghouti F, Mausbach BT, Thompson WK, et al. Augmenting psychoeducation with a mobile intervention for bipolar disorder: a randomized controlled trial. J Affect Disord 2015 Mar 15;174:23-30. [doi: 10.1016/j.jad.2014.10.053] [Medline: 25479050]
12. Mind Strong. URL: https://mindstrong.com/ [accessed 2021-08-18]
13. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
14. Husky M, Swendsen J, Ionita A, Jaussent I, Genty C, Courtet P. Predictors of daily life suicidal ideation in adults recently discharged after a serious suicide attempt: A pilot study. Psychiatry Res 2017 Oct;256:79-84. [doi: 10.1016/j.psychres.2017.06.035] [Medline: 28624676]
15. Fazel S, Runeson B. Suicide. N Engl J Med 2020 Jan 16;382(3):266-274. [doi: 10.1056/NEJMra1902944] [Medline: 31940700]
16. Suicide data in 2018. World Health Organisation. 2018 Nov 15. URL: http://www.who.int/mental_health/prevention/suicide/suicideprevent/en/ [accessed 2021-08-25]
17. Patton GC, Coffey C, Sawyer SM, Viner RM, Haller DM, Bose K, et al. Global patterns of mortality in young people: a systematic analysis of population health data. Lancet 2009 Sep 12;374(9693):881-892 [FREE Full text] [doi: 10.1016/S0140-6736(09)60741-8] [Medline: 19748397]
18. Burns CD, Cortell R, Wagner BM. Treatment compliance in adolescents after attempted suicide: a 2-year follow-up study. J Am Acad Child Adolesc Psychiatry 2008 Aug;47(8):948-957 [FREE Full text] [doi: 10.1097/CHI.Ob013e3181799e84] [Medline: 18596554]
19. Zalsman G, Hawton K, Wasserman D, van HK, Arensman E, Sarchiapone M, et al. Suicide prevention strategies revisited: 10-year systematic review. Lancet Psychiatry 2016 Jul;3(7):646-659. [doi: 10.1016/S2215-0366(16)30030-X] [Medline: 27289303]
20. Vaiva G, Berrouiguet S, Walter M, Courtet P, Ducrocq F, Jardon V, et al. Combining Postcards, Crisis Cards, and Telephone Contact Into a Decision-Making Algorithm to Reduce Suicide Reattempt: A Randomized Clinical Trial of a Personalized Brief Contact Intervention. J Clin Psychiatry 2018 Sep 25;79(6):1-13. [doi: 10.4088/JCP.17m11631] [Medline: 30256552]
21. Bzdok D, Meyer-Lindenberg A. Machine Learning for Precision Psychiatry: Opportunities and Challenges. Biol Psychiatry Cogn Neurosci Neuroimaging 2018 Dec;3(3):223-230. [doi: 10.1016/j.bpsc.2017.11.007] [Medline: 29486863]
22. Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature 2015 May 28;521(7553):452-459. [doi: 10.1038/nature14541] [Medline: 26017444]

23. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med 2016 Sep 29;375(13):1216-1219 [FREE Full text] [doi: 10.1056/NEJMp1606181] [Medline: 27682033]
24. Bourla A, Ferreri F, Ogorzelec L, Peretti C, Guinchard C, Mouchabac S. Psychiatrists' Attitudes Toward Disruptive New Technologies: Mixed-Methods Study. JMIR Ment Health 2018 Dec 14;5(4):e10240 [FREE Full text] [doi: 10.2196/10240] [Medline: 30552086]
25. Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. Radiology 2017 Dec;284(2):574-582. [doi: 10.1148/radiol.2017162326] [Medline: 28436741]
26. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer 2018 Dec;18(8):500-510 [FREE Full text] [doi: 10.1038/s41568-018-0016-5] [Medline: 29777175]
27. Contreras I, Vehi J. Artificial Intelligence for Diabetes Management and Decision Support: Literature Review. J Med Internet Res 2018 May 30;20(5):e10775 [FREE Full text] [doi: 10.2196/10775] [Medline: 29848472]
28. Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism 2017 Apr;69S:S36-S40. [doi: 10.1016/j.metabol.2017.01.011] [Medline: 28126242]
29. Ramyead A, Studerus E, Kometer M, Uttinger M, Gschwandtner U, Fuhr P, et al. Prediction of psychosis using neural oscillations and machine learning in neuroleptic-naïve at-risk patients. World J Biol Psychiatry 2016 Jun;17(4):285-295. [doi: 10.3109/15622975.2015.1083614] [Medline: 26453061]
30. Arbabshirani MR, Plis S, Sui J, Calhoun VD. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. Neuroimage 2017 Jan 15;145(Pt B):137-165 [FREE Full text] [doi: 10.1016/j.neuroimage.2016.02.079] [Medline: 27012503]
31. Mechelli A, Lin A, Wood S, McGorry P, Amminger P, Tognin S, et al. Using clinical information to make individualized prognostic predictions in people at ultra high risk for psychosis. Schizophr Res 2017 Jun;184:32-38 [FREE Full text] [doi: 10.1016/j.schres.2016.11.047] [Medline: 27923525]
32. Woo C, Chang LJ, Lindquist MA, Wager TD. Building better biomarkers: brain models in translational neuroimaging. Nat Neurosci 2017 Feb 23;20(3):365-377. [doi: 10.1038/nn.4478] [Medline: 28230847]
33. Dwyer DB, Falkai P, Koutsouleris N. Machine Learning Approaches for Clinical Psychology and Psychiatry. Annu Rev Clin Psychol 2018 May 07;14:91-118. [doi: 10.1146/annurev-clinpsy-032816-045037] [Medline: 29401044]
34. Schnack HG, Kahn RS. Detecting Neuroimaging Biomarkers for Psychiatric Disorders: Sample Size Matters. Front Psychiatry 2016;7:50 [FREE Full text] [doi: 10.3389/fpsyt.2016.00050] [Medline: 27064972]
35. Du Y, Guo Y. Evidence reasoning method for constructing conditional probability tables in a Bayesian network of multimorbidity. Technol Health Care 2015;23 Suppl 1:S161-S167. [doi: 10.3233/thc-150950] [Medline: 26410321]
36. Satterthwaite TD, Xia CH, Bassett DS. Personalized Neuroscience: Common and Individual-Specific Features in Functional Brain Networks. Neuron 2018 Apr 18;98(2):243-245. [doi: 10.1016/j.neuron.2018.04.007] [Medline: 29673476]
37. App evaluation model. American Psychiatric Association. 2019 Nov 5. URL: https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/app-evaluation-model [accessed 2021-08-25]
38. Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Massachusetts, United States: Morgan Kaufmann; 1988.
39. Kjærulff U, Madsen A. Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis. New York: Springer-Verlag; 2013:A.
40. O?Hagan A, Buck C, Daneshkhah A, Eiser J, Garthwaite P, Jenkinson D. Uncertain Judgements: Eliciting Experts' Probabilities. Chichester: John Wiley & Sons; 2006:E.
41. Veen D, Stoel D, Zondervan-Zwijnenburg M, van de Schoot R. Proposal for a Five-Step Method to Elicit Expert Judgment. Front Psychol 2017 Dec 05;8:1-11. [doi: 10.3389/fpsyg.2017.02110]
42. Bentley KH, Franklin JC, Ribeiro JD, Kleiman EM, Fox KR, Nock MK. Anxiety and its disorders as risk factors for suicidal thoughts and behaviors: A meta-analytic review. Clinical Psychology Review 2016 Feb;43:30-46. [doi: 10.1016/j.cpr.2015.11.008]
43. Calati R, Nemeroff C, Lopez-Castroman J, Cohen L, Galynker I. Candidate Biomarkers of Suicide Crisis Syndrome: What to Test Next? A Concept Paper. Int J Neuropsychopharmacol 2020;23(3):192-205. [doi: 10.1093/ijnp/pyz063]
44. Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, et al. The Columbia–Suicide Severity Rating Scale: Initial Validity and Internal Consistency Findings From Three Multisite Studies With Adolescents and Adults. AJP 2011 Dec;168(12):1266-1277. [doi: 10.1176/appi.ajp.2011.10111704]
45. Kjærulff U, Madsen A. Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis. New-York: Springer-Verlag; 2013:A.
46. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. Journal of Psychiatric Research 2018 Dec;107:73-78. [doi: 10.1016/j.jpsychires.2018.10.006]
47. Larsen ME, Nicholas J, Christensen H. A Systematic Assessment of Smartphone Tools for Suicide Prevention. PLoS One 2016;11(4):e0152285 [FREE Full text] [doi: 10.1371/journal.pone.0152285] [Medline: 27073900]

48. Kazmi SR, Jun R, Yu M, Jung C, Na D. In silico approaches and tools for the prediction of drug metabolism and fate: A review. Comput Biol Med 2019 Mar;106:54-64. [doi: 10.1016/j.compbiomed.2019.01.008] [Medline: 30682640]
49. healthskouts. FDA/CE certified apps directory. FDA/CE certified apps directory. 2019. URL: https://apps.healthskouts.com [accessed 2021-08-25]
50. The Depression App. Flow Neuroscience. URL: https://flowneuroscience.com/home/app/ [accessed 2021-08-18]
51. Redefining Medicine: Prescription Digital Therapeutics for the Treatment of Serious Disease. Pear Therapeutics. URL: https://peartherapeutics.com/ [accessed 2021-08-18]

Abbreviations

AI: artificial intelligence
APA: American Psychiatric Association
CF: cumulated value of factor
CPD: conditional probability distribution
CPT: conditional probability table
CS: contextual severity
DAG: directed acyclic graph
EMA: ecological momentary assessment
EMI: ecological momentary intervention
IF: immediate value of factor
NOM: noisy-or model
SA: suicide attempt
VAS: visual analog scale

Edited by R Kukafka; submitted 24.09.20; peer-reviewed by S Steingrimsson, T Yasaka; comments to author 21.12.20; revised version received 11.02.21; accepted 16.03.21; published 30.09.21.

Please cite as:
Mouchabac S, Leray P, Adrien V, Gollier-Briant F, Bonnot O
Prevention of Suicidal Relapses in Adolescents With a Smartphone Application: Bayesian Network Analysis of a Preclinical Trial Using In Silico Patient Simulations
J Med Internet Res 2021;23(9):e24560
URL: https://www.jmir.org/2021/9/e24560
doi:10.2196/24560
PMID:34591030

Original Paper

Radiation Oncologists' Perceptions of Adopting an Artificial Intelligence–Assisted Contouring Technology: Model Development and Questionnaire Study

Huiwen Zhai[1,2*], MBA; Xin Yang[3*], PhD; Jiaolong Xue[4,5], PhD; Christopher Lavender[1], MA; Tiantian Ye[6], MD; Ji-Bin Li[7], PhD; Lanyang Xu[2], BPhil; Li Lin[3], MD, PhD; Weiwei Cao[8], MBA; Ying Sun[3], MD, PhD

[1]Office of Research Management and Education Administration, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

[2]Department of Anthropology, School of Sociology and Anthropology, Sun Yat-sen University, Guangzhou, China

[3]Department of Radiation Oncology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

[4]School of Management, Sun Yat-sen University, Guangzhou, China

[5]School of Management, Guangdong Ocean University, Zhanjiang, China

[6]School of Medicine, Southern University of Science and Technology, Shenzhen, China

[7]Department of Clinical Research, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

[8]Management Office of Huangpu Campus, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Ying Sun, MD, PhD
Department of Radiation Oncology
Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China
Collaborative Innovation Center for Cancer Medicine
651 Dongfeng Road
Guangzhou, 510060
China
Phone: 86 02087343066
Email: sunying@sysucc.org.cn

Abstract

Background: An artificial intelligence (AI)–assisted contouring system benefits radiation oncologists by saving time and improving treatment accuracy. Yet, there is much hope and fear surrounding such technologies, and this fear can manifest as resistance from health care professionals, which can lead to the failure of AI projects.

Objective: The objective of this study was to develop and test a model for investigating the factors that drive radiation oncologists' acceptance of AI contouring technology in a Chinese context.

Methods: A model of AI-assisted contouring technology acceptance was developed based on the Unified Theory of Acceptance and Use of Technology (UTAUT) model by adding the variables of perceived risk and resistance that were proposed in this study. The model included 8 constructs with 29 questionnaire items. A total of 307 respondents completed the questionnaires. Structural equation modeling was conducted to evaluate the model's path effects, significance, and fitness.

Results: The overall fitness indices for the model were evaluated and showed that the model was a good fit to the data. Behavioral intention was significantly affected by performance expectancy (β=.155; *P*=.01), social influence (β=.365; *P*<.001), and facilitating conditions (β=.459; *P*<.001). Effort expectancy (β=.055; *P*=.45), perceived risk (β=−.048; *P*=.35), and resistance bias (β=−.020; *P*=.63) did not significantly affect behavioral intention.

Conclusions: The physicians' overall perceptions of an AI-assisted technology for radiation contouring were high. Technology resistance among Chinese radiation oncologists was low and not related to behavioral intention. Not all of the factors in the Venkatesh UTAUT model applied to AI technology adoption among physicians in a Chinese context.

(J Med Internet Res 2021;23(9):e27122) doi:10.2196/27122

KEYWORDS

artificial intelligence; technology acceptance model; intension; resistance

Introduction

Background

The artificial intelligence (AI) race has long since begun; its silicone roots date back to 1956 [1]. In China's "Development Plan for The New Generation of Artificial Intelligence," the goal was to synchronize China's overall technology and AI applications with the world's advanced level of technology by 2020. The total value of China's AI industry market is expected to reach US $127 billion by 2025 [2], and the medical AI industry is forecasted to account for one-fifth of the overall AI market [3]. In the health care industry, AI has been developed for administrative assistance, clinical decision-making support, patient monitoring, and robotic surgeries [4]. In clinical decision-making support, computer software is used to analyze clinical data and assist medical professionals in making more precise decisions on disease diagnoses and treatments [5]. Sketching tumors is one such subfield.

Delineating tumors is tricky and requires much education, guidance, and expertise. For radiation oncologists, the tumor target contouring process for precision head and neck radiation therapy is time consuming. It requires, on average, 180 minutes of intense work for each patient, and contouring areas that are created by different radiation oncologists for the same tumor can be highly variable [6]. Moreover, contouring accuracy is key for achieving successful treatment outcomes. Suboptimal tumor coverage and poor-quality radiation therapy plans are major factors of disease relapse and inferior survival [7]. An AI contouring tool that automates the primary gross tumor volume contouring process for patients with cancer was developed at Sun Yat-sen University Cancer Center, wherein it has been applied clinically since 2019; over 12,500 cases of radiation planning have been carried out with the tool's assistance. The contouring system has a high level of accuracy (ie, when compared with the manual contouring accuracy of radiation oncologists). Although oncologists still need to edit automated contours by making minor adjustments, the system substantially saves time by up to 39.4%, and intraobserver and interobserver variations have been reduced by 36.4% and 54.5%, respectively [8].

Although the benefits of AI assistance appear to be obvious, like all forms of change, the adoption of AI technology might be met with resistance, especially within health care settings; it has been said that the only constant is change [9]. There is much hope and fear surrounding such technologies, and this fear can manifest as resistance from both health care professionals and patients [10]. One particular fear that may promote resistance is the fear that AI tools will replace workers (even skilled workers like doctors)—the much discredited, but still prevalent, Luddite fallacy [11]. In a *British Medical Journal* head-to-head debate, the topic was as follows: "Could machines using artificial intelligence make doctors obsolete?" Dr Jörg Goldhahn said "yes" and foreboded that humans simply cannot keep up with the data crunching advantage that computers have over humans [12]. Further, Vanessa Ramptom and Professor Giatgen Spinas claimed that the role of doctors will not become obsolete; instead, these roles will change to specifically focus on treating a patient as a whole [12]. These perspectives encompass part of the larger narratives that describe AI technology as either utopian or dystopian.

There are also claims that doctors are slow to uptake new technologies. In 2004, there was evidence of resistance across the United States against using a computerized order entry system to overcome issues, such as doctors' illegible handwriting, and reduce the number of incorrect prescriptions. The Cedars-Sinai Medical Center notably forced the withdrawal of the system, which was installed in two-thirds of the 870-bed hospital [13]. Similarly, in 2006, Bhattacherjee and Hikmet [14,15] found that physicians were resistant to adopting administrative health care information technologies, and more recently, in 2014, Hsieh [16] discovered 5 factors (attitude, subjective norm, perceived behavior control, institutional trust, and perceived risk) that influenced the adoption of such technologies. Notably, these studies all concerned administrative technologies. There is, however, also evidence of doctors resisting AI technology for other uses, such as patient diagnosis and treatment plan development.

Even though the National Institute of Health claims that AI technology is trending in cancer care [17], a number of notable AI projects have failed in the United States. The MD Anderson Cancer Center, one of the top cancer centers in the world, cancelled a project that involved IBM Corporation's oncology expert advisor tool, Watson, in 2016 after spending US $62 million [18]. In February 2021, IBM Corporation also tried to sell their Watson Health business, as it was not profitable [19].

In this study, we assessed whether there has been any resistance to the uptake of an AI contouring system in China and examined the system's pros and cons. Few prior studies have focused on the technology acceptance of AI tools that provide clinical decision-making support in cancer care from physicians' perspectives. AI health care researchers have primarily analyzed technical innovation and clinical results but have seldom considered the human contexts that are invariably involved in use behavior. Social and implemental issues need to be addressed when an AI tool is introduced in hospitals. Further, it is especially important to examine the extent of perceived risks and technology resistance bias.

Objective

The objective of this study was to develop and test a model for investigating the factors that drive Chinese radiation oncologists' acceptance of an AI contouring tool for automating the primary gross tumor volume and normal tissue contouring process. This study analyzes how expectancy, social influence, facilitating conditions, perceived risk, and resistance bias contribute to the Technology Acceptance Model (TAM).

Theoretical Background and Hypothesis Development

The theoretical models in this study included the TAM and the Unified Theory of Acceptance and Use of Technology (UTAUT) model. The perceived risk and status quo bias theory variables were added to the UTAUT model.

The first TAM was developed by Fred D Davis in 1989, and it was based on the theory of reasoned action. The TAM was used to analyze consumers' behaviors in the context of health information technology based on the following two major variables: perceived usefulness and perceived ease of use [20]. Later, in 2000, Venkatesh and Davis [21] developed the TAM2, which also addressed perceived usefulness and use intentions in terms of social influence and cognitive instrumental processes. They added the subjective norm, voluntariness, image, job relevance, output quality, and result demonstrability constructs, which significantly influenced user acceptance. The evolution of the TAM3 came next; the TAM3 was a comprehensive nomological network (ie, an integrated model) for evaluating individual-level (information technology) adoption and use [22].

In 2003, Venkatesh et al [23] also created a UTAUT model, which is an integrated tool based on 8 original models that were published in major information systems journals from a range of disciplines. These models included the theory of reasoned action model, the TAM, the motivation model, the theory of planned behavior model, a combination of the TAM and the theory of planned behavior model, the model of PC utilization, the innovation diffusion theory model, and the social cognitive theory model. The UTAUT originally explained technology acceptance among employees, and in 2012, a UTAUT2 model was developed, which focused on consumer technologies [24]. As AI-assisted contouring technology has been developed for medical staff in hospitals, hedonic motivation and price value, which were explored in the UTAUT2 model, were considered inapplicable to explaining oncologists' acceptance behaviors in this study. Therefore, we used the original UTAUT model in this study. The original model and its various modifications have been popularly used in studies of technology adoption in the health care industry over the past 5 years [25-28].

The UTAUT model revealed the following four variables: performance expectancy, effort expectancy, social influence (which is used to evaluate the behavioral intention to use a new technology), and facilitating conditions (which are used to evaluate use behavior). Most research has proven that these four constructs positively affect potential users' behavioral intentions [29,30]. Performance expectancy is defined as "the degree to which an individual believes that using a new technology will help him or her to attain gains in job performance." This construct was integrated from a similar concept in the TAM—perceived usefulness [31]. Effort expectancy is defined as "the degree of ease of use associated with IT" [23]. This concept is similar to the perceived ease of use construct in the TAM and innovation diffusion theory and the complexity of technology construct in the model of PC utilization. Social influence is defined as "the degree to which an individual perceives the importance of how others believe that he or she should use new IT." This concept was developed based on the notion that "an individual's behavior is influenced by the way in which one believes others will view him/her as a result of having used health IT" [23]. The Venkatesh model integrated subjective norms from the theory of reasoned action, the TAM2, and the theory of planned behavior; social factors from the model of PC utilization; and the image construct from the innovation diffusion theory into the social influence construct. The facilitating conditions construct is defined as "the degree to which an individual believes that an organizational and technical infrastructure exists to support the use of the system" [23]. The UTAUT's facilitating conditions concept contains the perceived behavior control concept from the theory of planned behavior and the combined TAM and theory of planned behavior model, the facilitating conditions concept from the model of PC utilization, and the compatibility construct from the innovation diffusion theory [23]. Lastly, the intention-usage behavior relationship has long been discussed in many studies of technology acceptance, and we expected that behavioral intentions to use AI-assisted technology would have a significant influence on use behavior. Based on these variables, we proposed the following hypotheses:

- Hypothesis 1: Performance expectancy positively affects oncologists' behavioral intentions to use an AI contouring tool for developing radiation therapy plans.
- Hypothesis 2: Effort expectancy positively affects oncologists' behavioral intentions to use an AI contouring tool for developing radiation therapy plans.
- Hypothesis 3: Social influence positively affects oncologists' behavioral intentions to use an AI contouring tool for developing radiation therapy plans.
- Hypothesis 4: Oncologists' behavioral intentions positively affect their use behaviors when using an AI contouring tool for developing radiation therapy plans.
- Hypothesis 5A: Facilitating conditions positively affect oncologists' use behavior when using an AI contouring tool for developing radiation therapy plans.
- Hypothesis 5B: Facilitating conditions positively affect oncologists' behavioral intentions to use an AI contouring tool for developing radiation therapy plans.

Perceived risk has long been discussed as a negative factor in health care information technology adoption studies [10,16], and it refers to the combination of uncertainty and the seriousness of an outcome [32]. Featherman and Pavlou [33] identified the following seven facets of e-service–related perceived risk: performance risk, financial risk, time risk, physiological risk, social risk, privacy risk, and overall risk. In this study, we addressed perceived risk based on the following four aspects: (1) performance risk, which we defined as the probability of malfunction and performance failure and whether the AI contouring tool could mislead physicians with inaccurate contouring; (2) time risk, which we defined as the probability of physicians needing more time to correct errors caused by the AI system; (3) psychological risk, which we analyzed in terms of whether using the AI contouring tool may cause psychological discomfort due to having a negative effect on clinicians' self-perceptions of the treatment plan; and (4) privacy risk, which we defined as a potential loss of the confidential patient data in the AI-assisted system. As the AI systems were

purchased by public hospitals without input from individual physicians, no financial or social risk assessment was required.

Resistance to change refers to people's attempts to maintain previous behaviors or habits that are connected to their past experiences when facing changes [15,34,35], and this has been discussed as a major barrier to health care information technology adoption [15,36]. The fear of job loss has also been discussed as a factor of the slow adoption of new information technologies [37]. We suggest that radiation oncologists might believe that AI-assisted technologies might replace them and that the adoption of such technologies will lead to an increase in unemployment rates. Thus, we included these two factors in the resistance bias construct, which we defined as people's resistance to using a new technology resulting from their resistance to change and concerns about unemployment. Therefore, we proposed the following hypotheses:

- Hypothesis 6: Perceived risk negatively affects oncologists' behavioral intentions to use an AI contouring tool for developing radiation therapy plans.
- Hypothesis 7: Resistance bias negatively affects oncologists' behavioral intentions to use an AI contouring tool for developing radiation therapy plans.

Methods

Participants and Sampling

The majority of participants were radiation oncologists (227/307, 73.9%) and medical students (80/307, 26.1%) from the Department of Radiation Oncology at Sun Yat-sen University Cancer Center who had clinical experience in using the computational system for contouring. The sample pool also included radiation oncologists from other hospitals in China. Participants were recruited between July 2020 and November 2020. A convenience sampling method was used in this study, as only a few hospitals in China have AI-assisted contouring systems. As our survey had 29 items, the required number of participants was more than 290. The questionnaire was conducted by sending direct website links via email or WeChat (a social media platform developed by Tencent Holding Limited). Questionnaires were only considered valid if (1) an account responded only once and (2) the total response time for completing the questionnaire was more than 60 seconds. In total, 322 radiation oncologists replied to the web-based survey. After filtering out the invalid responses, 307 responses—176 from Sun Yat-sen University Cancer Center and 131 from other hospitals—were deemed valid for data analysis.

Measurements

Our model included 8 constructs with 29 questionnaire items. Each item assessed 1 construct and was sourced from relevant literature related to new technology acceptance research. Some changes were made to fit the AI-assisted contouring system context (Table 1). After we developed a list of questions, a team discussion was held to reach a consensus. Afterward, we asked a group of 10 oncologists with experience in working with the AI system to answer the questions and provide comments for system improvement. A 7-point Likert scale, which ranged from 1 (strongly disagree) to 7 (strongly agree), was used to assess the items for all constructs except those for the use behavior construct, in which the answers were "yes" or "no."

The first page of the questionnaire provided an overview of this study's background, purpose, voluntary nature, and anonymity and an introduction of the AI-assisted tumor volume contouring system, which included information on its general functions and operating procedures and images of the system to help the participants gain a clear understanding. We developed 2 questionnaires to assess oncologists' perceptions. All participants first answered the following question: "Have you used the AI-assisted contouring system?" If the answer was "yes," then they were directed to answer the questions for those who have already used the system (Table 1); if the answer was "no," then they were directed to answer the questions for those who have not used the system. The questionnaires were identical apart from the usage of verb tenses. The questions were translated from English to Mandarin Chinese for the survey. Ethical approval was obtained from the Ethics Committee of the Department of Anthropology at Sun Yat-sen University.

Table 1. Questionnaire items related to the research model constructs.

Variables and items	Questionnaire items	Sources
PE[a,b]		
PE1	• I find the automated contouring system useful in my job.[c] • I think the automated contouring system will be useful in my job.[d]	Venkatesh et al [23] and Kijsanayotin et al [25]
PE2	• Using the automated contouring system enables me to accomplish tasks quicker.[c] • I think using the automated contouring system will enable me to accomplish tasks quicker.[d]	Venkatesh et al [23] and Kijsanayotin et al [25]
PE3	• Using the automated contouring system increases my productivity.[c] • I think using the automated contouring system will increase my productivity.[d]	Venkatesh et al [23] and Kijsanayotin et al [25]
PE4	• Using the automated contouring system improves the outcomes of my work.[c] • I think using the automated contouring system will improve the outcomes of my work.[d]	Venkatesh et al [23] and Kijsanayotin et al [25]
EE[a,e]		
EE1	• The automated contouring system is clear and understandable to me.[c] • I think the automated contouring system will be clear and understandable to me.[d]	Venkatesh et al [23] and Kijsanayotin et al [25]
EE2	• It was easy for me to become skillful at using the automated contouring system.[c] • I think it will be easy for me to become skillful at using the automated contouring system.[d]	Venkatesh et al [23] and Kijsanayotin et al [25]
EE3	• I find the automated contouring system easy to use.[c] • I think the automated contouring system will be easy to use.[d]	Venkatesh et al [23] and Kijsanayotin et al [25]
EE4	• Learning to operate the automated contouring system is easy for me.[c] • I think learning to operate the automated contouring system will be easy for me.[d]	Venkatesh et al [23] and Kijsanayotin et al [25]
SI[a,f]		
SI1	• People (eg, colleagues and friends) who influence my behavior think that I should use the automated contouring system.[c,d]	Venkatesh et al [23] and Kijsanayotin et al [25]
SI2	• People who are important to me (eg, department heads, tutors, superiors, and hospital leaders) think that I should use the automated contouring system.[c,d]	Venkatesh et al [23] and Kijsanayotin et al [25]
SI3	• Social propaganda has been helpful in the use of the automated contouring system.[c,d]	Venkatesh et al [23] and Kijsanayotin et al [25]
SI4	• In general, my hospital and my department have supported the use of the automated contouring system.[c,d]	Venkatesh et al [23] and Kijsanayotin et al [25]
FC[a,g]		
FC1	• I have the resources (devices, support from colleagues, etc) necessary to use the automated contouring system.[c,d]	Venkatesh et al [23] and Kijsanayotin et al [25]
FC2	• I have the knowledge (the clinical and computer skills) necessary to use the automated contouring system.[c,d]	Venkatesh et al [23] and Kijsanayotin et al [25]
FC3	• Health and information technology personnel in the hospital are available to assist with system difficulties.[c,d]	Venkatesh et al [23] and Kijsanayotin et al [25]
FC4	• I have adequate knowledge resources (eg, books, documents, and consultants) to help me learn about the automated contouring system.[c,d]	Venkatesh et al [23] and Kijsanayotin et al [25]

Variables and items	Questionnaire items	Sources
BI[a,h]		
BI1	• I intend to use the automated contouring system in the next 2 months.[c,d]	Venkatesh et al [23] and Kijsanayotin et al [25]
BI2	• I will use the automated contouring system in the next 2 months.[c,d]	Venkatesh et al [23] and Kijsanayotin et al [25]
BI3	• I plan to use the automated contouring system in the next 2 months.[c,d]	Venkatesh et al [23] and Kijsanayotin et al [25]
PR[a,i]		
PR1	• There is a possibility of malfunction and performance failure, so the system might fail to deliver accurate contouring areas and could mislead my work with inaccurate contouring.[c,d]	Ye et al [10], Bhattacherjee and Hikmet [15], and Andrews et al [38]
PR2	• There is a probability that I need more time to fix the errors and nuances of the artificial intelligence system.[c,d]	Ye et al [10], Bhattacherjee and Hikmet [15], and Andrews et al [38]
PR3	• I think using the automated contouring system may cause psychological distress, as it could have a negative effect on my self-perception of the treatment plan.[c,d]	Ye et al [10], Bhattacherjee and Hikmet [15], and Andrews et al [38]
PR4	• I am concerned that my patients' personal information and health details are/would be insecure and could be accessed by stakeholders or unauthorized persons, leading to lawsuits for physicians and the hospital.[c,d]	Ye et al [10], Bhattacherjee and Hikmet [15], and Andrews et al [38]
RB[a,j]		
RB1	• I do not want the automated contouring system to change how I develop my treatment plan for manual contouring because the new system is unfamiliar to me.[c,d]	Bhattacherjee and Hikmet [15] and Dou et al [34]
RB2	• I do not want to use the automated contouring system because of past experience; these new high-tech products always fall flat during practical application.[c,d]	Bhattacherjee and Hikmet [15] and Dou et al [34]
RB3	• I do not want to use the automated contouring system because there is a possibility of losing my job, as artificial intelligence–assisted technology may do my work better than me.[c,d]	Raza et al [37]
UB[k,l]		

Variables and items	Questionnaire items	Sources
UB1	• I have already used the automated contouring system.[c] • I am preparing to use the automated contouring system.[d]	Venkatesh et al [23]
UB2	• I recommend that others should use the automated contouring system.[c,d]	Venkatesh et al [23]
UB3	• Have you ever overridden the contouring system after using it for some time?[c] • Would you ever override the contouring system?[b]	Venkatesh et al [23]

[a]7-point scale (1=strongly disagree; 2=quite disagree; 3=slightly disagree; 4=neither agree or disagree; 5=slightly agree; 6=quite agree; 7=strongly agree).

[b]PE: performance expectancy.

[c]Questions for those who have already used the system.

[d]Questions for those who have not used the system.

[e]EE: effort expectancy.

[f]SI: social influence.

[g]FC: facilitating conditions.

[h]BI: behavioral intention.

[i]PR: perceived risk.

[j]RB: resistance bias.

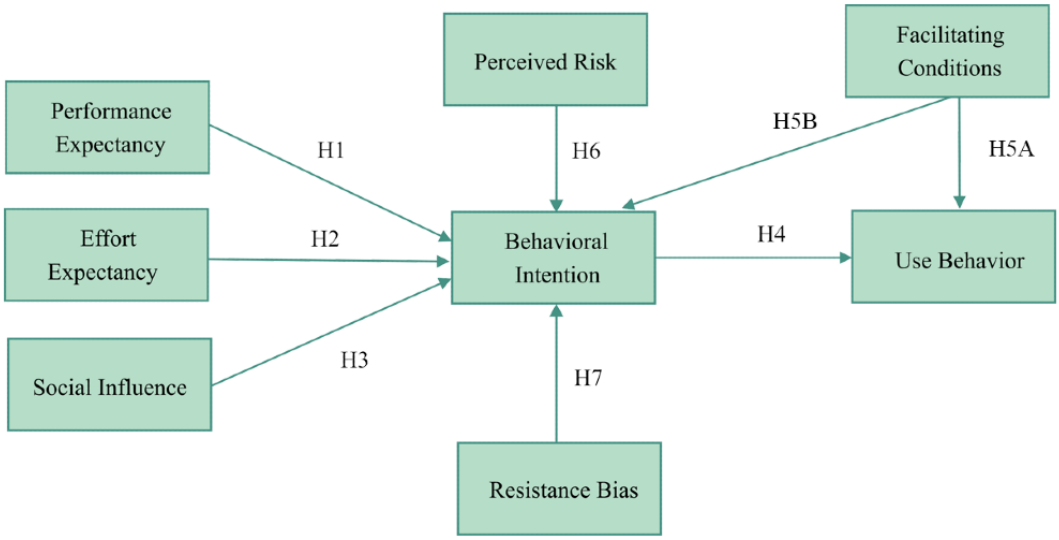
[k]2-point scale (1=yes; 2=no).

[l]UB: use behavior.

Data Analysis

Structural equation modeling analysis was conducted by using IBM SPSS Amos 21.0 (IBM Corporation), and this was a 2-step analysis [39]. First, items and constructs were evaluated for reliability and validity via a confirmatory factor analysis of the measurement model. Second, the structural model's path effects, significance, and goodness of fit and mediation effects were evaluated. More specifically, the path analysis was performed by conducting a series of multiple regression analyses to derive the standardized regression coefficients (ie, β) for all proposed paths. Figure 1 shows the conceptual model for the analysis; the arrows with solid lines are the paths indicating the relationships among the specified concepts in the hypothesis. IBM SPSS Version 25.0 (IBM Corporation) was used for the descriptive statistics analysis.

Figure 1. The development of our model for investigating radiation oncologists' acceptance of an artificial intelligence–assisted contouring system. H: hypothesis.



Results

Demographic Results

The demographic characteristics of the 307 participants are shown in Table 2. About an equal number of men (151/307, 49.2%) and women (156/307, 50.8%) participated in this study. The majority (269/307, 87.6%) of the participants' ages ranged between 18 and 40 years, and all of the participants had at least a bachelor's degree. In China, radiation oncologists who work on contouring are often aged under 40 years, as senior oncologists mainly focus on checking the accuracy of contouring based on the treatment plans, which are mostly developed by

junior oncologists. Thus, it was unsurprising that we had more participants aged under 40 years. In our study, 185 participants had at least 4 years of experience in radiation oncology, 123 participants had already used the AI-assisted contouring technology, and 184 had not yet used it.

Table 2. Demographic characteristics (N=307).

Characteristics	Participants, n (%)
Sex	
Male	151 (49.2)
Female	156 (50.8)
Age (years)	
18-25	121 (39.4)
26-30	67 (21.8)
31-40	81 (26.4)
41-50	30 (9.8)
>50	8 (2.6)
Education	
Bachelor's degree	137 (44.6)
Master's degree or doctoral degree	170 (55.4)
Experience in radiation oncology (years)	
≤3	122 (39.7)
4-6	61 (19.9)
7-10	23 (7.5)
>10	101 (32.9)
Artificial intelligence–assisted contouring technology usage	
Yes	123 (40.1)
No	184 (59.9)

Measurement Model

To test the reliability and validity of this study, we examined common method bias, as all items in our survey were collected from the same questionnaire. We connected all of the items to a common latent factor and used a confirmatory factor analysis to conduct a single-factor test [40]. The results showed that the fit of the single-factor model was poor (χ^2_{377}=4294.616; χ^2/df=11.392; adjusted goodness of fit index=0.354; comparative fit index=0.628; Tucker–Lewis index=0.599; root mean square error of approximation=0.184), indicating that common method bias was insignificant in this study.

We also conducted some tests to examine the constructs' reliability, and the results are summarized in Table 3. The Cronbach α and composite reliability values were higher than the threshold of 0.7 [41], which indicated good construct reliability. For convergent validity, all item loadings ranged from 0.633 to 0.980, and good composite reliability (>0.7) was achieved [42].

Convergent validity was assessed with average variance extracted (AVE) values; all construct AVE values were greater than 0.5 [43] and thus showed good convergent validity. In Table 4, the square roots of the AVE values were higher than the numbers in the off-diagonal direction (correlations between a particular construct in the same column and other constructs in different rows) in the corresponding columns, indicating the good discrimination validity of all constructs. The results of a single-factor analysis of variance showed that the main effect that experience had on behavioral intention was not significant ($F_{3,303}$=2.587; P=.053).

The mean scores for behavioral intention (5.484 out of 7.0) and use behavior (1.713 out of 2.0) were high, which showed that the participants had high intentions to use the AI-assisted contouring system. However, even though the total degree of resistance was low, 17.6% (54/307) of the respondents showed a degree of resistance (ranked as an average score of ≥5 for resistance bias). Specifically, 13 out of the 123 (10.6%) participants who had accessed the AI system showed a degree of resistance, and 41 out of the 184 (22.3%) participants who had not accessed the AI system also showed a degree of resistance.

Table 3. Descriptive statistics of variables, items, composite reliability, and convergent validity.

Variable and item	Item loadings	Mean score of the items	Composite reliability	Convergence validity	Cronbach α
PE[a]		5.969	0.976	0.910	.976
PE1	0.946				
PE2	0.970				
PE3	0.957				
PE4	0.942				
EE[b]		5.669	0.970	0.891	.970
EE1	0.921				
EE2	0.933				
EE3	0.967				
EE4	0.954				
SI[c]		5.481	0.938	0.790	.937
SI1	0.909				
SI2	0.906				
SI3	0.829				
SI4	0.909				
PR[d]		4.418	0.826	0.544	.824
PR1	0.633				
PR2	0.735				
PR3	0.783				
PR4	0.788				
RB[e]		3.528	0.925	0.805	.825
RB1	0.865				
RB2	0.941				
RB3	0.883				
FC[f]		5.446	0.929	0.766	.926
FC1	0.821				
FC2	0.886				
FC3	0.899				
FC4	0.892				
BI[g]		5.484	0.980	0.942	.980
BI1	0.952				
BI2	0.980				
BI3	0.979				
UB[h]		1.713	0.820	0.604	.800
UB1	0.736				
UB2	0.860				
UB3	0.729				

[a]PE: performance expectancy.

[b]EE: effort expectancy.

[c]SI: social influence.

[d]PR: perceived risk.

[e]RB: resistance bias.

[f]FC: facilitating conditions.

[g]BI: behavioral intention.

[h]UB: use behavior.

Table 4. Discriminant validity of the constructs. The square roots of the average variance extracted values for comparisons among constructs are shown.[a]

Constructs	Performance expectancy	Effort expectancy	Social influence	Facilitating conditions	Perceived risk	Resistance bias	Behavioral intention	Use behavior
Performance expectancy	0.954	—[b]	—	—	—	—	—	—
Effort expectancy	0.767	0.944	—	—	—	—	—	—
Social influence	0.717	0.784	0.889	—	—	—	—	—
Facilitating conditions	0.665	0.712	0.716	0.875	—	—	—	—
Perceived risk	−0.334	−0.378	−0.358	−0.314	0.738	—	—	—
Resistance bias	−0.373	−0.435	−0.360	−0.368	0.563	0.897	—	—
Behavioral intention	0.724	0.750	0.782	0.802	−0.369	−0.394	0.971	—
Use behavior	0.600	0.644	0.611	0.630	−0.326	−0.424	0.785	0.777

[a]The items on the diagonal represent the square root of the average variance extracted value; the off-diagonal elements are correlation estimates.

[b]Not applicable.

Structural Model Analysis

The fitness of the model was evaluated with the following criteria: (1) the standardized root mean square residual (0.0317) had to be smaller than 0.08, (2) the comparative fit index (0.968) had to be greater than 0.90, and (3) the root mean square error of approximation (0.056) had to also be smaller than 0.08. As shown in Table 5, most of the indicators satisfied the criteria and the combination rule [44], which demonstrated that the hypothesized model was a good fit to the data.

Figure 2 shows a graphic description of the regression analysis, and Table 6 shows the numerical results of the path coefficient analysis. Standardized factor loadings (ie, β) were the standardized regression coefficients for each proposed path, which indicated the relationships among the variables in this study. *P* values were the parameters used to determine the results of the hypothesis test [45]. Behavioral intention was positively affected by performance expectancy (β=.155; *P*=.014), social influence (β=.365; *P*<.001), and facilitating conditions (β=.459; *P*<.001), and use behavior was significantly affected by behavioral intention (β=.235; *P*<.001). Effort expectancy (β=.055; *P*=.450), perceived risk (β=−.048; *P*=.348), and resistance bias (β=−.020; *P*=.634) did not significantly affect behavioral intention, and facilitating conditions (β=.004; *P*=.845) did not significantly affect use behavior. Behavioral intention also moderated the effects that social influence (*P*=.001) and facilitating conditions (*P*<.001) had on use behavior. R^2 values were calculated to assess the validity of the research model; 78.1% of the effects of behavioral intention could be explained by perceived risk, effort expectancy, social influence, facilitating conditions, performance expectancy, and resistance bias, and 75.3% of the effects of use behavior could be explained by behavioral intention and facilitating conditions (Figure 2).

Table 5. Research model fit.

Model fit indices	Criteria	Research model fit values
Chi-square value	The smaller the better	692.543
Degrees of freedom	The larger the better	354
Normed chi-square value (χ^2/df)	1<χ^2/df<3	1.956
Root mean square error of approximation	<0.08	0.056
Standardized root mean square residual	<0.08	0.0317
Comparative fit index	>0.9	0.968
Goodness of fit index	>0.85	0.859
Adjusted goodness of fit index	>0.8	0.827

Figure 2. Estimates from the regression analysis. H: hypothesis.

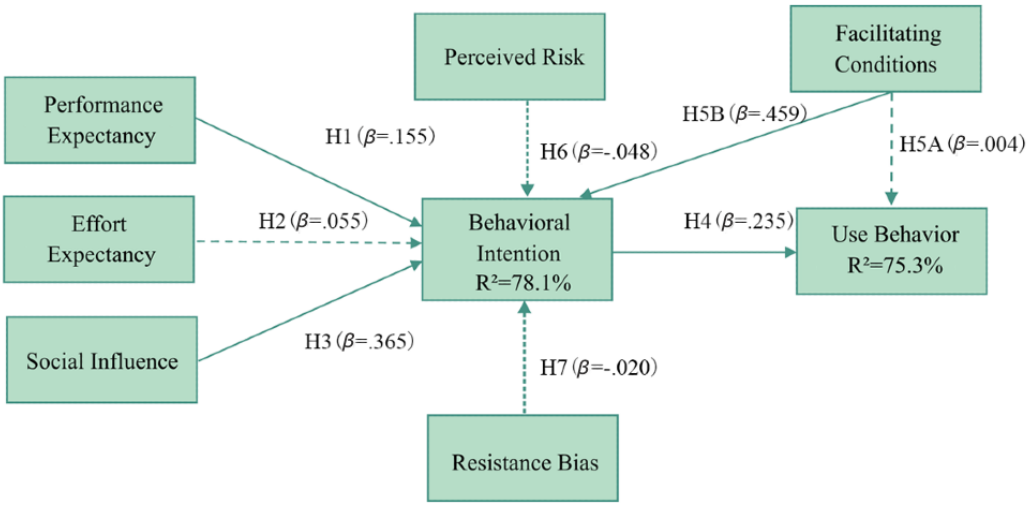


Table 6. Summary of support for the hypotheses.

Variables and hypotheses	Standardized factor loadings (SE)	Critical ratio (T value)	*P* value	Results
Hypothesis 1: Behavioral intention is affected by performance expectancy.	0.155 (0.063)	2.464	.01	Supported
Hypothesis 2: Behavioral intention is affected by effort expectancy.	0.055 (0.073)	0.755	.45	Not supported
Hypothesis 3: Behavioral intention is affected by social influence.	0.365 (0.076)	4.795	<.001	Supported
Hypothesis 4: Use behavior is affected by behavioral intention.	0.235 (0.024)	9.778	<.001	Supported
Hypothesis 5A: Use behavior is affected by facilitating conditions.	0.004 (0.022)	0.195	.85	Not supported
Hypothesis 5B: Behavioral intention is affected by facilitating conditions.	0.459 (0.061)	7.561	<.001	Supported
Hypothesis 6: Behavioral intention is affected by perceived risk.	−0.048 (0.051)	−0.939	.35	Not supported
Hypothesis 7: Behavioral intention is affected by resistance bias.	−0.020 (0.042)	−0.476	.63	Not supported

Discussion

Principal Findings

In this study, we investigated factors that may affect the adoption of an AI-assisted technology for developing treatment plans for cancer radiation therapy patients. The constructs of the research model were developed based on the UTAUT model, and perceived risk and resistance bias were added as potential barriers to new technology adoption. There were 3 principal findings, as follows: (1) the physicians' overall perceptions of the AI-assisted technology for radiation contouring were high; (2) social influence and facilitating conditions played a much more important role in AI adoption among radiation oncologists than performance expectancy, but effort expectancy was not significantly related to behavioral intention (*P*=.45); and (3) perceived risk (*P*=.35) and resistance bias (*P*=.63) were not significantly related to behavioral intention.

The Physicians' Overall Perceptions of the AI-Assisted Technology for Radiation Contouring Were High

Participants had high behavioral intentions (score: mean 5.484 out of 7.0) to use the new AI-assisted contouring technology; 90.9% (279/307) of the total sample gave scores of 4 or above, meaning that they were planning to use the system within 1 year. This result nearly matched those of a survey conducted in 2019, in which 84.62% of the participating medical departments planned to try the AI-assisted contouring system; 10.26% of these departments were in the process of purchasing the system, and 46.15% had purchased plans [46].

Participants' performance expectancy and effort expectancy were high, as they believed that the AI technology could improve their work performance and efficiency. This is similar to the perceptions of Canadian radiation oncologists, who believe that AI could save time and deliver better results to benefit patients [47].

The Relationship Among Performance Expectancy, Social Influence, and Facilitating Conditions Affected Behavioral Intention

Many studies have proven that performance expectancy, effort expectancy, social influence, and facilitating conditions have positive effects on use behavior [8,11,20,48]. For example, in a study of health care robot adoption, these four constructs significantly predicted the usage and uptake of robots [30]. In our study, performance expectancy (*P*=.01), social influence (*P*<.001), and facilitating conditions (*P*<.001) were significantly related to behavioral intention, while effort expectancy did not significantly affect behavioral intention (*P*=.45). Social influence

and facilitating conditions affected behavioral intention much more strongly than performance expectancy. These results indicate that in China, when physicians encounter AI-assisted technology for treatment plans, their opinions are more likely to be driven by people with which they have a close relationship (ie, the items of the social influence construct), such as friends and colleagues, department heads, tutors, superiors, and hospital leaders. This phenomenon reflects the culture of vertical collectivism, which is aligned with the ideology of "utilitarian Guanxi"—a Chinese philosophy that combines the pursuit of profit with objective goals [49]. Further, social propaganda, such as news stories about successfully using AI technology from professionals in the health care industry, likely influences physicians' perceptions toward using such technology.

Facilitating conditions did not have a direct effect on use behavior. However, this construct was the most important factor that affected behavioral intention. As shown by our results, behavioral intention moderates the effect that facilitating conditions have on use behavior (P<.001), indicating that providing support, such as support from the information technology team in a hospital, and having the knowledge and skills needed for using new technology were necessarily important and influenced radiation oncologists' use intentions. On the other hand, most researchers that use the UTAUT model agree that effort expectancy has a significant effect on behavioral intention. However, there have been some exceptions. For example, perceived ease of use did not have an effect on attitudes related to health care professionals' decisions to accept telemedicine technology [50]. In our study, effort expectancy did not affect use intention. In China, physicians in level 3 public hospitals (the highest standard in China) usually have heavy workloads. For example, at Sun Yat-sen University Cancer Center, the Department of Radiation Oncology works on approximately 1200 cases each day. The participants claimed that they did not mind putting in more effort and spending more time on learning about the new technology in order to later reduce their workloads. Determining whether this approach reduces or redistributes workloads is an area worthy of future investigation. Additionally, the AI technology was developed based on participants' previously used software. As a result, they found it easier to gain fluency with the system.

Perceived Risk and Resistance Bias Were Not Significantly Related to Behavioral Intention

In most research on new technology adoption, perceived risk has negatively affected the public's behavioral intentions [11,29]. However, in our study, perceived risk did not affect radiation oncologists' behavioral intentions. Even though they perceived risks (score: mean 4.418 out of 7.0) such as the possibility of being misled with inaccurate contouring and privacy issues, oncologists still had the intention to use the system, as did others (colleagues and superiors). The main reason for this was that it could effectively reduce their heavy workloads. In interviews with some of the participants, physicians explained that they could manually fix the errors in the contouring area because the amount of patient case data has increased and machine learning has made the AI-assisted contouring system more and more accurate; therefore, more time will be saved in the future.

As discussed in a study by Ye et al [10], resistance bias has had a negative impact on the technological acceptance of AI ophthalmic devices. However, different from the general public's perceptions, the low score for resistance bias (score: mean 3.528 out of 7.0) in our study showed that radiation oncologists were not likely to reject AI-assisted contouring technology, regardless of their unfamiliarity with the technology, their past experiences with the failure of new technology products, or the possibility of AI replacing and displacing jobs.

Our study demonstrated that resistance to using an AI-based contouring system to improve patient outcomes and work efficiency was low. There was little to no apprehension among users to learn about and implement this new technology. In previous studies of technology acceptance, resistance bias was observed among medical staff when the benefits of technology were not clear or immediate [36,51,52]. In past research on health care robot adoption, resistance bias did not have a direct negative effect on medical staff's behavioral intentions, as the health care robots were convenient to use and provided significantly better results than those provided by traditional methods [30]. In our study, the AI platform was developed by using the previous software that the oncologists used for work. This resulted in immediate benefits for improving contouring results and saving time. Therefore, negative perceptions surrounding new technologies with clear and immediate benefits, particularly fear and apprehension toward using such technologies (regardless of the degree of these perceptions), may not be prevalent in China. On the other hand, the AI contouring system in our study is used as an assistance tool for providing image contouring results as part of the process for developing a radiation therapy plan. In our case, oncologists still have to work on modifying the contouring areas and deciding on the amounts of radiation for each area. Wong et al [47] conducted a study among Canadian radiation oncologists that showed similar results; 77% of the participants agreed that human oversight was necessary for optimizing the planning process and delivering better quality care.

In China, using AI-assisted technology to relieve physicians' workloads will be a major trend in the future, as such technology tries to fill the gap between the public's demands and the limited amount of medical resources. Chinese hospitals have the need and motivation to adopt AI to improve efficiency, solve the health care professional shortage problem, and reduce patient waiting times. At Sun Yat-sen University Cancer Center, a breakthrough point for the AI product was the potential research opportunities with the AI development company for improving treatment planning accuracy, as the AI product not only helps oncologists improve their work efficiency but also serves as a valuable tool for scientific research.

Hospitals in China, like those in other Asian countries such as India, South Korea, and Thailand, have also proudly indicated their usage and acceptance of AI technology in their marketing campaigns, which have expounded to patients how much their medical services are improved by such technologies [18]. However, anecdotal reports on AI products in cancer care developed by IBM Corporation, one of the largest AI technology providers in the industry, have shown that the company has had trouble with finding customers in the United States [53] and

that technology resistance exists among oncologists who claim that they trust their own judgment over that of software [18]. We suggest that cultural differences may result in different perceptions on AI adoption in Western and Asian countries. Asian physicians may be more likely to accept the idea that AI technology is an assistant that performs certain routine tasks and more likely to reject AI doctors. In this case, the perception is that the AI system will not displace or replace the clinical practitioner; instead, the system will get rid of routine tasks and thus allow for more time to focus on patient care.

Implications for Practice

This study provides 3 suggestions for health care AI developers and hospitals that are planning to introduce AI for clinical decision-making. First, when a hospital introduces a new AI-assisted technology for clinical decision-making in China, the leadership should consider a top-down management process. Hospital leaders and department heads should demonstrate how a new technology works, advocate for the new technology, and encourage subordinate physicians to use the new technology. The hospital could also share AI success stories to influence public opinions on the use of the new technology and highlight the benefits of the system. At Sun Yat-sen University Cancer Center, breakthrough news stories about AI clinical research have been shown not only on social media platforms and local news programs but also at special events such as AI versus human competitions, which have been held at a number of academic conferences, to prove the accuracy of their AI system to both the public and experts. Second, the user interface of the new AI-assisted system should be as consistent as possible with the previously used interface to reduce the time required for users to adapt to the new technology. The hospital information technology center should also work closely with AI technology companies to provide technical support and data security for the use of the new AI platform. Third, per the lesson learned from IBM Corporation's failure in the United States, AI technology development (ie, progressing from perceptual intelligence to cognitive intelligence) should be a step-by-step process. IBM Corporation's oncology expert advisor project used natural language processing to provide treatment recommendations. However, the ambitions of such project were, in hindsight, a step too far. Instead of installing an AI doctor, Chinese hospitals first created an AI technology to act as an assistant tool for medical imaging to free physicians from routine work. When a clear application scenario for medical AI is imagined and created, people are more likely to realize the value of such technology.

Limitations and Future Research

This study has some limitations. Our results might be biased due to the age of participants, as 87.6% (269/307) of the participants were aged under 40 years. The majority of the participants (176/307, 57.3%) were also from Sun Yat-sen University Cancer Center, where medical capacity is high. Additionally, there might be elements of sampling bias, as convenience sampling was used. Further research needs to be carried out among medical professionals in other countries for comparisons of AI technology acceptance in health care services in different cultural contexts.

Conclusion

In this study, technology resistance among Chinese radiation oncologists was low and not related to behavioral intention. Negative perceptions of new technologies with clear and immediate benefits, such as fear and apprehension toward using such technologies (regardless of the degree of these perceptions), may not be prevalent in China. Further, not all of the factors in the Venkatesh UTAUT model applied to AI technology adoption among physicians in a Chinese context. Oncologists' opinions are more likely to be driven by people with which they have a close relationship (the items of the SI construct), such as friends and colleagues, department heads, tutors, superiors, and hospital leaders, and be based on receiving technology support from information technology departments and companies.

Acknowledgments

This study was supported by the National Key Research and Development Program of China (grants 2020YFC1316900 and 2020YFC1316904), the Young Creative Talent Program of Sun Yat-sen University Cancer Center (grant PT21100201), the PhD Start-up Fund of the Natural Science Foundation of the Guangdong Province of China (grant 2018A030310005), the China Postdoctoral Science Foundation (grant 2019M663348), and the National Natural Science Foundation of China (grants 72102238, 71572207, 71832015, and 72072191). The authors would like to thank Ms Wanfei Xie and Ms Le Xia from Sun Yat-sen University Cancer Center for their help with questionnaire collection.

Conflicts of Interest

None declared.

References

1. Ahuja AS. The impact of artificial intelligence in medicine on the future role of the physician. PeerJ 2019;7:e7702 [FREE Full text] [doi: 10.7717/peerj.7702] [Medline: 31592346]
2. 2020 Blue Book of Artificial Intelligence Medical Industry Development. China Academy of Information and Communication. 2020. URL: http://www.caict.ac.cn/kxyj/qwfb/ztbg/202009/t20200908_323708.htm [accessed 2021-07-16]
3. White paper on global development of artificial intelligence. Deloitte China. 2019. URL: https://www2.deloitte.com/content/dam/Deloitte/cn/Documents/technology-media-telecommunications/deloitte-cn-tmt-ai-report-zh-190919.pdf [accessed 2020-11-16]

4. Reddy S, Fox J, Purohit MP. Artificial intelligence-enabled healthcare delivery. J R Soc Med 2019 Jan;112(1):22-28. [doi: 10.1177/0141076818815510] [Medline: 30507284]
5. Musen MA, Shahar Y, Shortliffe EH. Clinical decision-support systems. In: Inhortliffe EH, Cimino JJ, editors. Biomedical Informatics: Computer Applications in Health Care and Biomedicine, 3rd Edition. New York, New York: Springer; 2006:698-736.
6. Teguh DN, Levendag PC, Voet PW, Al-Mamgani A, Han X, Wolf TK, et al. Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck. Int J Radiat Oncol Biol Phys 2011 Nov 15;81(4):950-957. [doi: 10.1016/j.ijrobp.2010.07.009] [Medline: 20932664]
7. Chen AM, Chin R, Beron P, Yoshizaki T, Mikaeilian AG, Cao M. Inadequate target volume delineation and local-regional recurrence after intensity-modulated radiotherapy for human papillomavirus-positive oropharynx cancer. Radiother Oncol 2017 Jun;123(3):412-418. [doi: 10.1016/j.radonc.2017.04.015] [Medline: 28511960]
8. Lin L, Dou Q, Jin YM, Zhou GQ, Tang YQ, Chen WL, et al. Deep learning for automated contouring of primary tumor volumes by MRI for nasopharyngeal carcinoma. Radiology 2019 Jun;291(3):677-686. [doi: 10.1148/radiol.2019182012] [Medline: 30912722]
9. Defnet AM, Davis SS. The robot in general surgery-change is the only constant. JAMA Surg 2021 Jan 01;156(1):30. [doi: 10.1001/jamasurg.2020.4590] [Medline: 33084850]
10. Ye T, Xue J, He M, Gu J, Lin H, Xu B, et al. Psychosocial factors affecting artificial intelligence adoption in health care in China: Cross-sectional study. J Med Internet Res 2019 Oct 17;21(10):e14316 [FREE Full text] [doi: 10.2196/14316] [Medline: 31625950]
11. Ataka K, Chen D, Yamaguchi T, Yamamoto I. Living in harmony with artificial intelligence: Will AI replace your work? Journal of Information Processing and Management 2018;60(12):865-881. [doi: 10.1241/johokanri.60.865]
12. Goldhahn J, Rampton V, Spinas GA. Could artificial intelligence make doctors obsolete? BMJ 2018 Nov 07;363:k4563. [doi: 10.1136/bmj.k4563] [Medline: 30404897]
13. Lapointe L, Rivard S. Getting physicians to accept new information technology: insights from case studies. CMAJ 2006 May 23;174(11):1573-1578 [FREE Full text] [doi: 10.1503/cmaj.050281] [Medline: 16717265]
14. Bhattacherjee A, Hikmet N, Menachemi N, Kayhan VO, Brooks RG. The differential performance effects of healthcare information technology adoption. Information Systems Management 2006 Dec 22;24(1):5-14. [doi: 10.1080/10580530601036778]
15. Bhattacherjee A, Hikmet N. Physicians' resistance toward healthcare information technology: a theoretical model and empirical test. Eur J Inf Syst 2017 Dec 19;16(6):725-737. [doi: 10.1057/palgrave.ejis.3000717]
16. Hsieh PJ. Physicians' acceptance of electronic medical records exchange: an extension of the decomposed TPB model with institutional trust and perceived risk. Int J Med Inform 2015 Jan;84(1):1-14. [doi: 10.1016/j.ijmedinf.2014.08.008] [Medline: 25242228]
17. Noorbakhsh-Sabet N, Zand R, Zhang Y, Abedi V. Artificial intelligence transforms the future of health care. Am J Med 2019 Jul;132(7):795-801 [FREE Full text] [doi: 10.1016/j.amjmed.2019.01.017] [Medline: 30710543]
18. Strickland E. How IBM Watson overpromised and underdelivered on AI health care. IEEE Spectrum. 2019 Apr 02. URL: https://spectrum.ieee.org/biomedical/diagnostics/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care [accessed 2020-12-16]
19. Cooper L, Lombardo C. IBM explores sale of IBM Watson Health. Morningstar. 2021 Feb 18. URL: https://www.morningstar.com/news/dow-jones/20210218166 50/ibm-explores-sale-of-ibm-watson-health-update [accessed 2021-07-01]
20. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
21. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: Four longitudinal field studies. Manage Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
22. Venkatesh V, Bala H. Technology acceptance model 3 and a research agenda on interventions. Decis Sci 2008 May 09;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
23. Venkatesh V, Morris M, Davis G, Davis F. User acceptance of information technology: Toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
24. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157-178. [doi: 10.2307/41410412]
25. Kijsanayotin B, Pannarunothai S, Speedie SM. Factors influencing health information technology adoption in Thailand's community health centers: applying the UTAUT model. Int J Med Inform 2009 Jun;78(6):404-416. [doi: 10.1016/j.ijmedinf.2008.12.005] [Medline: 19196548]
26. Shiferaw KB, Mehari EA. Modeling predictors of acceptance and use of electronic medical record system in a resource limited setting: Using modified UTAUT model. Inform Med Unlocked 2019;17:100182 [FREE Full text] [doi: 10.1016/j.imu.2019.100182]
27. Bawack RE, Kala Kamdjoug JR. Adequacy of UTAUT in clinician adoption of health information systems in developing countries: The case of Cameroon. Int J Med Inform 2018 Jan;109:15-22. [doi: 10.1016/j.ijmedinf.2017.10.016] [Medline: 29195701]

28. Jang SH, Kim RH, Lee CW. Effect of u-healthcare service quality on usage intention in a healthcare service. Technol Forecast Soc Change 2016 Dec;113:396-403. [doi: 10.1016/j.techfore.2016.07.030]
29. Martins C, Oliveira T, Popovič A. Understanding the internet banking adoption: A unified theory of acceptance and use of technology and perceived risk application. Int J Inf Manage 2014 Feb;34(1):1-13. [doi: 10.1016/j.ijinfomgt.2013.06.002]
30. Vichitkraivin P, Naenna T. Factors of healthcare robot adoption by medical staff in Thai government hospitals. Health Technol (Berl) 2020 Nov 06;11(1):139-151. [doi: 10.1007/s12553-020-00489-4]
31. Berg M, Aarts J, van der Lei J. ICT in health care: sociotechnical approaches. Methods Inf Med 2003;42(4):297-301. [Medline: 14534625]
32. Bauer RA. Consumer behavior as risk taking. In: Dynamic Marketing for a Changing World. 1960 Presented at: The 43rd National Conference of the American Marketing Association; June 15-17, 1960; Chicago, Illinois p. 389-398.
33. Featherman MS, Pavlou PA. Predicting e-services adoption: a perceived risk facets perspective. Int J Hum Comput Stud 2003 Oct;59(4):451-474. [doi: 10.1016/S1071-5819(03)00111-3]
34. Dou K, Yu P, Deng N, Liu F, Guan Y, Li Z, et al. Patients' acceptance of smartphone health technology for chronic disease management: A theoretical model and empirical test. JMIR Mhealth Uhealth 2017 Dec 06;5(12):e177 [FREE Full text] [doi: 10.2196/mhealth.7886] [Medline: 29212629]
35. Guo X, Sun Y, Wang N, Peng Z, Yan Z. The dark side of elderly acceptance of preventive mobile health services in China. Electronic Markets 2012 Dec 11;23(1):49-61. [doi: 10.1007/s12525-012-0112-4]
36. Deng Z, Mo X, Liu S. Comparison of the middle-aged and older users' adoption of mobile health services in China. Int J Med Inform 2014 Mar;83(3):210-224. [doi: 10.1016/j.ijmedinf.2013.12.002] [Medline: 24388129]
37. Raza MH, Adenola AF, Nafarieh A, Robertson W. The slow adoption of cloud computing and IT workforce. Procedia Comput Sci 2015;52:1114-1119. [doi: 10.1016/j.procs.2015.05.128]
38. Andrews L, Gajanayake R, Sahama T. The Australian general public's perceptions of having a personally controlled electronic health record (PCEHR). Int J Med Inform 2014 Dec;83(12):889-900. [doi: 10.1016/j.ijmedinf.2014.08.002] [Medline: 25200198]
39. Anderson JC, Gerbing DW. Structural equation modeling in practice: A review and recommended two-step approach. Psychol Bull 1988 May;103(3):411-423. [doi: 10.1037/0033-2909.103.3.411]
40. Malhotra NK, Kim SS, Patil A. Common method variance in IS research: A comparison of alternative approaches and a reanalysis of past research. Manage Sci 2006 Dec;52(12):1865-1883. [doi: 10.1287/mnsc.1060.0597]
41. Rivard S, Huff SL. Factors of success for end-user computing. Commun ACM 1988 May;31(5):552-561. [doi: 10.1145/42411.42418]
42. Hair JF, Ringle CM, Sarstedt M. PLS-SEM: Indeed a silver bullet. Journal of Marketing Theory and Practice 2014 Dec 08;19(2):139-152. [doi: 10.2753/MTP1069-6679190202]
43. Fornell C, Larcker DF. Structural equation models with unobservable variables and measurement error: Algebra and statistics. J Mark Res 1981 Aug;18(3):382-388. [doi: 10.2307/3150980]
44. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Modeling 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
45. Wasserstein RL, Lazar NA. The ASA statement on p-values: Context, process, and purpose. Am Stat 2016 Jun 09;70(2):129-133. [doi: 10.1080/00031305.2016.1154108]
46. 2019 China Radiotherapy AI Contouring Product Usage Survey Report. Radiotherapy Front. 2019. URL: https://mp.weixin.qq.com/s/6N9tMen7768OxpmMmzJRvg [accessed 2020-12-23]
47. Wong K, Gallant F, Szumacher E. Perceptions of Canadian radiation oncologists, radiation physicists, radiation therapists and radiation trainees about the impact of artificial intelligence in radiation oncology - national survey. J Med Imaging Radiat Sci 2021 Mar;52(1):44-48. [doi: 10.1016/j.jmir.2020.11.013] [Medline: 33323332]
48. Chang I, Hwang H, Hung W, Li Y. Physicians' acceptance of pharmacokinetics-based clinical decision support systems. Expert Syst Appl 2007 Aug;33(2):296-303. [doi: 10.1016/j.eswa.2006.05.001]
49. Du J, Vantilborgh T. Cultural differences in the content of employees' psychological contract: A qualitative study comparing Belgium and China. Psychol Belg 2020 Jun 01;60(1):132-151 [FREE Full text] [doi: 10.5334/pb.498] [Medline: 32523710]
50. Chau PY, Hu PJ. Investigating healthcare professionals' decisions to accept telemedicine technology: an empirical test of competing theories. Information & Management 2002 Jan;39(4):297-311. [doi: 10.1016/S0378-7206(01)00098-2]
51. Panteleimonitis S, Parvaiz A. Robotic colorectal surgery. Surgery (Oxf) 2017 Mar;35(3):151-156. [doi: 10.1016/j.mpsur.2016.12.004]
52. Brewster L, Mountain G, Wessels B, Kelly C, Hawley M. Factors affecting front line staff acceptance of telehealth technologies: a mixed-method systematic review. J Adv Nurs 2014 Jan;70(1):21-33. [doi: 10.1111/jan.12196] [Medline: 23786584]
53. Ross C, Swetlitz I. IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close. STAT. 2017 Sep 05. URL: https://www.statnews.com/2017/09/05/watson-ibm-cancer/ [accessed 2020-12-23]

Abbreviations

AI: artificial intelligence
AVE: average variance extracted
TAM: Technology Acceptance Model
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by R Kukafka; submitted 13.01.21; peer-reviewed by I Gabashvili, F Liu, J Ropero; comments to author 17.05.21; revised version received 07.07.21; accepted 27.07.21; published 30.09.21.

Please cite as:
Zhai H, Yang X, Xue J, Lavender C, Ye T, Li JB, Xu L, Lin L, Cao W, Sun Y
Radiation Oncologists' Perceptions of Adopting an Artificial Intelligence–Assisted Contouring Technology: Model Development and Questionnaire Study
J Med Internet Res 2021;23(9):e27122
URL: https://www.jmir.org/2021/9/e27122
doi:10.2196/27122
PMID:34591029

Viewpoint

Chloe for COVID-19: Evolution of an Intelligent Conversational Agent to Address Infodemic Management Needs During the COVID-19 Pandemic

Sophia Siedlikowski[1], BSc (Hons), MSci; Louis-Philippe Noël[1], BEng, MSci; Stephanie Anne Moynihan[1], BSc (Hons), MD; Marc Robin[1], BSc, MD

Dialogue Health Technologies Inc, Montreal, QC, Canada

Corresponding Author:
Marc Robin, BSc, MD
Dialogue Health Technologies Inc
390 Rue Notre-Dame Ouest #200
Montreal, QC, H2Y 1T9
Canada
Phone: 1 613 806 0671
Email: marc.robin@dialogue.co

Abstract

There is an unprecedented demand for infodemic management due to rapidly evolving information about the novel COVID-19 pandemic. This viewpoint paper details the evolution of a Canadian digital information tool, *Chloe* for COVID-19, based on incremental leveraging of artificial intelligence techniques. By providing an accessible summary of *Chloe's* development, we show how proactive cooperation between health, technology, and corporate sectors can lead to a rapidly scalable, safe, and secure virtual chatbot to assist public health efforts in keeping Canadians informed. We then highlight *Chloe's* strengths, the challenges we faced during the development process, and future directions for the role of chatbots in infodemic management. The information presented here may guide future collaborative efforts in health technology in order to enhance access to accurate and timely health information to the public.

(J Med Internet Res 2021;23(9):e27283) doi:10.2196/27283

KEYWORDS

chatbot; COVID-19; conversational agents; public health; artificial intelligence; infodemic; infodemiology; misinformation; digital health; virtual care

Introduction

In the early stages of a global health crisis caused by a novel disease such as COVID-19, misinformation tends to propagate rapidly due to the absence of easily accessible and reliable information from authoritative sources [1]. According to the World Health Organization, controlling the circulation of information about a novel disease requires infodemic management, defined as the application of "evidence-based interventions that bring understandable, localized, evidence-based information to citizens and drive positive health-seeking behaviour" [2]. At the onset of the COVID-19 pandemic in Canada, however, public health hotlines were overwhelmed with calls, impeding the infodemic management process. For example, on March 11, 2020, Quebec's provincial health information 811 phone line received 16,840 calls compared to an average of 6000 calls managed on a typical day [3].

In recent years, chatbots have emerged as a promising means to support public health efforts by easily increasing access to accurate, tailored, and free information to millions of people at once [4]. Leveraged with artificial intelligence (AI) tools, chatbots are intelligent conversational agents that result from the application of natural language processing techniques to analyze speech combined with an engine that is trained to provide human-like responses [5].

The recent surge in chatbot use across industries and business functions such as customer service, sales, marketing, and automation of internal processes stems from an evolution in the way humans communicate [6]. The current digital age, in which most individuals have access to a mobile phone and are accustomed to holding multiple conversations at once through short online interactions, represents an ideal environment for chatbot development [6]. Since the onset of the COVID-19 pandemic, chatbots are increasingly being developed to

efficiently deliver evidence-based information about COVID-19 to the public [4]. Despite these advances, the possibility of chatbots causing harm by communicating inaccurate information to users warrants ongoing monitoring of chatbot quality and safety [7].

The development of Dialogue Health Technologies Inc's chatbot *Chloe* represents a successful Canadian cooperation among health, technology, and corporate sectors, in order to assist the public during the pandemic by addressing their questions and concerns about COVID-19 [8]. Founded in 2016, Dialogue is a leading provider of virtual health care in Canada, offering its services exclusively to Canadian organizations through an online care platform using technologies developed in-house. This care platform consists of an interface in both desktop and mobile application formats that connects members to a care team consisting of health providers and coordinators. Our experience with chatbots in health care includes the design of various AI-enhanced bots used to collect and intuitively interpret patient information in order to determine urgency, appropriateness for virtual care, and the type of provider required. These chatbots were built by combining medical knowledge with state-of the art algorithms and are continuously monitored, updated, and trained to enhance their effectiveness and efficiency, including bot-to-human handoff when required [9].

Chloe was initially conceived to integrate within the Dialogue care platform as a way to help manage questions about COVID-19 from our clients but was quickly converted into a stand-alone digital application in order to be deployed as a free tool for the general public. After Health Canada confirmed the first Canadian case of COVID-19 on January 27, 2020, our medical operations and customer success teams noted an increased flow of questions from members and organizations across the country about this novel disease, looking to us as a health care company for guidance. The urgent need to slow the spread of the virus and provide large-scale access to free information about COVID-19 from authoritative sources, as well as the desire to assist flooded provincial health information phone lines in answering common questions from Canadians, prompted us to pursue an AI approach.

We launched the chatbot *Chloe*, named after Dialogue's first nurse, on March 9, 2020. *Chloe* was designed to provide free, up-to-date information to all Canadians on COVID-19 from trusted Canadian authorities, following federal, provincial, and territorial regulations based on each user's location. To ensure transparency of our technological approach, we rendered the online service and content management system open-source. *Chloe* was designed following best dialogue management principles [10] and built with Rasa, a state-of-the-art open-source framework [11].

This report first outlines the development of *Chloe*, an information and self-assessment tool to respond to a rapidly evolving health crisis. This was achieved by incrementally leveraging AI tools in order to address the challenges associated with the volume of demands and constantly changing information during the early phases of the COVID-19 pandemic. This paper then highlights *Chloe*'s strengths, the challenges we faced, and future directions for chatbot development to support infodemic management.

We wish to note that *Chloe* was designed and launched in an emergency effort to support public health authorities in keeping Canadians informed during the first wave of the COVID-19 pandemic before any governmental information tools had been deployed. Following the rollout of official and local public health online self-assessment tools providing features similar to ours, *Chloe* stopped being regularly updated.

Chloe's Development: From a Frequently Asked Questions Page to an Intelligent Chatbot

Increased uncertainty surrounding the previously unknown virus and disease resulted in a deluge of questions not only for our own care team but also for the entire Canadian health care system. In response to this surge, our Clinical Quality Committee led efforts to provide official answers to the most frequently asked questions (FAQs) from members on our care platform. This led to the release of the first version of our FAQ section on our website on March 5, 2020.

An FAQ section addresses predictable queries, but the constant flow of new information associated with the COVID-19 pandemic required a dynamic tool. In response to this challenge, we designed a self-assessment flow based on official available medical information surveyed by our care team. This allowed us to efficiently sort through and respond to the diverse concerns our members and clients had about COVID-19. We then created a smart virtual medical assistant that we named *Chloe for COVID-19* that was released publicly for all Canadians on March 9, 2020. In accordance with Dialogue's rigorous data security and privacy policies, every effort was made to ensure that Canadians could use *Chloe* in an anonymous and secure fashion. To that effect, no information that could help identify the user (including but not limited to IP addresses and approximate location) was collected by default.

Chloe would ask a user about their symptoms, location, travel history, and recent contacts. This assessment resulted in a personalized recommendation including links to local resources such as the user's provincial government's COVID-19 guidelines. Figure 1 presents one example of the start of the assessment flow for a user in Quebec.

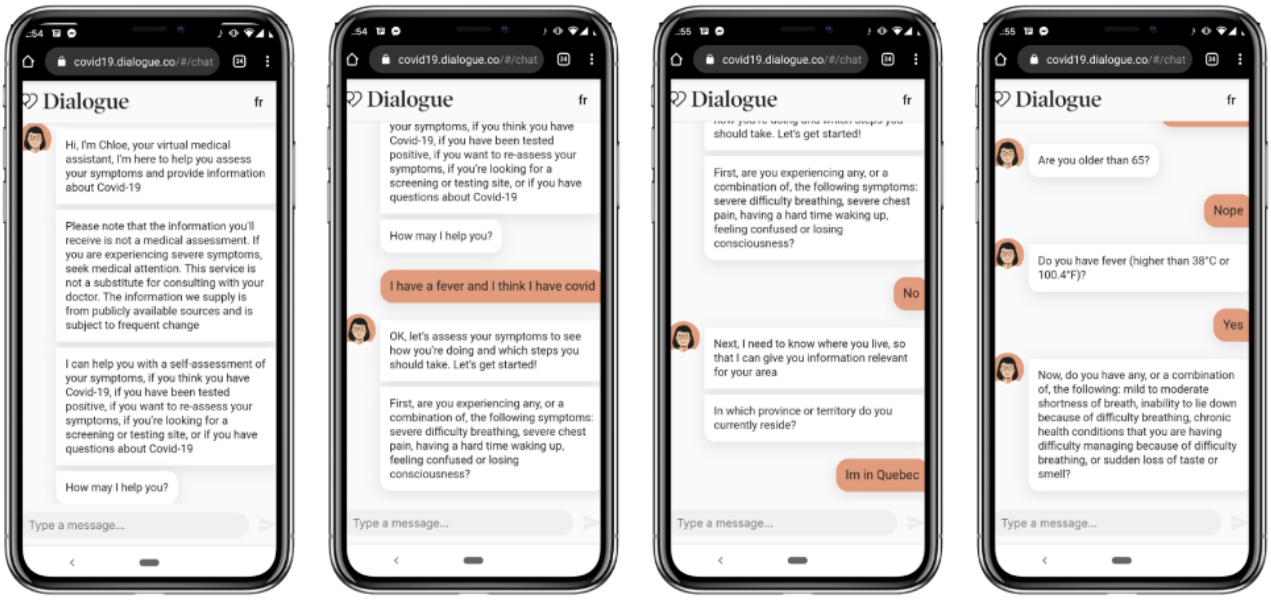
When provincial online self-assessment tools were later deployed, we integrated this information into *Chloe*. Our self-assessment tool saw the highest traffic on March 15, 2020, with 5945 new users accessing the tool, coinciding with the overflow of public health information phone lines. By April 18, 2020, it had been used 73,000 times, further indicating *Chloe's* contribution to overall public health efforts in keeping Canadians informed.

In order to keep up with constantly evolving information spanning multiple Canadian jurisdictions and to discriminate between official information and disinformation, we designed and built a web scraper. This automated service was programmed to extract the latest data from reliable sources

identified by our medical team. During the peak of the first wave, it ran every hour. When the flow of new information later slowed, it was then activated every 4 hours to ensure *Chloe* as well as our online FAQ still offered the most up-to-date information to Canadians.

In addition to the self-assessment tool and FAQ, we developed a question and answer (Q&A) model within *Chloe* to further meet the information needs of Canadians regarding COVID-19.

Figure 1. Example of self-assessment flow for a user in Quebec.



Given the time-consuming nature of this process and the need to gain a better understanding of what questions were most commonly asked, we used clustering to divide the approximately 4000 questions into 40 groups of similar questions. Clustering is a form of unsupervised machine learning, which aims to segregate groups with similar traits and assign them into clusters. The clustering algorithm used word embeddings to represent each of the questions in a multidimensional way. We were then able to accurately measure the similarity among the questions and split them into groups accordingly. For instance, all questions asking about how long the coronavirus could survive on surfaces, one of our most commonly asked questions, were grouped into one cluster.

On April 28, 2020, we released our first interactive Q&A model, marking our transition from an FAQ system to a customized conversational experience. As a first iteration of a natural language Q&A model, we implemented a distance-computing, similarity-matching algorithm. In this model, all clustered questions were embedded in a multidimensional space. Figure 2 represents this step of the process, in which each colored circle is a question and its position is determined by the attributes computed by the algorithm. Averaging techniques were used to obtain the center of each cluster. This identified the average features of the questions within the same cluster.

For example, it was determined that the average question regarding the coronavirus on surfaces holds a combination of the word "surface" or "object" and a declension of the word "time." Next, we embedded the user's question using the same pretrained model, to have it in the same multidimensional space as the cluster averages. The distance between a user's question and the clusters could be computed mathematically. The closest cluster could then be identified and the question categorized.

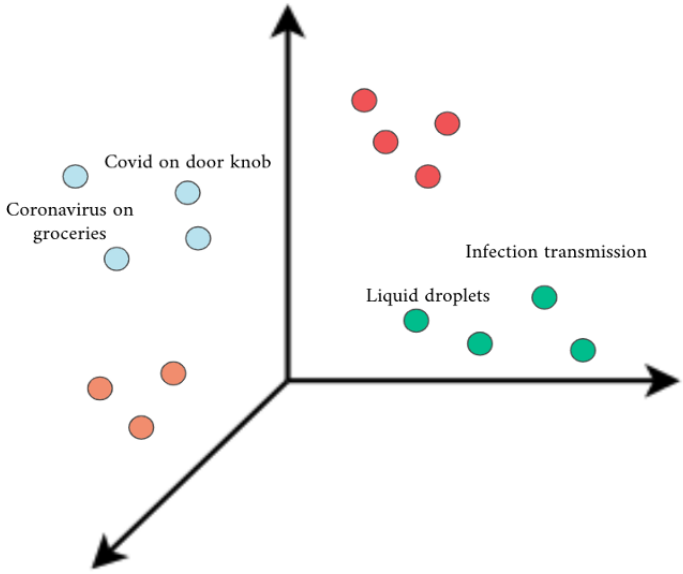
For users' questions that matched a related cluster, we could then return the associated Q&A. For instance, the question "How long does it take for coronavirus to vanish from doorknobs?" was matched to the cluster that dealt with the virus on surfaces. We could then return the associated question-answer pair that would look like this:

> *How long can coronavirus stay on surfaces? It has been demonstrated that it can last for X hours on Y surface.*

If the distance between the user's question and each of the cluster's centers was too great, however, the question would not be categorized, and an answer was not returned. In such cases, the user would be offered a link to our FAQ page, and *Chloe* would ask the user if they had another question. Our monitoring team also took note of these situations to further train *Chloe* to be able to answer an increasing number of user questions. This was achieved by having our team analyze these additional questions and add them to the set of questions used to retrain the clustering algorithm.

The up-to-date COVID-19 information identified by our web scraper was used to refine the Q&A model, which was then further optimized using machine learning techniques. Our team of applied researchers developed a machine learning model that would find the correct answer to a given question about COVID-19 by feeding it approximately 4000 questions acquired through crowdsourcing and several thousand more from data collection tools offered by Amazon Web Services [12].

Figure 2. Representation of questions in a multidimensional space.



Canada being a bilingual country, we also needed this model to accurately return question-answer pairs to all Canadians in both English and French. We therefore followed the aforementioned process twice in order to cover questions in both languages.

While developing the previously described similarity model, we also worked on a more advanced model consisting of 2 main components: a Bidirectional Encoder Representations from Transformers (BERT) model [13] that ranked all COVID-19 FAQ questions from trusted sources according to their relevance with a user's question and an out-of-distribution (OOD) detector using a Local Outlier Factor algorithm [14]. The purpose of this OOD detector was to eliminate questions that could not be answered with the available information from the vetted sources the chatbot used. This improved the number of questions answered and the accuracy of the answers provided while minimizing manual updating work from our care team. Figure 3 presents a general diagram of the model architecture.

Figure 3. Diagram of the final question & answer (Q&A) model architecture using the example of a question about the symptoms of COVID-19. BERT: Bidirectional Encoder Representations from Transformers; FAQ: frequently asked question; OOD: out-of-distribution.

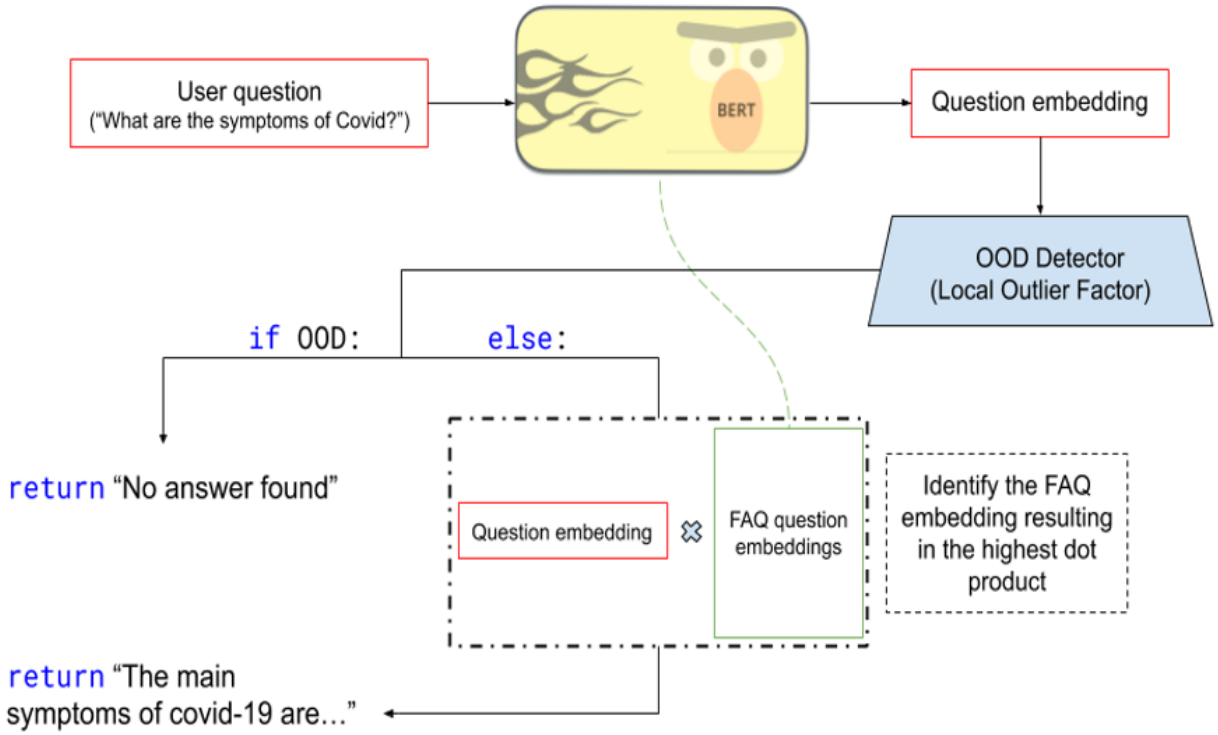


Table 1 describes the results obtained from this final Q&A model's accuracy. Overall accuracy represents the model's ability to give a correct answer to the user's question, and OOD accuracy represents its ability to detect situations in which the question asked has no answer in our data sources. The BERT pretrained model we used has a better performance in English, which explains the discrepancy between the English and French results. Although higher metrics can be obtained from Q&A models, we believe that these results are reasonable given time and data constraints. These results were also deemed acceptable because our Q&A model sought to provide answers to commonly asked questions and not directly inform patient care.

Table 1. Overall and out-of-distribution (OOD) accuracies of the final question & answer (Q&A) model in English and in French.

Language	Overall accuracy, %	OOD accuracy, %
English	81.83	90.48
French	71.32	70.60

We nevertheless addressed the error rates of our model to the best of our abilities by consistently displaying the question and answer as a pair when responding to the user. For example, if the user asked about coronavirus and pets and *Chloe* replied with information about the coronavirus and surfaces, the user would know that the model was responding about surfaces and not pets. This ensured the user would not be misled by false information. Lastly, the model was validated using a standard process [15]. The dataset was divided into training, validation, and test sets, followed by an evaluation of how the test set (not previously seen by the model) performed. Further details about the Q&A model are detailed in a poster presented at the Montreal AI Symposium 2020 [16].

Two other important features were implemented to increase *Chloe's* ability to meet the evolving needs of Canadians: a daily self-health assessment and COVID-19 testing navigation. The self-health check feature was added to the assessment flow at the end of May 2020. Users choosing to enroll in this feature received a daily SMS message with a link to their self-health check-in and had the choice to opt out at any time. Anyone using this self-health assessment service had their first name and phone numbers securely stored on our data servers located in Canada. Having regular contact with users helped tailor the assessment to the user's particular situation and needs. When starting the assessment flow, users' previously reported symptoms and concerns helped the chatbot identify relevant questions to be asked and offer appropriate recommendations. For example, if the user reported a fever the day before, the chatbot might ask if it was still present and if it had gotten better or worse.

The testing navigation feature was added on June 2, 2020 to the assessment flow to support public health officials and public health information phone lines. This feature was built in partnership with Clinia [17], a technology company enabling health care resource navigation. By integrating Clinia's application programming interface into *Chloe*, any user whose self-assessment resulted in a recommendation to get COVID-19 testing would be prompted with the option to search for testing sites. Once the user entered their desired location to get tested, they would receive a list of sites offering COVID-19 testing, their opening hours, special instructions, as well as a link to obtain directions to the sites. Figure 4 presents an example of local testing sites offered to a user in Quebec seeking navigation assistance.

To make the conversation feel as natural as possible for users, we leveraged a machine-learning technology called natural language understanding (NLU), a subtopic of natural language processing, which converts human-readable text inputs into machine language that can be processed to extract an understanding. In our particular case, we extracted intents, which represent the user's query, and entities, which represent key categorized values. For example, if a user said, "I want to find a test site close to Montreal," NLU would help us map the user's intent to something like "find_test_site" and extract the entity "Montreal," which we know is a city. Test sites in Montreal could then easily be suggested to the user. The use of NLU allowed for more open-ended questions akin to normal conversations and minimized clicking. As of June 29, 2020, *Chloe* could ask users: "How may I help you?" NLU components were associated with each possible question, providing a natural conversational experience with an intelligent chatbot. In terms of dialogue management strategies, we used a combination of NLU and static flows defined by a multidisciplinary team of medical experts, software and machine learning engineers, and conversational designers. Figure 5 presents what the latest version of *Chloe* looks like to users upon first connecting with the chatbot. In response to *Chloe*'s initial "How may I help you?", if a user's response falls outside the clusters of questions known to *Chloe*, the chatbot presents 5 clear clickable options from which the user can choose. Figure 6 shows *Chloe's* main features as presented to users.

Figure 4. Example of test navigation results for a user in Quebec.

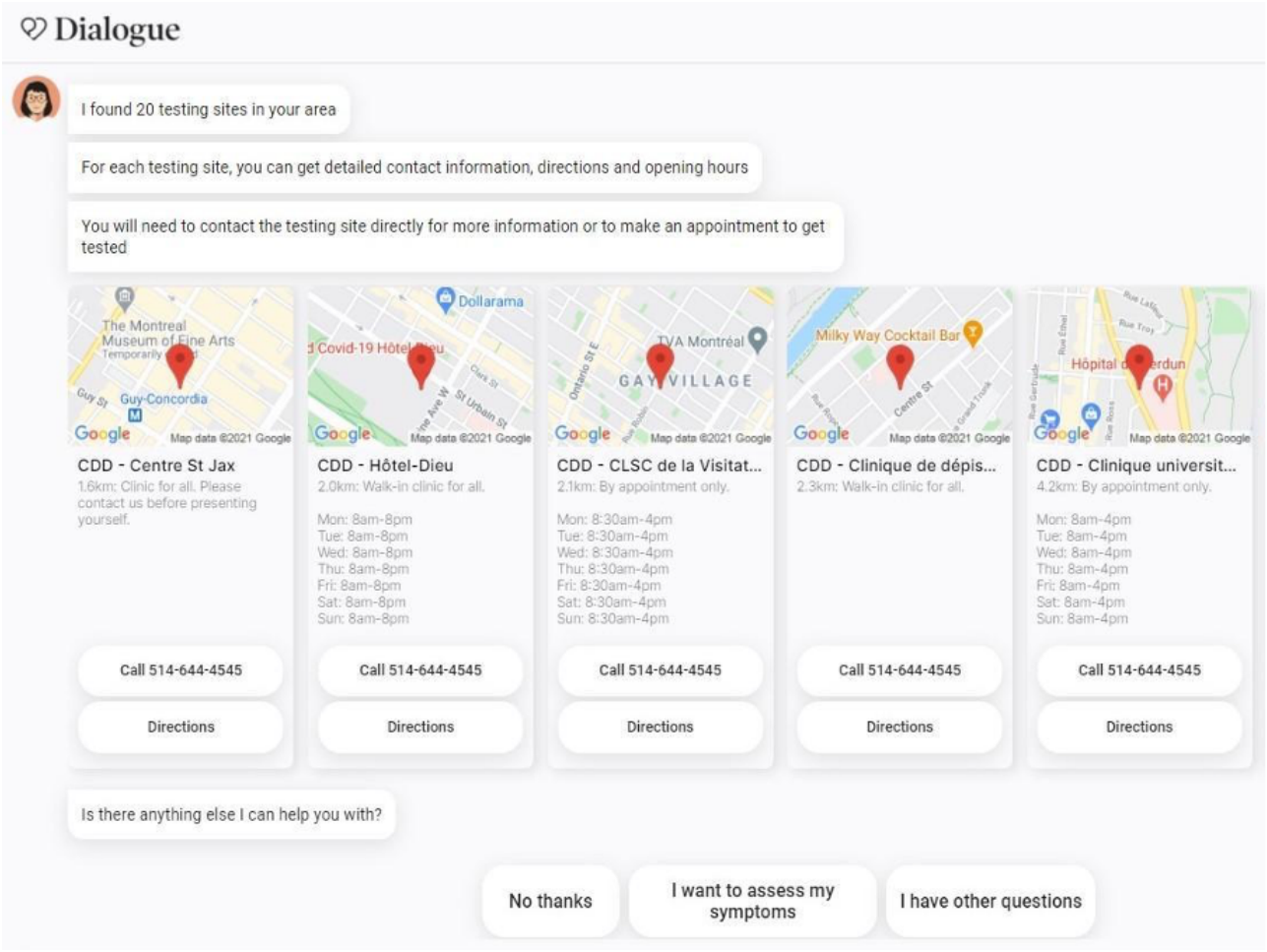


Figure 5. *Chloe's* welcome message to users explaining their options.

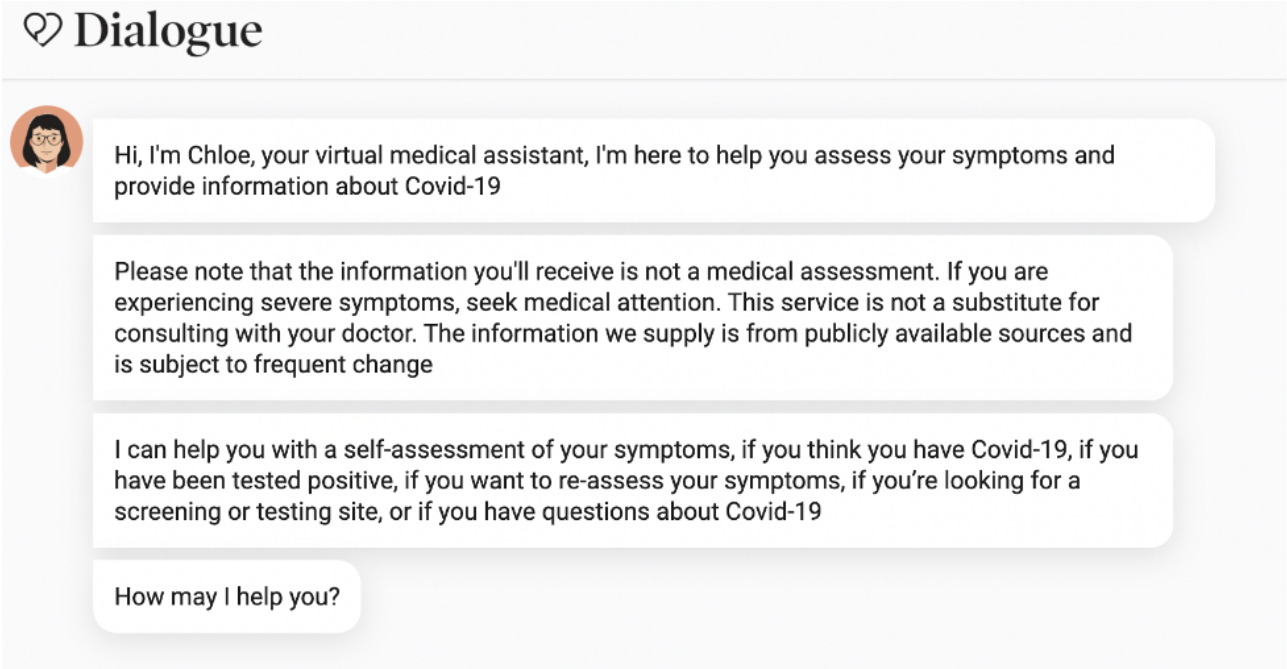
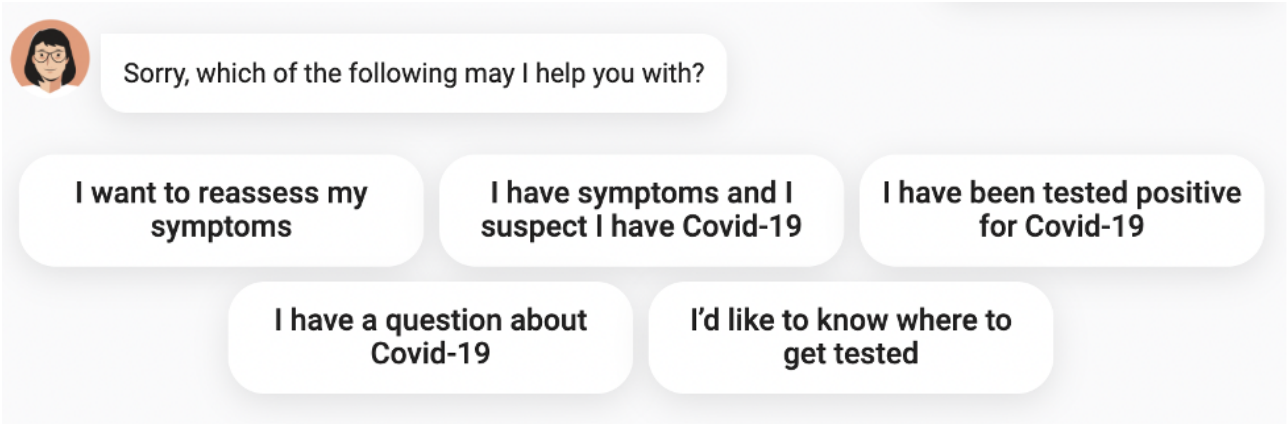


Figure 6. Display of *Chloe's* features available to users.



Strengths, Challenges, and Future Directions

To our knowledge, at the onset of the COVID-19 pandemic, there were no other accessible chatbots that provided free, local, up-to-date information on COVID-19 in both English and French to the Canadian public. As of early March 2020, *Chloe for COVID-19* filled this important gap by supporting efforts in keeping Canadians informed when public health hotlines were most overwhelmed. *Chloe* underwent rigorous training by a dedicated team of medical and technology experts who were involved from inception to deployment. This team continually monitored the chatbot using analytics data, making adjustments and updates when needed. Informal user testing was conducted each time a major update was introduced. This involved approximately 10 individuals recruited internally who tested out the different available flows and provided feedback.

As the pandemic unfolded and more information emerged about COVID-19, *Chloe* adapted to address evolving questions from the public. In her most advanced version in June 2020, *Chloe* uniquely integrated 5 features into 1 bilingual chatbot: a self-assessment tool, resources for users who tested positive for COVID-19, answers to common questions about COVID-19, testing navigation, and a daily self-health check-in.

Despite *Chloe's* success in providing tailored information to the Canadian public early on in the pandemic, we encountered several challenges during *Chloe's* development and after *Chloe's* launch, of which the most notable were keeping up with the increasingly vast and rapidly changing information available on COVID-19 as well as time and data constraints within an emergency health crisis context.

Rapidly changing information on COVID-19 (eg, how the disease presents in individuals, how the SARS-CoV-2 virus is transmitted, and when individuals should get tested), complicated by Canada's multijurisdictional landscape with guidelines changing from one region to another, required constant monitoring from our team to ensure accurate information was provided to users across the country. Furthermore, as mentioned earlier in this report, chatbots have the potential to address the questions of millions of people at once [4], but the risk of amplifying misinformation and the lack of research on chatbot effectiveness present significant challenges, particularly in a pandemic context [7]. By reinforcing local provincial and territorial guidelines, we ensured that *Chloe's* messaging to users aligned with the most up-to-date public health information tailored to each user's location and health situation. We learned that having a multidisciplinary team of technology, nursing, physician, customer success, and clinical quality experts, assisted by AI tools, was the best approach to quickly mobilize all necessary information and make any required adjustments to *Chloe* as the pandemic situation evolved.

Given that we developed *Chloe* as an informational tool in an emergency effort to support overwhelmed public health hotlines, we lacked the time and resources needed to rigorously evaluate *Chloe*'s implementation and measure *Chloe's* impact on users. In a noncrisis context, it would have been ideal to set and track specific implementation metrics other than general chatbot traffic to improve *Chloe* more systematically. In this instance, we prioritized the need to act quickly in assisting public health efforts at the height of a health crisis. We also considered that implementation metrics are particularly important for chatbots that directly inform patient care, such as those that refer users to consultations with health professionals. Those chatbots require closer monitoring than chatbots such as *Chloe*, whose main goal was to reliably answer the public's questions about COVID-19 and to redirect individuals to testing services and local public health guidelines when needed.

Moreover, we were unable to precisely measure *Chloe's* acceptability, but we acknowledge the importance of evaluating this metric [18]. Current metrics used to determine the acceptability of AI chatbots in health care include trustworthiness of the information provided as well as a user experience that is perceived as empathetic to the patient [19]. Despite not formally evaluating *Chloe's* acceptability, *Chloe* was designed to be trustworthy, by offering up-to-date official information from public health sources, and empathetic, by leveraging AI techniques to make the conversation feel as natural as possible with users. We also knew from prior experience designing and implementing health technology tools for the Dialogue platform, as well as extensive member feedback, that chatbots are an acceptable and easy-to-use tool for collecting, assessing, and sharing health information. That being said, rigorous user testing combined with the systematic collection of user feedback, for instance through the integration

of a permanent feedback link in the user interface of a chatbot as done by Herriman et al [4], may have allowed us to better understand user needs and make quicker improvements to *Chloe's* features.

In further building on *Chloe's* development and envisioning the future of chatbots, we would also prioritize evaluating the impact of chatbots on how well individuals feel informed. In a scoping review of the technical metrics used to evaluate healthcare chatbots, Abd-Alrazaq et al [20] found that most chatbots were assessed according to 27 different metrics, with global usability and survey designs as the most commonly used metrics. Although chatbots like *Chloe* are not "healthcare chatbots" in that they do not deliver healthcare services, the evaluation of their usability and performance is warranted in order to improve their ability to provide the right information to users at the right time.

In summary, we believe *Chloe's* development is useful in illustrating the potential for rapid collaboration among health, technology, and corporate sectors to assist public health efforts in keeping individuals informed at the height of a health crisis. Addressing the public's need for accurate health information will remain important within the current pandemic context and during any future health crises. As COVID-19 vaccination campaigns continue to target the general public, intelligent conversational agents have the potential to play a critical role in debunking misinformation surrounding vaccines and directing individuals to public health resources [21,22]. One study in the preprint stage, which examined a chatbot designed to address arguments against the COVID-19 vaccine, promisingly showed that individuals' attitudes towards vaccination became more positive after receiving information from this chatbot [23]. Moreover, in a recent Nature survey of infectious disease researchers, virologists, and immunologists in 23 countries, the majority of respondents predicted that the SARS-CoV-2 virus will likely become endemic over time, with its future closely depending on virus mutations and the type of immunity individuals will acquire [24]. In anticipation of the upcoming waves of uncertainty regarding this virus in the coming years, we envision an ongoing and growing role for chatbots in supporting public health efforts in infodemic management in order to keep the public safe and informed.

Conclusion

The devastating and widespread health consequences of the still ongoing COVID-19 pandemic and continued uncertainty surrounding this novel disease are unprecedented challenges to public health. Since the start of the pandemic, managing the ensuing deluge of both information and misinformation about COVID-19 has also required unparalleled efforts from health authorities.

In detailing the development of *Chloe for COVID-19*, we provide an encouraging example of collaboration leading to a rapidly scalable, safe, secure, and useful human-enhanced digital tool to address the information needs of Canadians during this pandemic. Continuing to leverage the potential of such chatbots to become trusted public health allies will support ongoing infodemic management strategies. As research on the safety and acceptability of these tools advances, we anticipate a growing role for chatbots, both standalone and integrated with human-provided care, in the current pandemic, in particular with vaccination rollout and ongoing efforts to combat misinformation.

Acknowledgments

The authors wish to acknowledge NuEcho and Clinia for their collaboration and their important contribution to *Chloe's* development. Financial support for this project was received from ScaleAI and Mitacs.

Conflicts of Interest

All four authors of this paper work at Dialogue Health Technologies Inc. but do not have any personal financial interests related to the contents of this paper.

References

1. Vraga EK, Jacobsen KH. Strategies for Effective Health Communication during the Coronavirus Pandemic and Future Emerging Infectious Disease Events. World Medical & Health Policy 2020 Jul 29;12(3):233-241. [doi: 10.1002/wmh3.359]
2. Infodemic Management. World Health Organization. 2021. URL: https://www.who.int/teams/risk-communication/infodemic-management [accessed 2021-01-10]
3. Info-Santé 811 ne suffit pas à la demande. Radio Canada. 2020 Mar 12. URL: https://ici.radio-canada.ca/nouvelle/1662861/811-engorge-ressources-mccann [accessed 2021-01-20]
4. Herriman M, Meer E, Rosin R, Lee V, Washington V, Volpp KG. Asked and answered: Building a chatbot to address covid-19-related concerns. NEJM Catalyst Innovations in Care Delivery 2020:1 [FREE Full text] [doi: 10.1056/cat.20.0230]
5. Abdul-Kader SA, Woods JC. Survey on Chatbot Design Techniques in Speech Conversation Systems. International Journal of Advanced Computer Science and Applications 2015;6(7):1. [doi: 10.14569/ijacsa.2015.060712]
6. Dale R. The return of the chatbots. Nat. Lang. Eng 2016 Sep 13;22(5):811-817. [doi: 10.1017/s1351324916000243]
7. Miner AS, Laranjo L, Kocaballi AB. Chatbots in the fight against the COVID-19 pandemic. NPJ Digit Med 2020;3:65 [FREE Full text] [doi: 10.1038/s41746-020-0280-0] [Medline: 32377576]
8. Chloe. Dialogue Health Technologies. URL: https://covid19.dialogue.co/#/ [accessed 2021-01-10]

9. Liu J, Gao Z, Kang Y, Jiang Z, He G, Sun C, et al. Time to Transfer: Predicting and Evaluating Machine-Human Chatting Handoff. AAAI 2021:5841-5849 [FREE Full text]
10. Harms J, Kucherbaev P, Bozzon A, Houben G. Approaches for Dialog Management in Conversational Agents. IEEE Internet Comput 2019 Mar;23(2):13-22. [doi: 10.1109/mic.2018.2881519]
11. Rasa. 2021. URL: https://rasa.com/ [accessed 2021-04-02]
12. Amazon Mechanical Turk. URL: https://www.mturk.com/ [accessed 2021-01-10]
13. Devlin J, Chang M, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. Cornell University. 2019. URL: https://arxiv.org/abs/1810.04805 [accessed 2021-01-10]
14. Breunig M, Kriegel H, Ng R, Sander J. LOF. SIGMOD Rec 2000 Jun;29(2):93-104. [doi: 10.1145/335191.335388]
15. Brownlee J. What is the Difference Between Test and Validation Datasets? Machine Learning Mastery. 2020 Aug 14. URL: https://machinelearningmastery.com/difference-test-validation-datasets/ [accessed 2021-06-25]
16. Bronzi M, Pinto J, Ghosn J, Subakan C, Sharma P, Lu X. Poster 58: A Question Answering System in Response to the COVID-19 Crisis. 2020 Presented at: Montreal AI Symposium 2020; September 11, 2020; Montreal, Canada URL: https://drive.google.com/file/d/1chJyp6mEhieL13EzbX_Xpf8m3MenDuDE/view?pli=1
17. Clinia. URL: https://clinia.com/en-ca [accessed 2021-01-10]
18. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y, et al. Conversational Agents in Health Care: Scoping Review and Conceptual Analysis. J Med Internet Res 2020 Aug;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]
19. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. Digit Health 2019;5:2055207619871808 [FREE Full text] [doi: 10.1177/2055207619871808] [Medline: 31467682]
20. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M. Effectiveness and Safety of Using Chatbots to Improve Mental Health: Systematic Review and Meta-Analysis. J Med Internet Res 2020 Jul 13;22(7):e16021 [FREE Full text] [doi: 10.2196/16021] [Medline: 32673216]
21. Vedhara K. Experts create 'chatbot' to address people's concerns about Covid-19 vaccines. University of Nottingham. 2021 Apr 21. URL: https://www.nottingham.ac.uk/news/vaccine-hesitancy [accessed 2021-06-20]
22. Kinsella B. Nuance Launches COVID-19 Vaccine Bot and Vaccine Assistant. Voicebot.ai. 2020 Dec 28. URL: https://voicebot.ai/2020/12/28/nuance-launches-covid-19-vaccine-bot-and-vaccine-assistant/ [accessed 2021-06-20]
23. Altay S, Hacquin A, Chevallier C, Mercier H. Information Delivered by a Chatbot Has a Positive Impact on COVID-19 Vaccines Attitudes and Intentions. PsyArXiv. 2021 Apr. URL: https://psyarxiv.com/eb2gt/ [accessed 2021-06-25]
24. Phillips N. The coronavirus is here to stay - here's what that means. Nature 2021 Feb;590(7846):382-384. [doi: 10.1038/d41586-021-00396-2] [Medline: 33594289]

Abbreviations

AI: artificial intelligence
BERT: Bidirectional Encoder Representations from Transformers
FAQ: frequently asked questions
NLU: natural language understanding
OOD: out-of-distribution
Q&A: question and answer

Edited by C Basch; submitted 20.01.21; peer-reviewed by A Kocaballi, H Belli; comments to author 17.03.21; revised version received 27.06.21; accepted 10.07.21; published 21.09.21.

Please cite as:
Siedlikowski S, Noël LP, Moynihan SA, Robin M
Chloe for COVID-19: Evolution of an Intelligent Conversational Agent to Address Infodemic Management Needs During the COVID-19 Pandemic
J Med Internet Res 2021;23(9):e27283
URL: https://www.jmir.org/2021/9/e27283
doi:10.2196/27283
PMID:34375299

Original Paper

Development, Feasibility, Acceptability, and Utility of an Expressive Speech-Enabled Digital Health Agent to Deliver Online, Brief Motivational Interviewing for Alcohol Misuse: Descriptive Study

Maya Boustani[1], PhD; Stephanie Lunn[2], MSc; Ubbo Visser[3], PhD; Christine Lisetti[2], PhD

[1]Department of Psychology, Loma Linda University, Loma Linda, CA, United States

[2]Knight Foundation School of Computing and Information Sciences, Florida International University, Miami, FL, United States

[3]Department of Computer Science, University of Miami, Miami, FL, United States

Corresponding Author:
Maya Boustani, PhD
Department of Psychology
Loma Linda University
11130 Anderson St
Suite 117
Loma Linda, CA, 92350
United States
Phone: 1 909 558 7680
Email: mboustani@llu.edu

Abstract

Background: Digital health agents — embodied conversational agents designed specifically for health interventions — provide a promising alternative or supplement to behavioral health services by reducing barriers to access to care.

Objective: Our goals were to (1) develop an expressive, speech-enabled digital health agent operating in a 3-dimensional virtual environment to deliver a brief behavioral health intervention over the internet to reduce alcohol use and to (2) understand its acceptability, feasibility, and utility with its end users.

Methods: We developed an expressive, speech-enabled digital health agent with facial expressions and body gestures operating in a 3-dimensional virtual office and able to deliver a brief behavioral health intervention over the internet to reduce alcohol use. We then asked 51 alcohol users to report on the digital health agent acceptability, feasibility, and utility.

Results: The developed digital health agent uses speech recognition and a model of empathetic verbal and nonverbal behaviors to engage the user, and its performance enabled it to successfully deliver a brief behavioral health intervention over the internet to reduce alcohol use. Descriptive statistics indicated that participants had overwhelmingly positive experiences with the digital health agent, including engagement with the technology, acceptance, perceived utility, and intent to use the technology. Illustrative qualitative quotes provided further insight about the potential reach and impact of digital health agents in behavioral health care.

Conclusions: Web-delivered interventions delivered by expressive, speech-enabled digital health agents may provide an exciting complement or alternative to traditional one-on-one treatment. They may be especially helpful for hard-to-reach communities with behavioral workforce shortages.

(J Med Internet Res 2021;23(9):e25837) doi:10.2196/25837

KEYWORDS

digital health agent; virtual health assistant; online intervention; alcohol abuse; brief intervention; motivational interviewing; intelligent virtual agent; embodied conversational agent

Introduction

Background

Alcohol use disorder (AUD) affects 10%-20% of men and 5%-10% of women over their lifetime, and 26.4% of adults engage in binge drinking. AUD is the third leading preventable cause of death [1], with driving under the influence accounting for 31% of driving fatalities. In addition to personal costs associated with AUD, alcohol abuse costs the US economy an average of $249 billion per year. Motivational interviewing (MI) [2] is an effective and scalable intervention for AUD [3].

It is a client-centered counseling style that is directive and elicits behavior change by helping clients explore ambivalence and resolve it in order to develop *intrinsic* motivation to change. Adaptations of MI have bourgeoned to meet the need for motivational interventions that are brief and thus compatible within primary care settings [4]. Brief motivational interviewing (BMI) interventions include MI's style of communication (communicating empathy, increasing discrepancy, rolling with resistance, and supporting self-efficacy) with the common underlying elements of effective brief interventions (eg, feedback, menus of options for changing respectful of current readiness to change, supportive advice). BMI can be delivered in multiple settings, as both a standalone intervention and in combination with other strategies for substance use disorders, such as cognitive-behavioral therapy, and has been found to be effective across meta-analyses [3-5]. Despite the high rates of alcohol use and availability of these effective interventions, only 1 in 10 individuals with AUD receive care [6,7].

Barriers to Care

A number of barriers prevent individuals from accessing the treatment they need, including acknowledging the need for treatment [8], availability of trained providers [9], proximity of providers, access to transportation, affordability, insurance coverage, scheduling, and stigma [10]. Individuals living in rural settings or in poverty — where alcohol abuse is more prominent — are disproportionately impacted by these barriers [9]. In rural settings in particular, anonymity is more difficult [11] and increases stigma around help-seeking. Lack of flexible scheduling options for individuals who work full time further exacerbate barriers to treatment [9]. Finally, when individuals do access treatment, it is not always an evidence-based treatment — further complicating issues around access to quality care.

Digital Health Interventions

Digital health interventions (DHIs) are interventions that are delivered via digital platforms (eg, applications, websites, mobile devices). Unlike telehealth (where a live provider meets with a consumer via a video chat), DHIs do not rely on a human provider to deliver services. As such, they have the potential to reduce a number of barriers associated with location (can be accessed from anywhere), scheduling (can be accessed at any time), stigma (can be accessed anonymously from the privacy of one's home), and cost (most are affordable or free). Past research indicates that consumers tend to be more truthful when disclosing possibly stigmatizing information such as a drug or alcohol disorder to a computer versus a human [12-16] — providing another advantage to DHIs as they can be more informed about consumers' at-risk behaviors than a human provider.

A review of DHIs [17] indicates that these interventions range from brief screening tools to several months of structured activities. Content includes screening and self-monitoring, personalized normative feedback, goal-setting activities, and interactive journaling. Benefits include reductions in alcohol consumption and consequences of heavy drinking. Despite all the benefits associated with DHIs, they are associated with high dropout rates [18]. For instance, a systematic review of the use of mental health support smartphone applications indicates that only 4% of users engage daily with the applications [19]. Researchers suggest that the use of a DHI without the support or recommendation of a mental health professional may limit its use [19,20]. Mohr et al [21] pointed out that improvement in mental health conditions tends to require continued behavior change over many weeks or months, yet mental health technologies are mainly didactic, thereby not ideal for supporting engagement and behavior change. Most behavioral health technologies require some human backing from a mental health professional to sustain engagement. Qualitative studies point to lack of motivation due to frustrating technology, inadequate content, competing priorities, and lack of face-to-face encounters [22,23]. This limits the promise of DHIs as a scalable solution to increase access to care, which our approach aims to improve.

A review of DHIs designed specifically for MI [24] further points out that, given the important emphasis on the interpersonal therapeutic communication style that is a core aspect of MI, delivering MI through these different modalities is difficult. For instance, can the "MI spirit," or relational aspects, happen digitally? Therefore, comprehending the type of technology used to deliver relationally focused treatments provides an understanding of how technology may be used to replace face-to-face contact. The study found that DHIs for MI vary greatly in terms of technology and richness of the media used, ranging from text-only to audio files, video files, and interactive animated characters, and that the most common feature of these technologies was personalized feedback to the participant based on their input. Only a subset of a few programs used videos (eg, a "video-doctor" actor playing a doctor's responses in MI style) or animations (eg, a robot's head with facial expressions supporting participants during the intervention). These media were always developed with the aim to mirror interpersonal communication. Our current focus on developing and evaluating 3D virtual characters able to deliver a BMI, with appropriate facial expressions, body gestures, speech synthesis, and speech recognition in real time, aims at providing awareness into how technology may be used to replace face-to-face contact.

The review by Shingleton and Palfai [24] also found that, while surveyed articles explained methods for some aspects (eg, automatic computer prompts, chat rooms, emails, videos, animated characters) to deliver MI, most articles did not explain how they translated MI principles into the DHI nor whether or how the relational components were resolved. Translating aspects that require the MI spirit such as "expressing empathy" or "collaboration" to technology — versus consolidating commitment to change and developing discrepancy, on which most studies focused — is particularly difficult to implement in a piece of software. One conclusion was that future researchers need to detail both, not only how the technical aspects (eg, chat rooms, emails) are delivered but also how the relational aspects (eg, emoticons, videos of talking narrators) are delivered in order to increase the human-like discourse with the DHI. Asking questions to help understand how participants felt about and during the interaction were also encouraged as important "soft" outcomes to uncover ways to increase the "spirit" of MI within technology. As highlighted by Mohr et al [21], while usability testing has increased in recent years, the

design of DHIs has generally not included input from end users. Our focus on technology outcomes in this article aimed at providing insight into these "soft" outcomes, by explicitly asking users of our DHI-specific survey questions about their experience with an expressive, animated embodied conversational agent (ECA) in order to inform the impact of our DHI current design, our future redesigns, and other researchers' DHIs.

Embodied Conversational Agents

ECAs (also known as virtual intelligent agents or virtual humans) are simulated human characters that may have the potential to increase consumer engagement in DHIs [25]. Unlike avatars — which are virtual entities that represent and are controlled by the user (popular in video gaming) — ECAs are virtual entities of their own that interact with a consumer autonomously and anonymously. They are developed with the aim to look, sound, and behave as closely to humans as possible. Their ability to hold conversations is still limited [26] but advancements in natural language processing and artificial intelligence (AI) hold promise in the future of ECAs as an alternative solution to traditional therapy for mental health and substance abuse concerns [27]. ECAs have the ability to use sophisticated multimodal communication to build rapport [28-41], communicate empathically [32-35,38], and engage in social talk [42-46]. Despite the promise of ECAs, research around the acceptability, feasibility, and utility of such technology by consumers of behavioral health interventions is limited. Exceptions can be found in a few studies using 3D ECAs with realistic animated facial expressions and mirroring of the users' facial expressions [25], a study including an ECA with a dialog management system allowing users to answer freely to the ECA (albeit without full robustness for broad dissemination without synchronous human support) [26], and a few others using simple ECAs [43,47-50].

In spite of their success, however, ECA development did not scale with the now abundant internet devices (smartphones, laptops) and the latest progress in 3D graphics. Some attempts have been made to build web-based, 3D ECAs [51-53]. However, their implementation is still very basic, and they do not offer an integrated framework for web-based ECA development, including the ability to recognize and synthesize social cues in real time during spoken dialog, which is a significant technical challenge and which our ECA provides.

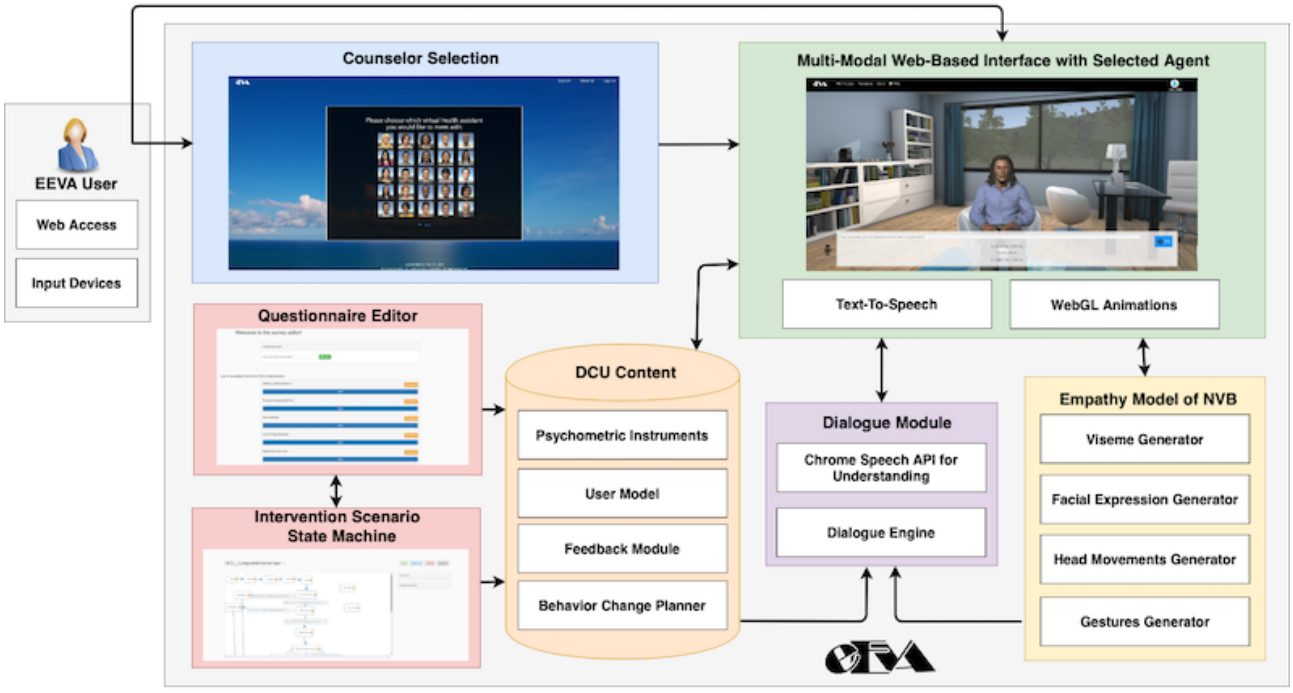
Current Study

This study aimed to fill the gap in knowledge of using ECAs in behavioral health contexts by establishing the acceptability, feasibility, and utility of using ECAs by consumers undergoing a BMI intervention for alcohol abuse. BMIs are highly structured (*assessment* of, followed by normative *feedback*, then *menu* of change options), making them amenable to delivery via DHI [22], particularly if the "MI spirit," or relational components, can be captured without face-to-face contact. One such BMI, namely the Drinker's Check-Up (DCU) [2] is the intervention used for this work. DCU has been computerized as a menu-based, text-only program delivered online that targets alcohol abuse, reducing drinking by an average of 50% at a 12-month follow-up [50]. The DCU is one of the 2 English-language, web-based DHIs designed for the public that have been tested in randomized controlled trials (RCTs) [7]. We therefore chose to study how the delivery of the DCU by an ECA will be perceived by its users, given that its nonverbal and other media features aim to address the observed limitations of the use of avatars in DHIs for MI that do not strengthen the social relationship with the user [50].

Using the technology acceptance model (TAM) [54,55] to guide our work, this study enabled us to determine if the ECA designed by our team using the empathic embodied virtual agent (eEVA; see Figure 1 and Figure 2) framework for building digital health agents [56] has enough personal characteristics and social abilities (eg, open-minded, supportive, respectful, friendly) to give users a positive experience (acceptability). The TAM stipulates that user acceptance can be predicted by the perceived usefulness (utility) and perceived ease of use (feasibility) of the technology. As such, we were interested in learning more about consumers' perceived positive experience (acceptability), usefulness of eEVA (utility), and ease of use (feasibility) to better understand acceptability and potential for adoption of the technology. Having technology that consumers like and find easy to use and helpful increases the potential for adoption, which, in turn, increases access to care. Increases in access to care have the potential to improve health outcomes for alcohol users. Prior studies have found that MI for alcohol use (including online delivery via textual interface) improves health outcomes [57].

Figure 1. Our empathic embodied virtual agent (eEVA) delivering a brief motivational interviewing behavior change session.

Figure 2. Empathic embodied virtual agent (eEVA) system overview. API: application programming interface; DCU: Drinker's Check-Up; NVB: nonverbal behavior.



Methods

Intervention

DCU

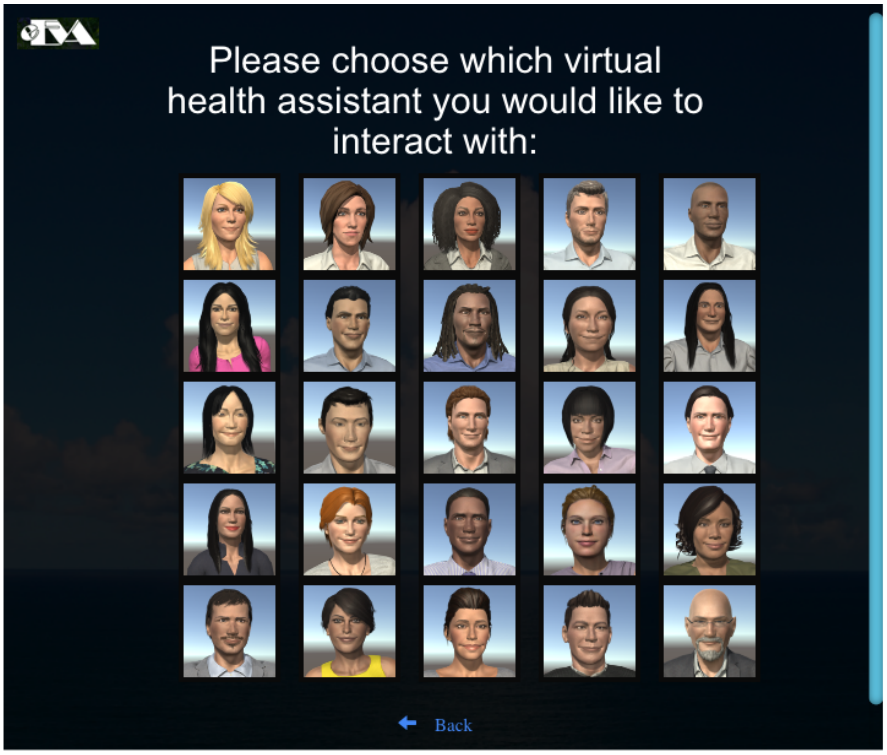
The intervention is based on DCU — an evidence-based intervention that uses strategies from MI [57]. First, users provide detailed information about their drinking (eg, own drinking patterns or issues, family history of alcohol use). Next, they receive individualized feedback about their drinking habits, including information about risk factors and consequences [58]. Finally, they resolve their ambivalence about whether to change their drinking, plan for a change, and set goals for change. The intervention provides resources to help users with changing their drinking [59]. However, the DCU does not tell the clients what to do or not to do — it is up to the user whether they want to change their alcohol use. The DCU has been studied extensively and led to reductions in the quantity and frequency of drinking by 50% throughout a 12-month follow-up period, when compared to a waitlist control group [57]. The DCU is 1 of the 2 web-based DHIs that have the strongest evidence supporting their efficacy based on RCTs [7].

In this study, the DCU was delivered via an ECA (namely eEVA) rather than via its traditional textual interface. As a BMI, the eEVA intervention combines MI style of communication with the common underlying elements of effective brief interventions characterized by the acronym FRAMES [2]: *Feedback* about client's individual status is personalized and stored in a user model database, keeping a record of users' answers for the next session(s); *Responsibility* for changing is left with the individual, and the language used throughout the intervention reflects this (eg, "I will not pressure you in any way"); *Advice* is provided in a supportive manner, with empathic choice of words and supportive body gestures (eg, leaning forward, head nodding); *Menus* of different options for changing that respect an individual's readiness to change are offered; *Empathic* style of communication is central to the individual-clinician relationship, and it is conveyed by the ECA's verbal utterances (eg, spoken reflections), nonverbal behavior (NVB; eg, smiling facial expressions, lean forwards, hand flips, nodding at appropriate times), and empathic choice of wording (eg, "It might be surprising to you to know that you are in the top percentile in drinking compared to people of your gender and age; you might want to review your answers again …"); and *Self-efficacy* is nurtured and emphasized throughout, including with choice of words and positive facial expressions (eg, various head nods and smiling facial expressions).

Technical Implementation of the Intervention

Our eEVA framework (shown in Figure 2) provides (1) a realistic 3D WebGL graphics virtual environment with a realistic virtual office environment that can be "inhabited" by 1 of the 25 available racially diverse ECAs (shown in Figure 3), each with physiologically realistic Facial Action Coding System–validated facial expression animations and full body animations; (2) real-time speech recognition of the user's answers; (3) text and multiple choice input; (4) voice synthesis for the ECA's spoken utterances; (5) ECA's lip synchronization between phonemes and visemes; (6) ECA's adaptive nonverbal responses such as head nods or leaning forward depending upon the utterance dialog act; (7) configurable dialogue content; and (8) ability to capture and process users' social cues such as facial expression recognition (which will be enabled in a future study).

Figure 3. Menu of diverse empathic embodied virtual agent (eEVA) embodied conversational agents consumers can select.



Technically speaking, the framework consists of 3 main components. First, the application layer consists of a modular client-side JavaScript mainframe that controls the multimodal user interface, audio and video input, graphical user interface (GUI) interaction, and services such as speech recognition and speech synthesis. Second, the JavaScript mainframe handles execution of a scenario (the content of the DCU in this study) — a collection of state machines that are created by developers. Third, the scenario states can be constructed to pull information (eg, the ECA's speech, graphics to show) from the data layer — a database of content.

The backbone of the client-side application is a JavaScript framework that handles the formation of a group of modules and the communication between them. Each module then implements various functionalities, including gaining feedback from the user (eg, asking to access microphone and camera) and processing input information (eg, analyzing users' responses, extracting facial expressions if desired); determining how to answer the user (eg, words agent should say, NVBs); and answering the user through a multimodal 3D-embodied ECA, with speech synthesis, NVBs, and multimedia content (eg, text, images, and videos).

This results in an interactive online application that can run on numerous platforms such as desktop, cell phone, autonomous robotic agent, and potentially smartwatch integrations (Figure 4). In addition, the user can also choose between a collection of 3D virtual characters to interact with — of different genders, races, and appearances. To personalize the eEVA system further, favorite chosen characters are remembered and displayed after login during the next interaction with the system. Distributing the framework core (eg, 3D character, perception, behavior) to consumer devices makes this technology scalable, with little to no overhead with additional users. Computer-intensive functionality such as speech and face recognition is asynchronous via web services or with built-in functionalities in the browser.

Figure 4. Empathic embodied virtual agent (eEVA) running on different platforms: (a) desktop, (b) mobile phone, (c) autonomous robot, (d) smartwatch concept.

Unlike traditional ECAs, the eEVA design follows common modularity patterns found in robotics platforms such as ROS [60], allowing us to generate collections of modules to cover a diversity of application use cases, such as various browsers, many internet bandwidth limitations, and interaction capabilities. For instance, when using speech recognition, to provide the transcript of the user's spoken utterances to be used by the ECA application, based on browser capabilities, a specialized module can be used to either interface with the Web Speech API or to use another service such as Watson Speech to Text. The advantage of this design is the seamless passing from one module implementation to another, including at runtime, without affecting the rest of the application.

To model social interaction, 3 main categories of module functionality are necessary, namely input/sensing modules (for perceiving social cues from the user in real time); social interaction decision-making modules, including an ECA behavior module, vocal command interpretation, and the scenario controller; and output/actuator modules for actually expressing verbal and nonverbal cues to the user. The functionalities of the main modules used in the current version of eEVA are listed in Table 1. Most modules have simple functions to retrieve or display information from and to the user or call functions from libraries (third-party or in-house) or services.

Table 1. Listing of the most significant modules and their function descriptions used in the empathic embodied virtual agent (eEVA) for our health agent.

Module	Function description
Input/Sensing Module	• User microphone interface using WebRTC application programming interface (API) • Speech recognition using Google Chrome API • Interface with CoreNLP • Graphic user interface (GUI) for direct user input (text, buttons)
Social Interaction Decision-Making Modules	• Vocal command interpretation • Embodied conversational agent's (ECA) behavior (gesture and facial animations) • Scenario controller (state machine execution)
Output/Actuator Modules	• Speech synthesis • 25 webGL 3D eEVA ethnically diverse characters

We tested 2 types of network connections: broadband and 4G mobile data. The majority of the launch time consists of loading the 3D character and surrounding virtual environment, which takes about 30 seconds and 25 seconds on 4G and broadband, respectively. The experiments (Table 2) showed that the main distributed functionalities of the eEVA framework allow real-time interaction and adequate loading times. This was echoed by users, as discussed in our Results section.

Table 2. Average response time and standard deviation analysis for the empathic embodied virtual agent (eEVA) using 4G or broadband connections over the internet between North America and Europe, with caching disabled (first run).

Functionality	Time on 4G mobile data (milliseconds)	Time on broadband internet (milliseconds)
Unity 3D character, mean (SD)	30018 (663)	24626 (1910)
TTS[a] (sentence), mean (SD)	939 (381)	551 (141)
TTS (word), mean (SD)	72 (40)	44 (23)
Speech recognition	~30 (offline processing)	N/A[b]
Entire HTTP request, mean (SD)	1124 (166)	784 (66)
DOM[c] loading, mean (SD)	2313 (80)	1635 (224)

[a]TTS: text to speech.

[b]N/A: not applicable.

[c]DOM: Document Object Model.

Model of Empathic Verbal and Nonverbal Behavior

To simulate some of the communication psycholinguistic signals of a counselor delivering a BMI, we first videotaped BMI sessions between a live licensed counselor and a client. Then, a clinical expert reviewed the videotapes to code verbal reflections and NVB. From these, the expert generated a set of rules for basal behaviors of the health agents. Based on the codes of verbal and NVBs, eEVA was implemented with the following verbal reflections: "Ah." "Alright." "Okay." "Good." "Sounds Good." "Oh, okay." "Great!" "Thanks for letting me know." "Oh, I see." "Okay, thanks!"

In addition, the following NVB animations were synthesized on the agent's face: smile, facial expressions, hand gestures (typing on a computer at a desk, hands resting on the agents' legs, formless flick, two-handed flip, two-handed contrast gesture), body leans (forward), head gestures (nod, shake, nonshake), and eyebrow movements (up, neutral, and down), which our results (discussed later) showed are conducive of a positive experience for the user with the agent in the given health care context.

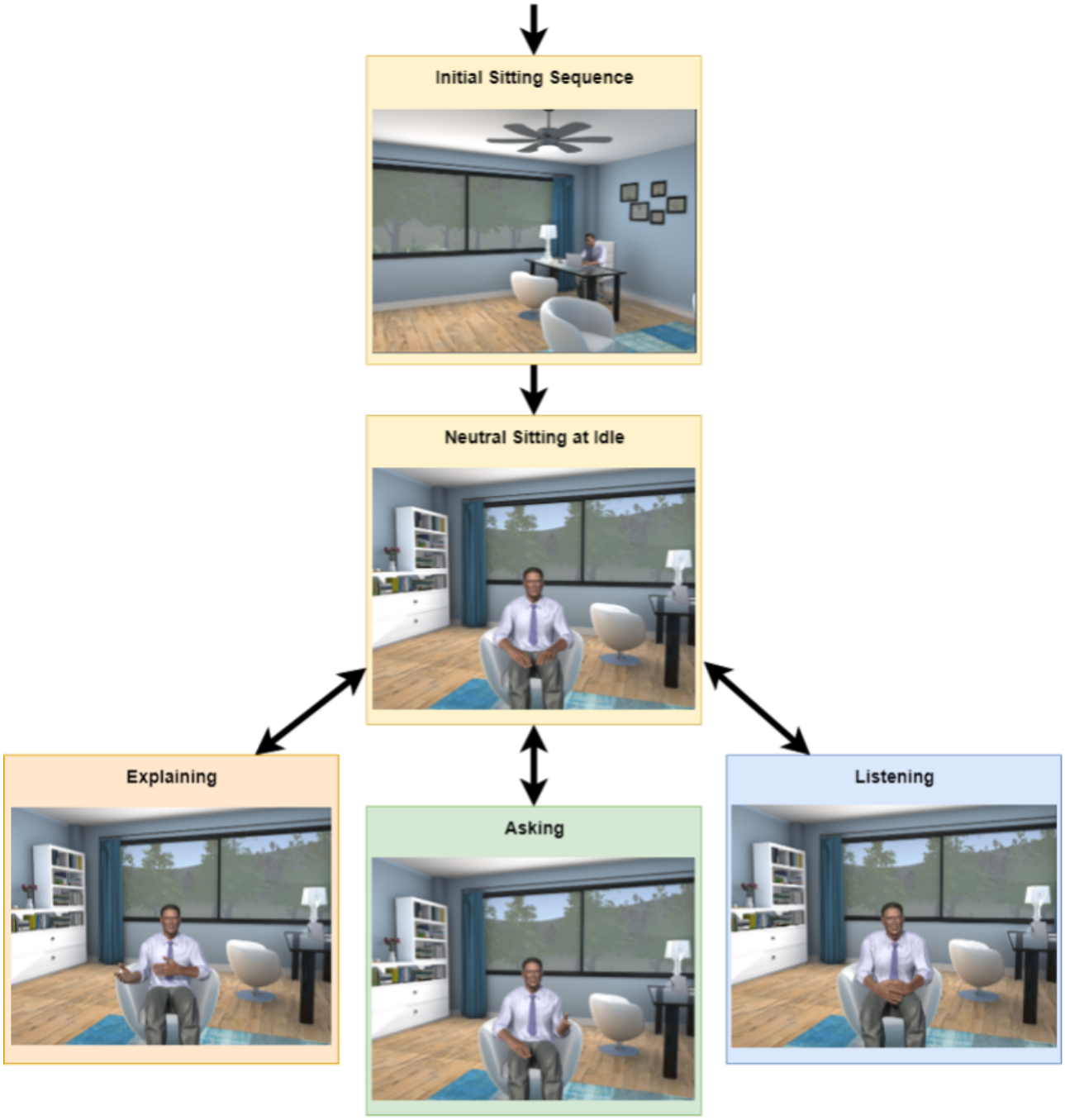
Since it was determined that head nods are critical to portraying (some level of) active listening, we sought to offer 3 variants depending on the user's chosen responses. We created a set of nods using established emotional expressions governed by activation of specific individual facial movement animations. All 3 of the head nods included activation of head down and eyes down. However, depending on the type of reaction required, these also included facial expressions (eg, smile, mild concern).

The patterns observed in the videotapes of the counselor-client session also directed us towards creating rules about when certain statements should be made, to ensure the counselor did not appear judgmental and to make the interaction appear more natural. In all scenarios, the counselor began seated at the desk while typing on the computer, then looked up and moved to the chair closer to where the user perceives they are sitting during an initial greeting. Once seated, the counselor began with a greeting introduction and then moved on to delivering the DCU. Verbal responses to user responses were applied based on the "score" of each question to provide nonjudgmental reactions for higher scores that might indicate a problem and positive reactions for scores that might suggest healthy consumption

levels. For example, “Sounds good” was used in response to a user mentioning that they wanted to change.

In parallel, NVBs were applied using a set of states that were determined as appropriate given the context of the interaction as shown in Figure 5: Neutral, Explaining, Asking, and Listening. In the *Neutral* state, the counselor spoke and used a smile, a gaze, facial expressions, head gestures, or eyebrow movements. In the *Explaining* state, which was activated during long periods of speaking, the two-handed flip and two-handed contrast gesture were applied. When the counselor posited a question to the user, the *Asking* state was initiated, which included a single hand formless flick. While the counselor waited for the user to respond to a question, the *Listening* mode was initiated, which included a leaning forward gesture. The leaning gesture remained in effect until a choice was made, at which point the body resumed an upright sitting position.

Figure 5. Defining nonverbal behaviors for virtual health agents. All undergo the initial sitting sequences and then assume a neural sitting at idle position. From here, the agent can enter either the be explaining, asking, or listening loop.



Participants

Participants were alcohol users aged 21 to 55 years and recruited online to participate in the intervention. Participants had to have engaged in heavy drinking (consumed 5 drinks in one sitting at least once in the past year), not currently be receiving treatment for their AUD, and not have a medical condition for which alcohol use would be contraindicated. Users were also screened for severe mental illness. We recruited 51 participants as part of a larger RCT of the effectiveness of this program in reducing

alcohol abuse. Participants were 62% (32/51) male, were 32% (19/51) female, and had a mean age of 28 (SD 15.8) years. Participants reported their race as White (21/51, 42%), Black (12/51, 24%), Asian (3/51, 7%), Other (2/51, 3%), and no response (12/51, 24%). Participants reported their ethnicity as non-Hispanic (43/51, 84%), Hispanic (8/51, 15%), and no response (2/51, 1%). Marital status was reported as married (23/51, 46%), single (17/51, 34%), divorced (3/51, 7%), widowed (1/51, 1%), or no response (6/51, 12%). Their education level was reported as high school (7/51, 14%), some college (13/51, 25%), Associate's degree (10/51, 19%), Bachelor's degree (18/51, 36%), and Graduate degree (3/51, 6%).

Procedures

Participants were recruited online via targeted advertising on Facebook, offering free treatment for alcohol users and compensation for research participation. When users clicked on the ad, they were redirected to an online screener (on Qualtrics) to ensure they were eligible for the study. If they were eligible, users were randomly assigned to receive the same DCU MI intervention delivered online either by an ECA (eEVA) or a textual interface. Participants in this study were those who were assigned to eEVA, since the control group assigned to the text-only interface could not comment on the ECA's social features that they did not see. Once randomized, participants were provided with a username and password to enter the DHI platform eEVA. Participants had to log on, enter demographic information, and begin the intervention. After completing the intervention (participants were given 1 week to complete), they were redirected to an online survey on Qualtrics to provide feedback about their experience with the intervention.

Measures

After completing the intervention, participants were asked to provide feedback on the feasibility, acceptability, and utility of the technology. Specifically, we sought feedback regarding their engagement (acceptability), perceived utility, and intent to use the technology (feasibility). Participants completed a questionnaire developed for this study. Since there does not exist, to date, a standardized instrument to evaluate interaction with ECAs of various levels of complexities, we used and adapted relevant existing questionnaires commonly used for the evaluation of human interaction with technologies involving some social cues, whether embodied with graphics or with robot technologies. Questions were based on a combination of the engagement model by O'Brien and Toms [61], Almere model by Heerink et al [62], and "Godspeed questionnaire" by Bartneck et al [63], which has been widely used to evaluate human-technology interactions using 5 key concepts — anthropomorphism, animacy, likability, perceived intelligence, and safety — that have been found useful for interacting with either ECAs or robots. All responses were on a 7-point Likert scale (1=Strongly Agree; 7=Strongly Disagree), with lower scores indicating more desirable findings. There was no cut-off as the measures were combined for the purpose of this study and were meant to provide descriptive feedback. In addition, each question included a blank space with a prompt of "Comments" for participants to provide optional qualitative feedback to each question.

The engagement model by O'Brien and Toms [61] explores acceptability of the ECA via constructs of user engagement with technology. Six attributes of a technology make it more likely that a user will engage with it (challenge using the technology, interest, motivation to use the technology, and appeal of the technology [eg, "I found the health assistant interesting"]). Responses were on a 7-point Likert scale (1=Strongly agree; 7=Strongly disagree).

The Almere model evaluates the user's acceptance of the digital health agent by relying on constructs from the Unified Theory of Acceptance and Use of Technology [12]. Users replied to 13 statements (eg, "I enjoyed participating in this session with the health assistant") that map to constructs that predict intent to use the technology and perceived usefulness of the technology. Responses were on a 7-point Likert scale (1=Strongly agree; 7=Strongly disagree).

The Godspeed questionnaire consists of 12 questions that capture 5 constructs that measure human-like traits of robots, which we adapted for ECA (eg, "The health agent seemed warm"): (1) anthropomorphism (eg, moving rigidly or moving elegantly), (2) animacy (eg, mechanical or organic), (3) likeability (eg, unfriendly or friendly), (4) perceived intelligence (eg, incompetent or competent), and (5) perceived safety (eg, anxious or calm). Each trait could be rated as being very human-like to very unhuman-like on a 7-point scale.

All questions included a fill-in option for participants to expand on their numerical responses with qualitative feedback if they wanted to. We highlight some of those comments in the Results section.

Results

Acceptability and Utility

Participants reported high acceptability and utility of the technology, as indicated by their scores on the Almere model questions (mean 2.31, SD 1.05). Most participants reported enjoying their interaction with the agent (44/51, 86%) stating "The questions she asked me, no one had asked me before and helped recognize my drinking problem." They thought the agent was both physically appealing (38/51, 74%) and had a pleasant voice (42/51, 82%), stating:

> *He is neat, he has a good haircut, he is well dressed.*
>
> *He appears wise, intelligent, and healthy.*
>
> *The voice was adequate, calm, and confident.*

Furthermore, participants indicated that they found it easy to interact with the health agent (45/51, 89%) and they learned to do so quickly (45/51, 89%), stating:

> *I just followed the instructions and voila!*
>
> *I just waited and followed his instructions; it was easy.*

Participants had more difficulty with the voice feature of the technology, with only 69% (35/51) reporting that they felt like the agent understood them when they spoke into the microphone.

Based on the qualitative feedback, this may have been due to issues with participants' microphone setting rather than the technology itself:

> *I set my microphone up, but it seems there is a problem with it.*
>
> *I tried to speak my answers, but it never worked so I ended up typing them.*

Those who were able to get their microphones to work seemed to have no difficulty speaking to the agent as echoed by their qualitative feedback (eg, "I don't even repeat my answers; the agent understands me very well."). Despite some difficulties with the microphone, 78% (40/51) felt like they could have a conversation with the agent, stating "I felt that he knows me, knows what I want" and reported that they sometimes felt like they were talking to a real person (36/51, 71%), stating that:

> *…his voice sounded quite real…*
>
> *…the gestures he made, way he moved around…*
>
> *he answered me like a real person…*

Participants further reported that the health assistant was friendly (45/51, 89%) and they found it to be useful (44/51, 88%) because:

> *…he explained things that I did not know…*
>
> *…it helped me recognize that I have a problem…*
>
> *I learned a lot.*

Participants reported that they were comfortable disclosing information about their drinking to their digital health assistant (48/51, 93%), with 83% (41/51) reporting that they were *more* comfortable disclosing their drinking to the digital health assistant over their medical doctor. Indeed, participants highlighted:

> *The assistant gives me a level of trust that I don't have with other humans.*
>
> *It is more easy talking to (the health agent) than to a real person.*
>
> *I don't feel like they are judging me.*

Finally, participants reported trusting the advice the health agent gave them (43/51, 85%) and that they planned on following that advice (44/51, 86%), stating:

> *I think she is sincere and wants to help me with my problems.*
>
> *…because it is based on facts and studies and that is real and valid information for me.*

One participant noted "I didn't feel like I was given advice, more like information to be able to make my own decision. I was the one with the power to give myself advice." — perfectly capturing the intent of MI.

Engagement

Participants were highly engaged with the DHI, as indicated by their score on the engagement questions (mean 2.86, SD –0.96), indicating that the majority of participants agreed with statements around how engaged they were. Specifically, 69% (35/51) were not worried about making mistakes while using the technology, stating:

> *At first, it was a little bit intimidating, but then I felt confident.*
>
> *The assistant feels understanding, attentive, very friendly.*

A majority (44/51, 86%) thought it was a good idea to use the health assistant, reporting:

> *He is kind of like a home counselor who works with reliable information and statistics.*
>
> *It is practical, easy to use, and guides the person on what to do without forcing us to make a final decision.*

Participants felt that the system could be adaptive to their needs (46/51, 90%), stating that "it could be adapted to other health problems like smoking."

Finally, 88% (45/51) found the health assistant to be interesting, indicating "I was impressed by the way it converts my answers into figures and important information for my health" and said they would interact with the agent again (43/51, 85%):

> *Setting a new exchange with the health assistant would help me to reach my goal.*

Impressions of the Digital Health Agent

Participants reported a high number of human-like traits on the Godspeed questions (mean 2.07, SD 0.89). Participants reported that the agent moved appropriately (43/51, 85%) and seemed warm (46/51, 90%), responsive (45/51, 89%), knowledgeable (47/51, 92%), relaxed (46/51, 90%), flexible (42/51, 83%), honest (46/51, 90%), respectful (46/51, 90%), confident (47/51, 92%), interested (44/51, 86%), open-minded or nonjudgmental (43/51, 84%), and supportive (45/51, 89%).

Overall, participants' responses to the questionnaires and qualitative feedback indicated that they found the delivery of an MI intervention by a digital health agent over the internet to be acceptable, be engaging, and have features that are close to human-like.

Discussion

Our goal with this study was to understand if the technology we developed was feasible (able to be implemented online), acceptable, useful, and easy to use by consumers. As such, we focused largely on the technology aspects of the intervention.

Principal Findings

This study provides an optimistic outlook for the use of digital health agents to deliver brief online interventions in the future. Consumers overwhelmingly reported positive experiences in their interactions with the agent, with many reporting that they trusted the agent and felt that they could more comfortably disclose information that they may not have disclosed to a human provider. This echoes what has been found in the literature around disclosing to computers versus humans [12,14,64,65].

Further, participants tended to attribute many human-like traits to their agent (eg, friendly, trustworthy, kind) and commented

positively on the physical appearance, voice, and physical gestures of the agent. Our team engaged in coding of verbal reflections, hand gestures, and facial expressions of a real therapist to enhance the digital agent's nonverbal communication to resemble what a therapist might do in session [66].

Implications for this work are important. Given the shortage of mental health workforces in many locations, digital health agents may provide an acceptable complement to traditional face-to-face therapy, reducing demand for higher levels of care, where a digital health agent can act as a clinician "extender" to deliver booster sessions. Similar to telehealth services, digital health agents resolve a number of barriers to care such as transportation and scheduling.

In addition, digital health agents reduce stigma around mental health care, are less costly than one-on-one therapy, and can be scaled out and disseminated. For individuals with high levels of social anxiety, digital health agents may provide them with a unique opportunity to get help. Given the digital nature of the agent, it is possible to adjust the programming to make the agent able to speak in multiple languages, reducing language barriers for minority and refugee populations. Already, consumers can pick a digital health agent from a library of diverse options of physical visual features (gender, age, race, and ethnicity; see Figure 2) and vocal features (gender). This is an exciting development given the lack of a diverse mental health workforce [67,68].

This study has provided the investigative team with valuable feedback to improve the technology, including improving the flow and tone of the voice, providing questions in text, and formatting the technology for use on mobile phones and with lower bandwidth.

Limitations

Despite these enthusiastic findings and implications, it is important to note a number of limitations. First, digital health agents cannot replace traditional therapy and certainly cannot manage crisis situations. They are well suited for brief, structured interventions, but cannot replace the complex nature of a therapeutic relationship and complex therapeutic interventions such as family therapy and emotion-focused therapy. They were considered and studied in this article as clinician "extenders." Second, this study was conducted with participants in the United States only. It is unclear if technology acceptability would be as high in other countries. Third, the impact of the DHI on actual alcohol outcomes remains unclear. A study is underway to better understand the effects of a digital health agent. Fourth, it is unclear whether ECAs are suitable for various health problems besides alcohol and for various other populations not studied here, such as the elderly or children. Finally, the access to and cost of reliable internet necessary to use ECAs may limit access to some — potentially further increasing the digital divide.

Conclusions

This type of intervention and research on digital health agents in virtual reality over the internet are still in their infancy, and there is much work to be done. For instance, the same BMI intervention delivered in this study could be adjusted to other health behaviors (eg, other substances, medication compliance, weight management). Other interventions could be delivered to screen or treat a variety of problems. Furthermore, our team is working on integrating natural language dialog management features such that the agent will respond to the consumer's answers without needing specific input from which to read. We conducted preliminary work [26] for a speech-enabled ECA for BMI interventions with promising results. However, natural language understanding is still a very open research area of computer science, and its use in DHIs is not robust enough to deploy with real users who need support and help, without the potential frustrations generated by unreliable agent's speech understanding. Our team has also worked on features based on AI that allow a PC-based agent (ie, operating on PC only and not web-based) to pick up on the consumer's facial expressions in real time (eg, if the consumer smiles, the agent smiles back), which has been shown to improve users' engagement with digital health agents [25], and we plan to add this type of feature to our web-based eEVA system.

Despite all these exciting potential developments, it is critical to evaluate DHIs with high levels of rigor before they can be deployed for the population at large. As pointed out by Carroll [7], very few of the many available internet-based interventions have been carefully evaluated in well-controlled clinical trials, and the majority of those studies have been conducted with college populations, bringing into questions the generalization of the results to broader society. The conclusions that can be drawn from many studies are constrained by high levels of dropout, high attrition, and weak control conditions (eg, waitlists). To that end, we will report on the results of the RCT we conducted to assess the efficacy of the eEVA DHI compared to a text-only version of the intervention.

We furthermore consider that these digital health agents and DHIs can only complement the unique experience of psychosocial therapy and serve as "clinician extenders" [7]. As suggested by Mohr et al [21], mental health technologies in general should be considered as sociotechnical systems (or technology-enabled services rather than mere products) that must fit within an ecosystem of mental health services (involving human support and organizational factors). Our aim is to provide relief to a clogged mental health system and provide online access to self-help to individuals who otherwise would not access traditional face-to-face care. Further research on whether DHIs increase access to care by removing the barriers identified earlier (eg, availability and proximity of trained providers, affordability, stigma) or whether they increase the digital divide is needed [21]. Digital health agents, even with integrated AI, will not replace human therapists; they should be considered therapist extenders.

Acknowledgments

The research described in this article was supported by a grant from the US National Science Foundation Award No.1423260 to Florida International University and Loma Linda University.

Conflicts of Interest

None declared.

References

1. Obesity and overweight. World Health Organization. 2011. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2021-07-13]
2. Miller WR, Rollnick S. Motivational interviewing: preparing people for change (2nd edition). New York, NY: The Guilford Press; 2002.
3. Rubak S, Sandbaek A, Lauritzen T, Christensen B. Motivational interviewing: a systematic review and meta-analysis. Br J Gen Pract 2005 Apr;55(513):305-312 [FREE Full text] [Medline: 15826439]
4. Burke BL, Arkowitz H, Menchola M. The efficacy of motivational interviewing: a meta-analysis of controlled clinical trials. J Consult Clin Psychol 2003 Oct;71(5):843-861. [doi: 10.1037/0022-006X.71.5.843] [Medline: 14516234]
5. Heckman CJ, Egleston BL, Hofmann MT. Efficacy of motivational interviewing for smoking cessation: a systematic review and meta-analysis. Tob Control 2010 Oct 30;19(5):410-416 [FREE Full text] [doi: 10.1136/tc.2009.033175] [Medline: 20675688]
6. Acevedo A, Panas L, Garnick D, Acevedo-Garcia D, Miles J, Ritter G, et al. Disparities in the treatment of substance use disorders: does where you live matter? J Behav Health Serv Res 2018 Oct 12;45(4):533-549 [FREE Full text] [doi: 10.1007/s11414-018-9586-y] [Medline: 29435862]
7. Carroll KM. Lost in translation? Moving contingency management and cognitive behavioral therapy into clinical practice. Ann N Y Acad Sci 2014 Oct;1327:94-111 [FREE Full text] [doi: 10.1111/nyas.12501] [Medline: 25204847]
8. Scott CK, Grella CE, Dennis ML, Nicholson L. Linking Individuals with Substance Use Disorders (SUDs) in Primary Care to SUD Treatment: the Recovery Management Checkups-Primary Care (RMC-PC) Pilot Study. J Behav Health Serv Res 2018 Apr 27;45(2):160-173 [FREE Full text] [doi: 10.1007/s11414-017-9576-5] [Medline: 29181779]
9. Wamsley M, Satterfield J, Curtis A, Lundgren L, Satre D. Alcohol and Drug Screening, Brief Intervention, and Referral to Treatment (SBIRT) Training and Implementation: Perspectives from 4 Health Professions. J Addict Med 2018;12(4):262-272. [doi: 10.1097/ADM.0000000000000410] [Medline: 30063221]
10. Valdez LA, Garcia DO, Ruiz J, Oren E, Carvajal S. Exploring structural, sociocultural, and individual barriers to alcohol abuse treatment among Hispanic men. Am J Mens Health 2018 Nov 27;12(6):1948-1957 [FREE Full text] [doi: 10.1177/1557988318790882] [Medline: 30051746]
11. Leggio L, de Witte P, Chick J. Alcohol and Alcoholism: Then, Now and The Future of 'The Red Journal'. Alcohol Alcohol 2018 Nov 01;53(6):637-638. [doi: 10.1093/alcalc/agy075] [Medline: 30339185]
12. Ahmad F, Hogg-Johnson S, Stewart DE, Skinner HA, Glazier RH, Levinson W. Computer-assisted screening for intimate partner violence and control: a randomized trial. Ann Intern Med 2009 Jul 21;151(2):93-102 [FREE Full text] [doi: 10.7326/0003-4819-151-2-200907210-00124] [Medline: 19487706]
13. Card W, Lucas R. Computer interrogation in medical practice. International Journal of Man-Machine Studies 1981 Jan;14(1):49-57. [doi: 10.1016/s0020-7373(81)80032-6]
14. Ghanem KG, Hutton HE, Zenilman JM, Zimba R, Erbelding EJ. Audio computer assisted self interview and face to face interview modes in assessing response bias among STD clinic patients. Sex Transm Infect 2005 Oct 01;81(5):421-425 [FREE Full text] [doi: 10.1136/sti.2004.013193] [Medline: 16199744]
15. Kissinger P, Rice J, Farley T, Trim S, Jewitt K, Margavio V, et al. Application of computer-assisted interviews to sexual behavior research. Am J Epidemiol 1999 May 15;149(10):950-954. [doi: 10.1093/oxfordjournals.aje.a009739] [Medline: 10342804]
16. Newman JC, Des Jarlais DC, Turner CF, Gribble J, Cooley P, Paone D. The differential effects of face-to-face and computer interview modes. Am J Public Health 2002 Feb;92(2):294-297. [doi: 10.2105/ajph.92.2.294] [Medline: 11818309]
17. Rogers MA, Lemmen K, Kramer R, Mann J, Chopra V. Internet-Delivered Health Interventions That Work: Systematic Review of Meta-Analyses and Evaluation of Website Availability. J Med Internet Res 2017 Mar 24;19(3):e90 [FREE Full text] [doi: 10.2196/jmir.7111] [Medline: 28341617]
18. Yeager CM, Benight CC. If we build it, will they come? Issues of engagement with digital health interventions for trauma recovery. Mhealth 2018;4:37 [FREE Full text] [doi: 10.21037/mhealth.2018.08.04] [Medline: 30363749]
19. Kerst A, Zielasek J, Gaebel W. Smartphone applications for depression: a systematic literature review and a survey of health care professionals' attitudes towards their use in clinical practice. Eur Arch Psychiatry Clin Neurosci 2020 Mar;270(2):139-152. [doi: 10.1007/s00406-018-0974-3] [Medline: 30607530]

20. Price M, Gros DF, McCauley JL, Gros KS, Ruggiero KJ. Nonuse and dropout attrition for a web-based mental health intervention delivered in a post-disaster context. Psychiatry 2012;75(3):267-284 [FREE Full text] [doi: 10.1521/psyc.2012.75.3.267] [Medline: 22913502]
21. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three Problems With Current Digital Mental Health Research . . . and Three Things We Can Do About Them. Psychiatr Serv 2017 May 01;68(5):427-429 [FREE Full text] [doi: 10.1176/appi.ps.201600541] [Medline: 28412890]
22. Lie SS, Karlsen B, Oord ER, Graue M, Oftedal B. Dropout From an eHealth Intervention for Adults With Type 2 Diabetes: A Qualitative Study. J Med Internet Res 2017 May 30;19(5):e187 [FREE Full text] [doi: 10.2196/jmir.7479] [Medline: 28559223]
23. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual Research Review: Digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr 10;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
24. Shingleton RM, Palfai TP. Technology-delivered adaptations of motivational interviewing for health-related behaviors: A systematic review of the current research. Patient Educ Couns 2016 Jan;99(1):17-35 [FREE Full text] [doi: 10.1016/j.pec.2015.08.005] [Medline: 26298219]
25. Lisetti C, Amini R, Yasavur U, Rishe N. I can help you change! An empathic virtual agent delivers behavior change health interventions. ACM Trans. Manage. Inf. Syst 2013 Dec;4(4):1-28. [doi: 10.1145/2544103]
26. Yasavur U, Lisetti C, Rishe N. Let's talk! speaking virtual counselor offers you a brief intervention. J Multimodal User Interfaces 2014 Sep 5;8(4):381-398. [doi: 10.1007/s12193-014-0169-9]
27. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. J Med Internet Res 2017 May 09;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
28. Gratch J, Wang N, Gerten J, Fast E, Duffy R. Creating Rapport with Virtual Agents. In: Pelachaud C, Martin JC, André E, Chollet G, Karpouzis K, Pelé D, editors. Intelligent Virtual Agents. IVA 2007. Lecture Notes in Computer Science, vol 4722. Berlin, Heidelberg: Springer Publishing Company; 2007:125-138.
29. Gratch J, Okhmatovskaia A, Lamothe F, Marsella S, Morales M, van der Werf R, et al. Virtual Rapport. In: Gratch J, Young M, Aylett R, Ballin D, Olivier P, editors. Intelligent Virtual Agents. IVA 2006. Lecture Notes in Computer Science, vol 4133. Berlin, Heidelberg: Springer Publishing Company; 2006:14-27.
30. Kang SH, Gratch J, Wang N, Watt JH. Agreeable People Like Agreeable Virtual Humans. In: Prendinger H, Lester J, Ishizuka M, editors. Intelligent Virtual Agents. IVA 2008. Lecture Notes in Computer Science, vol 5208. Berlin, Heidelberg: Springer; 2008:253-261.
31. Kang SH, Gratch J, Wang N, Watt JH. Does the contingency of agents' nonverbal feedback affect users' social anxiety? 2008 Presented at: 7th international joint conference on Autonomous agents and multiagent systems; May 12-16, 2008; Estoril, Portugal.
32. McQuiggan SW, Lester JC. Modeling and evaluating empathy in embodied companion agents. International Journal of Human-Computer Studies 2007 Apr;65(4):348-360. [doi: 10.1016/j.ijhcs.2006.11.015]
33. Boukricha H, Wachsmuth I. Empathy-based emotional alignment for a virtual human: a three-step approach. Künstl Intell 2011 May 19;25(3):195-204. [doi: 10.1007/s13218-011-0109-8]
34. Boukricha H, Wachsmuth I, Hofstätter A, Grammer K. Pleasure-arousal-dominance driven facial expression simulation. 2009 Presented at: 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops; September 10-12, 2009; Amsterdam, Netherlands. [doi: 10.1109/acii.2009.5349579]
35. Boukricha H, Becker C, Wachsmuth I. Simulating empathy for the virtual human max. 2007 Presented at: 2nd Workshop on Emotion and Computing - Current Research and Future Impact; 2007; Osnabrück, Germany URL: https://pub.uni-bielefeld.de/record/2276595
36. Pelachaud C. Modelling multimodal expression of emotion in a virtual agent. Philos Trans R Soc Lond B Biol Sci 2009 Dec 12;364(1535):3539-3548 [FREE Full text] [doi: 10.1098/rstb.2009.0186] [Medline: 19884148]
37. Pelachaud C. Studies on gesture expressivity for a virtual agent. Speech Communication 2009 Jul;51(7):630-639 [FREE Full text] [doi: 10.1016/j.specom.2008.04.009]
38. Prendinger H, Ishizuka M. The empathic companion: a character-based interface that addresses users' affective states. Applied Artificial Intelligence 2005 Mar 09;19(3-4):267-285. [doi: 10.1080/08839510590910174]
39. Von Der Pütten A, Krämer NC, Gratch J. Who's there? Can a Virtual Agent Really Elicit Social Presence? 2009 Presented at: 12th Annual International Workshop on Presence; November 11-13, 2009; Los Angeles, CA.
40. von den Putten A, Reipen C, Wiedmann A, Kopp S, Krämer NC. The Impact of Different Embodied Agent-Feedback on Users´ Behavior. 2009 Presented at: 9th International Conference on Intelligent Virtual Agents; September 14-16, 2009; Amsterdam, Netherlands. [doi: 10.1007/978-3-642-04380-2_86]
41. Wang N, Gratch J. Rapport and facial expression. 2009 Presented at: 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops; September 10-12, 2009; Amsterdam, Netherlands. [doi: 10.1109/acii.2009.5349514]
42. Bickmore T, Giorgino T. Health dialog systems for patients and consumers. J Biomed Inform 2006 Oct;39(5):556-571 [FREE Full text] [doi: 10.1016/j.jbi.2005.12.004] [Medline: 16464643]

43. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005 Oct;59(1):21-30. [doi: 10.1016/j.pec.2004.09.008] [Medline: 16198215]
44. Cassell J, Bickmore T. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. User Model User-Adap Inter 2003;13:89-132. [doi: 10.1023/A:1024026532471]
45. Klüwer T. "I Like Your Shirt"-Dialogue Acts for Enabling Social Talk in Conversational Agents. 2011 Presented at: International Workshop on Intelligent Virtual Agents; September 15-17, 2011; Reykjavik, Iceland. [doi: 10.1007/978-3-642-23974-8_2]
46. Schulman D, Bickmore T, Sidner C. An Intelligent Conversational Agent for Promoting Long-Term Health Behavior Change Using Motivational Interviewing. 2011 Presented at: AAAI Spring Symposium Series; March 21-23, 2011; Palo Alto, CA.
47. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a relational agent by hospital patients with depressive symptoms. Interact Comput 2010 Jul 01;22(4):289-298 [FREE Full text] [doi: 10.1016/j.intcom.2009.12.001] [Medline: 20628581]
48. Johnson WL, LaBore C, Chui YC. A pedagogical agent for psychosocial intervention on a handheld computer. 2004 Presented at: AAAI Fall Symposium on Dialogue Systems for Health Communication; October 22-24, 2004; Arlington, VA.
49. Silverman B, Holmes J, Kimmel S, Branas C, Ivins D, Weaver R, et al. Modeling emotion and behavior in animated personas to facilitate human behavior change: the case of the HEART-SENSE game. Health Care Manag Sci 2001 Sep;4(3):213-228. [doi: 10.1023/a:1011448916375] [Medline: 11519847]
50. Friederichs S, Bolman C, Oenema A, Guyaux J, Lechner L. Motivational interviewing in a Web-based physical activity intervention with an avatar: randomized controlled trial. J Med Internet Res 2014;16(2):e48 [FREE Full text] [doi: 10.2196/jmir.2974] [Medline: 24550153]
51. Llorach G, Blat J. Say Hi to Eliza. An Embodied Conversational Agent on the Web. 2017 Presented at: 17th International Conference Intelligence Virtual Agents (IVA); August 27-30, 2017; Stockholm, Sweden. [doi: 10.1007/978-3-319-67401-8_34]
52. Ramanarayanan V, Tilsen S, Proctor M, Töger J, Goldstein L, Nayak K, et al. Analysis of speech production real-time MRI. Computer Speech & Language 2018 Nov;52:1-22 [FREE Full text] [doi: 10.1016/j.csl.2018.04.002]
53. Schroeder J, Wilkes C, Rowan K, Toledo A, Paradiso A, Czerwinski M, et al. Pocket Skills: A Conversational Mobile Web App To Support Dialectical Behavioral Therapy. 2018 Presented at: Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal, Quebec, Canada. [doi: 10.1145/3173574.3173972]
54. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
55. Davis FD, Bagozzi RP, Warshaw PR. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
56. Polceanu M, Lisetti C. Time to Go ONLINE! A Modular Framework for Building Internet-based Socially Interactive Agents. 2019 Presented at: 19th ACM International Conference on Intelligent Virtual Agents; July 2-5, 2019; Paris, France. [doi: 10.1145/3308532.3329452]
57. Hester RK, Squires DD, Delaney HD. The Drinker's Check-up: 12-month outcomes of a controlled clinical trial of a stand-alone software program for problem drinkers. J Subst Abuse Treat 2005 Mar;28(2):159-169. [doi: 10.1016/j.jsat.2004.12.002] [Medline: 15780546]
58. Kilian Wells G, Moonie S, Pharr J. Association between alcohol screening and brief intervention during routine check-ups and alcohol consumption among adults living in California. Arch Psychiatr Nurs 2018 Dec;32(6):872-877. [doi: 10.1016/j.apnu.2018.07.001] [Medline: 30454631]
59. Stringer KL, Baker EH. Stigma as a barrier to substance abuse treatment among those with unmet need: an analysis of parenthood and marital status. Journal of Family Issues 2015 Apr 24;39(1):3-27. [doi: 10.1177/0192513x15581659]
60. Quigley M, Gerkey B, Conley K, Faust J, Foote T, Leibs J, et al. ROS: an open-source Robot Operating System. 2009 Presented at: ICRA workshop on open source software; May 12-17, 2009; Kobe, Japan.
61. O'Brien HL, Toms EG. What is user engagement? A conceptual framework for defining user engagement with technology. J. Am. Soc. Inf. Sci 2008 Apr;59(6):938-955. [doi: 10.1002/asi.20801]
62. Heerink M, Kröse B, Evers V, Wielinga B. Assessing Acceptance of Assistive Social Agent Technology by Older Adults: the Almere Model. Int J of Soc Robotics 2010 Sep 4;2(4):361-375. [doi: 10.1007/s12369-010-0068-5]
63. Bartneck C, Kulić D, Croft E, Zoghbi S. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. Int J of Soc Robotics 2008 Nov 20;1(1):71-81. [doi: 10.1007/s12369-008-0001-3]
64. Lucas GM, Rizzo A, Gratch J, Scherer S, Stratou G, Boberg J, et al. Reporting Mental Health Symptoms: Breaking Down Barriers to Care with Virtual Human Interviewers. Front. Robot. AI 2017 Oct 12;4:53-79. [doi: 10.3389/frobt.2017.00051]
65. Lucas G, Gratch J, King A, Morency L. It's only a computer: Virtual humans increase willingness to disclose. Computers in Human Behavior 2014 Aug;37:94-100 [FREE Full text] [doi: 10.1016/j.chb.2014.04.043]

66. Boustani M, Lunn S, Polcneau M, Lisetti C. Increasing consumer engagement with computer-delivered behavioral interventions: Incorporating verbal reflections and non-verbal gestures to a digital health agent. 2019 Presented at: 53rd Annual Convention of the Association of Behavioral and Cognitive Therapies (ABCT); November 21-24, 2019; Atlanta, GA.
67. Buche J, Beck AJ, Singer PM. Factors impacting the development of a diverse behavioral health workforce. University of Michigan School of Public Health Behavioral Health Workforce Research Center. 2017. URL: https://mha.ohio.gov/Portals/0/assets/HealthProfessionals/Training%20and%20Workforce%20Development/CareerPathways/Factors_Impacting_%20Behavioral_Health_Workforce-Diversity.pdf?ver=2019-03-13-102735-210 [accessed 2021-08-30]
68. Duffy FF, West J, Wilk J, Narrow W, Hales D, Thompson J. Mental health practitioners and trainees. Medicine 2002:327-368.

Abbreviations

AI: artificial intelligence
AUD: alcohol use disorder
BMI: brief motivational interviewing
DCU: Drinker's Check-Up
DHI: digital health intervention
ECA: embodied conversational agent
eEVA: empathic embodied virtual agent
GUI: graphical user interface
MI: motivational interviewing
NVB: nonverbal behavior
RCT: randomized controlled trial
TAM: technology acceptance model

Edited by G Eysenbach; submitted 17.11.20; peer-reviewed by L Kramer, J Satterfield, C Emezue; comments to author 09.12.20; revised version received 26.05.21; accepted 29.05.21; published 29.09.21.

Please cite as:
Boustani M, Lunn S, Visser U, Lisetti C
Development, Feasibility, Acceptability, and Utility of an Expressive Speech-Enabled Digital Health Agent to Deliver Online, Brief Motivational Interviewing for Alcohol Misuse: Descriptive Study
J Med Internet Res 2021;23(9):e25837
URL: https://www.jmir.org/2021/9/e25837
doi:10.2196/25837
PMID:34586074

Original Paper

Unique Features of a Web-Based Nutrition Website for Childhood Cancer Populations: Descriptive Study

Lisa Wartenberg[1], RD, LD, MFA; Margaret Raber[2,3], MPH, DPhil; Joya Chandra[1,3], PhD

[1]Department of Pediatrics Research, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

[2]Department of Behavioral Science, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

[3]Center for Energy Balance in Cancer Prevention and Survivorship, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

Corresponding Author:
Joya Chandra, PhD
Department of Pediatrics Research
The University of Texas MD Anderson Cancer Center
1515 Holcombe Blvd
Unit 853
Houston, TX, 77030-4000
United States
Phone: 1 17135635405
Email: jchandra@mdanderson.org

Abstract

Background: Children with cancer experience a myriad of nutritional challenges that impact their nutrition status during treatment and into survivorship. Growing evidence suggests that weight at diagnosis impacts cancer outcomes, but provider guidance on nutrition and diet during treatment varies. Nutrition literacy and culinary resources may help mitigate some common nutritional problems; however, many patients may face barriers to accessing in-person classes. Along with dietitian-led clinical interventions, web-based resources such as the newly updated electronic cookbook (e-cookbook) created by The University of Texas MD Anderson Cancer Center, @TheTable, may facilitate access to nutrition and culinary education during treatment and into survivorship.

Objective: We sought to define and describe the features and content of the @TheTable e-cookbook and compare it with analogous resources for a lay audience of patients with childhood cancer and childhood cancer survivors as well as their families.

Methods: We evaluated freely available web-based resources via a popular online search engine (ie, Google). These searches yielded three web-based resources analogous to @TheTable: the American Institute for Cancer Research's Healthy Recipes, The Children's Hospital of San Antonio's Culinary Health Education for Families Recipe for Life, and Ann Ogden Gaffney and Fred Hutchinson Cancer Research Center's Cook for Your Life. These sites were analyzed for the following: number of recipes, search functionality, child or family focus, cancer focus, specific dietary guidance, videos or other media, and miscellaneous unique features.

Results: Cook for Your Life and Culinary Health Education for Families Recipe for Life were the most comparable to @TheTable with respect to cancer focus and family focus, respectively. Healthy Recipes is the least user-friendly, with few search options and no didactic videos.

Conclusions: The @TheTable e-cookbook is unique in its offering of child- and family-focused content centered on the cancer and survivorship experience.

(J Med Internet Res 2021;23(9):e24515) doi:10.2196/24515

KEYWORDS

pediatric oncology; web-based resources; oncology nutrition; culinary education; oncology; children; pediatric; nutrition; culinary; education

Introduction

An estimated 300,000 children aged 0-19 years are diagnosed with cancer each year in the United States [1]. Although the incidence of cancer in children and adolescents in the United States has risen slightly over the past few decades, the 5-year survival rate has improved from 58% to over 80% since the mid-1970s [2-4]. This significant improvement in survival rate has resulted in the need to mitigate long-term treatment-related sequelae in childhood cancer survivors (CCS), such as secondary cancers, obesity, and cardiovascular disease. Many of these late effects may be exacerbated by excess weight and poor diet, which highlights the importance of nutrition education across the cancer care continuum [5]. Despite these issues, adherence to dietary guidelines is low across the population, including among children without a history of cancer [6], CCS [7,8], and adult survivors of childhood cancer [9-11]. Specifically, the diet of CCS is lacking in fruits, vegetables, and calcium [11,12]. Resources that promote diet quality are scarce for patients with childhood cancer undergoing treatment and CCS as well as their families.

Evidence suggests that children with a BMI outside a healthy weight range at presentation (ie, <5th or >85th percentile [13,14]) have significantly poorer survival than those with a BMI within the healthy weight range [14-17]. Diet quality commonly worsens during treatment in pediatric patients [18] and into survivorship [8,19]; however, provider guidance on diet modification and nutrition adequacy is rarely offered and greatly varies. Recently diagnosed patients and their families have voiced the need to incorporate guidance on appropriate nutrition and lifestyle changes in cancer care [20]. Younger patients may be at particular risk of experiencing adverse effects due to a poor diet that was present and continued through treatment—as many dietary habits are established by the age of 3 years—or one that started during treatment [18,21].

While diet quality in pediatric patients and CCS warrants further research, practical and accessible dietary guidance is clearly needed for patients during treatment. Parents of CCS refer to digital nutrition- and cooking-related content to inform their food choices [22]. Digital information provided by authorized sources could support families of CCS by dispelling diet myths, curbing the spread of misinformation, and promoting nutrition literacy and culinary knowledge. Nutrition-focused web-based resources may help increase the frequency of homemade meals, a factor that has been positively associated with diet quality. Web-based resources focused on healthy eating are promising means to mitigate common treatment-related sequelae and late effects as well as barriers to accessing information on nutrition among CCS.

This paper describes nutrition-focused web-based resources targeted to patients with childhood cancer and CCS as well as their families. The University of Texas MD Anderson Cancer Center's (MDACC) electronic cookbook (e-cookbook), @TheTable, was developed in 2012 by a team of research dietitians, culinary specialists, and other staff [23]. Since its original dissemination, @TheTable has undergone significant design changes to improve its user-friendliness. The collection of recipes and nutrition-related content has grown substantially, and the educational nutrition information is now offered in three additional languages (Spanish, Mandarin Chinese, Arabic) that reflect well-represented populations within the institution. In this study, we compared the features and content of @TheTable with those of other freely available web-based resources and highlighted the unique aspects of this e-cookbook.

Methods

Search Strategy for the Identification of Web-Based Resources

We mirrored the methods typically used by patients, survivors, and families to acquire information and conducted multiple internet searches between 2019 and 2020 using combinations of keywords such as "tools," "cancer," "recipes," "healthy," and "eating." An MDACC research librarian independent of our research team additionally assisted in the search and found comparable results.

Resource Selection

Resources were included based on the following criteria: a substantial (>100 recipes) and searchable recipe catalogue, cost-free use of resource, and authorized source (ie, content generated or reviewed by a registered dietitian or affiliated with a major institution). From the web-based search, the two closest analogues to @TheTable were selected for in-depth analysis. The included sites were analyzed for the following: number of recipes, search functionality, child or family focus, cancer focus, specific dietary guidance, videos or other media, and miscellaneous unique features. To compensate for the uniqueness of the resources targeted to childhood cancer populations, the Culinary Health Education for Families (CHEF) Recipe for Life site created by The Children's Hospital of San Antonio was included as a web-based resource with a pediatric focus that the research team was familiar with.

Features and Content Extraction

Website features are shown in Table 1. Content extracted included number of recipes available on the site and ability to search by certain criteria (eg, texture, symptom, nutrition need, meal type). Filters and search functionality were included as key features of web-based resources, as these allow organization of larger libraries or nutrition content for browsing and also allow users with specific questions or needs to find information quickly, thus increasing usability.

Table 1. Key features of web-based cookbooks and resources for patients.

Site title	Author	Recipes, n	Search by symptom	Search by nutrition	Search by meal type	Child focus	Cancer focus	Diet tips/guides	Videos	Other features
@TheTable [24]	The University of Texas MD Anderson Cancer Center	>770	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Question submission
Healthy Recipes [25]	American Institute for Cancer Research	250-300	No	No	Yes	No	Yes	Yes	No	Physical activity guidelines
Culinary Health Education for Families Recipe for Life [26]	The Children's Hospital of San Antonio	>75	No	Yes	Yes	Yes	No	Some	Yes	Teaching kitchens; provider referrals to culinary programs; home activities; blog; community resources
Cook for Your Life [27]	Ann Ogden Gaffney; Fred Hutchinson Cancer Research Center	>500	No	No	Yes	No	Yes	Yes	Yes	Culturally adapted menus; menu collections

The presence or absence of child or family focus was also noted. A "kid-friendly" website was one that had a sensory approach to food (eg, inclusion of color, shape, texture) to facilitate accommodating child-specific preferences and aversions (related or unrelated to treatment); focused on child-specific nutritional needs related or unrelated to the cancer experience; and acknowledged the role and journey of the caretaker or parent in treatment and survivorship. Cancer focus was gauged according to whether recipes could accommodate treatment-related adverse effects; the presence or absence of symptom- or treatment-specific content or search functions largely defined this category. The availability of nutrition tips or guidelines, the presence of freely available didactic videos, and miscellaneous features unique to the web-based resource (ie, overlapping in-person classes, home activities, blogs or newsletters, cultural inclusivity) were assessed.

Results

Included Web-Based Resources

To better examine the MDACC @TheTable e-cookbook, selected web-based resources were analyzed. We identified the following: Healthy Recipes from the American Institute for Cancer Research; CHEF Recipe for Life from the Children's Hospital of San Antonio; and Cook for Your Life, created by Ann Ogden Gaffney and owned by the Fred Hutchinson Cancer Research Center.

@TheTable Resource

@TheTable is a web-based resource that may be accessed without cost or subscription. It includes over 770 recipes that are searchable by symptom, nutritional need, texture, and flavor, in addition to standard cookbook features (Table 1). Nutrition-related filters allow selection of nutritious recipes specific to certain diets such as vegan, vegetarian, and low-sodium diets (Figure 1). Figure 1A shows a screenshot of the landing page for the @TheTable e-cookbook. From this page, site visitors can access a myriad of freely available resources specifically designed for patients with cancer. The family focus of the e-cookbook aims at bolstering the nutritional habits of family units. The salient features of the e-cookbook are recipes searchable by symptom, meal type, texture, and color, as well as nutrition tip sheets and didactic videos. Most recipes are in English, and some are in Spanish. Figure 1B shows nutrition tip sheets available in this e-cookbook. Freely available tip sheets on @TheTable can additionally be accessed in Spanish, Mandarin Chinese, and Arabic. This augments accessibility, especially for bilingual households or those living outside the United States. Tip sheets cover topics such as the nutritional needs of growing girls and boys undergoing treatment for cancer, the management of symptoms such as constipation, and practical tips for the grocery store. Figure 1C shows the @TheTable nutrition guide for pediatric patients with leukemia. The nutrition booklet follows the pedagogy of MyPlate, the nutrition guide likely used by families in public schools in the United States, and builds on that knowledge. It aims to establish a common language to enhance future interactions with clinicians and dietitians caring for pediatric patients with cancer. This booklet centered around the five food groups is also available in Mandarin Chinese, Arabic, and Spanish. Symptom-related filters show recipes that accommodate common treatment-related adverse effects such as diarrhea, nausea, dysgeusia, and constipation (Figure 2). Site visitors can also use the color filter to search by phytonutrient (eg, blue/purple filter for anthocyanin-rich foods). These specific filters were designed to complement a trifold brochure distributed at weekly in-person MDACC culinary classes. The texture filter allows the user to find comfort measures for symptoms such as mouth sores or mucositis (eg, soft or smooth/creamy) or select sensory aspects (eg, crunchy, dry, chewy). Videos on @TheTable show prerecorded recipes filmed in an onsite research kitchen. These 19 instructional videos feature pediatric patients, chefs or culinary experts, and research dietitians. An additional feature of @TheTable is the nutritional guidance on various topics delivered via tip sheets and booklets (N=26; Figure 2). As Table 2 illustrates, these are also available

in Spanish (19/26, 73%), Mandarin Chinese (18/26, 69%), and Arabic (18/26, 69%) to increase usability.

Figure 1. Representative images of @TheTable features for patients with cancer and survivors.

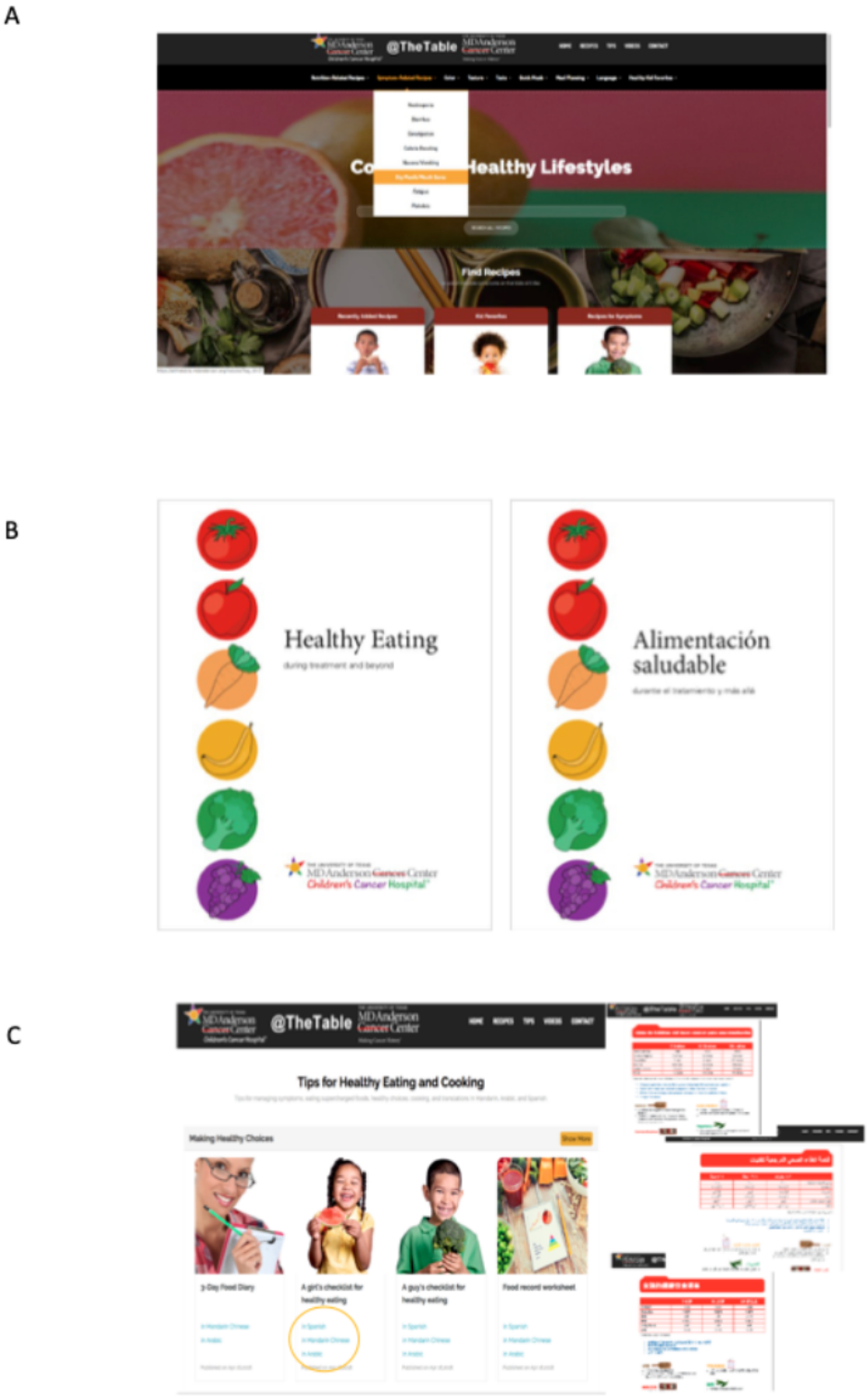


Figure 2. Schematic of additional @TheTable features highlighting its specialty focus and nutritional education, culinary, and multilingual content.

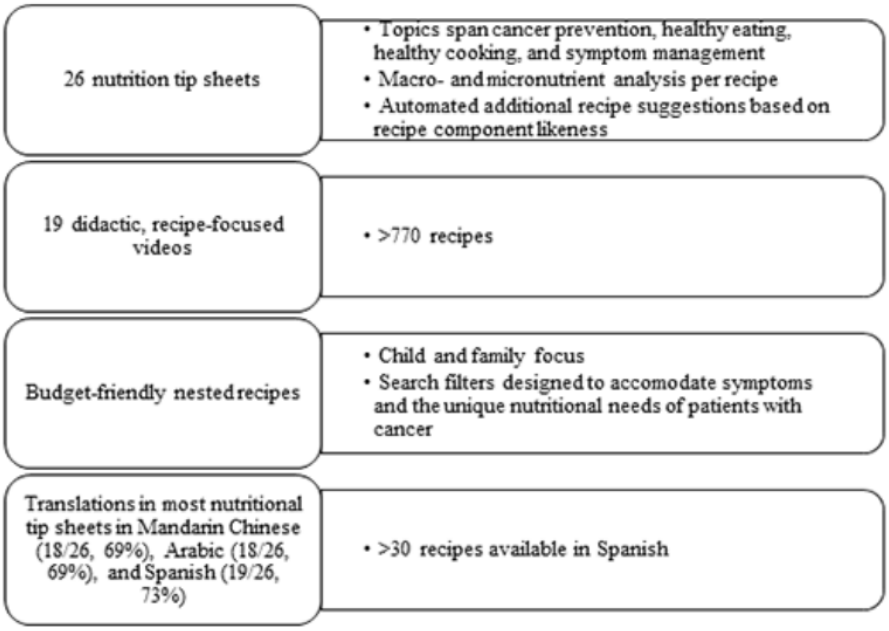


Table 2. Breakdown of nutrition education tip sheets available on @TheTable by broad topic.

Health topic	Example(s) of content	Tip sheet quantity	Available translations		
			Spanish	Mandarin Chinese	Arabic
Cancer prevention	• Add color to your diet! • Fallin' for pumpkin	4	3	3	3
Healthy eating	• Healthy eating during treatment and beyond booklet • What do I eat? • A girl's checklist for healthy eating	10	5	5	5
Healthy cooking	• Meal ideas • Prepping vegetables • Stock your spice cabinet	10	9	8	8
Symptom management	• Fiber and constipation	2	2	2	2

Healthy Recipes Resource

The Healthy Recipes site segments its content into the following two portals that branch off into more focused evidence-based guidelines: cancer prevention and cancer survival. Its recipe search tool is limited to ingredients or meal types; in this way, it does not explicitly focus on the cancer experience. Unlike @TheTable, this web-based resource does not have a child or family focus. Its nutrition-related content is centered around the New American Plate, a broadly used dietary model formulated under the guidance of the American Institute for Cancer Research's *Third Expert Report*, which translates cancer-specific, evidence-based nutrition recommendations into practical visual aids for the patient or consumer.

CHEF Recipe for Life Resource

All content on CHEF Recipe for Life is available in either English or Spanish. Recipes can be searched by difficulty level (ie, beginner, intermediate, advanced), meal type, and special diets (eg, dairy-free, low-fat, nut-free). Moreover, its didactic videos offer either recipe instructions or culinary skill tutorials. Users can sign up for newsletters, submit a form to ask a dietitian a specific question that is subsequently addressed in a site blog, or access a provider referral form to enroll in onsite evidence-based classes in the San Antonio area. However, the site focuses on management and prevention of chronic diseases, including type II diabetes and obesity. Its content is thus not cancer-specific, though this site had the strongest community focus of the included sites.

Cook for Your Life Resource

The web-based resource Cook for Your Life has a substantial recipe collection (>500 recipes) that can be searched by meal type, food preferences, or health considerations. Further, a "menus and collections" feature clusters recipes around commonalities such as "Vegan Kitchen Basics" and "Eat What You Can to Feel Better." This website stands out for its substantial video collection, which includes skill-building,

culinary technique, and recipe demonstrations. Its blog component delivers educational content related to specific food items or ingredients (eg, farro, avocado, curry powder) as well as nutrition guidelines (eg, adequate hydration, survivorship).

Comparison of Included Sources

All of the discussed web-based resources encourage plant-forward eating. Only @TheTable and Cook for Your Life include the option to filter recipes according to treatment-related symptoms. Like CHEF Recipe for Life, Cook for Your Life has all its content in either English or Spanish, whereas @TheTable has only selected content in other languages. Cook for Your Life and @TheTable both focus on cancer, though Cook for Your Life content is not based on the family or pediatric patient. CHEF Recipe for Life and @TheTable stand out among the included web-based resources for their focus on the family and pediatric patient. Notably, family-centered approaches to prevent obesity have proven to be most effective in pediatric nutritional interventions [28,29]. While @TheTable has cancer-specific content, CHEF Recipe for Life centers on chronic disease prevention and management. Further, the Healthy Kid Favorites search filter found in @TheTable, for instance, allows the user to search for recipes commonly enjoyed by children (eg, macaroni and cheese, pizza, nachos), which are altered to some extent to augment the nutritional profile. For example, the "Chicken Rainbow Nachos" recipe has fiber-packed vegetables with protein-rich chicken and black beans as nutritionally dense additions to the traditional ingredients. Moreover, the recipe is designed as part of a set of "nested recipes," which begin with a foundational "parent" recipe (eg, Tomato Poached Chicken Filling) that is used to create "child" recipes (eg, Chicken Rainbow Nachos). This approach stems from a community-based intervention with low-income Hispanic families. It allows for economizing and adds variety in home-prepared meals, though it may increase time spent in the kitchen. In addition, @TheTable has over 30 of approximately 770 recipes (~3.8% of the recipe collection) in Spanish, which allows broader user reach.

An additional aspect unique to @TheTable is the ability to search for recipes according to a color associated with a corresponding phytonutrient. This feature can be used independently, though it is meant to complement a physical trifold brochure given to patients and families at MDACC nutrition-focused events such as the onsite weekly cooking classes held in the inpatient unit at the Ronald McDonald House Kitchen. Users can search for orange or red foods, for example, which they learn are rich in carotenoids by reading the brochure. This offers the dual benefit of considering child food preferences as children may be particular about the color of a food or dish. Recipes from this e-cookbook are routinely used in MDACC onsite cooking classes delivered to patients with childhood cancer and CCS as well as their families [30,31].

Discussion

This paper describes the unique features of the redesigned @TheTable e-cookbook. We compared it to three analogous websites and found that it is unique in its focus on cancer in the pediatric context, as well as search functionality and recipe collection. Further, we found that Cook for Your Life and CHEF Recipe for Life are most similar to @TheTable with respect to cancer and family focus, respectively. Healthy Recipes is the least user-friendly because it has fewer search options than the other websites and no didactic videos.

Web-based resources with a nutrition focus are particularly relevant to pediatric oncology because of the poor diet quality commonly observed in both patients [32-36] and survivors [8,37]. Specialized nutrition resources addressing common treatment-related problems and late effects may help promote healthy eating patterns and improve patient and survivor outcomes [8,15,35,37-39]. While @TheTable and other web-based resources may play an important role in augmenting nutrition literacy and nutrition education in patients with cancer and survivors and their families, these do not supplant dietitian-led nutrition care. These resources are mainly descriptive, which limits conclusions regarding how these resources may be most effectively integrated into patient care.

Web-based resources, such as @TheTable, have the potential for broad utilization in both digital and in-person interventions. For example, @TheTable has been used across diverse settings such as in-person cooking classes at summer camps for CCS, the pediatric inpatient unit, childhood cancer awareness events, and several cooking- and nutrition-focused studies [22,30,31,40]. However, patients and families could face barriers to participating in in-person interventions. Beaulieu-Gagnon et al [41] identified common barriers faced by patients with childhood cancer in attending face-to-face nutrition classes. Notably, low attendance (17 participants over 1 year) resulted in the cancelation of 71% of their healthy cooking courses. Similarly, low attendance is a problem at the in-person weekly cooking classes in the pediatric inpatient unit at MDACC; plans to attend are disrupted for even the most engaged and interested dyads because of unforeseen medical appointments or because the child feels unwell.

A disproportionately higher incidence of treatment-related problems is seen in CCS with low-income or multiracial background [42]. Unfortunately, the need for culturally adapted nutrition- and diet-focused content remains largely unaddressed. Although it is merely a start, @TheTable has most of its educational nutrition content translated into three languages (ie, Spanish, Mandarin Chinese, Arabic), which helps broaden access to this information. Cultural adaptation is imperative for meeting the unique needs of culturally diverse demographic groups. Some innovative digital resources tackle the widespread issue of limited patient access to registered dietitian services [43]. Ina, developed by Savor Health, is a virtual dietitian directed by user input and artificial intelligence. This chat bot provides expert on-demand dietary counseling and recipes via SMS text messaging after patient data is provided. The tools described above highlight ways in which digital applications can potentially enhance patient and survivor care and in turn improve outcomes. Further research is needed to understand the impact of these novel technologies and digital resources.

Acknowledgments

We are grateful for support from the Gerber Foundation, the Children's Art Project, the 2011-2014 Santa's Elves Fundraiser sponsored by The University of Texas MD Anderson's Advance Team, David Herr & Family, the Farrah Fawcett Foundation, the James B. and Lois R. Archer Family Charitable Foundation, the Center for Energy Balance in Cancer Prevention and Survivorship, the Duncan Family Institute for Cancer Prevention and Risk Assessment, the National Institutes of Health, and the National Cancer Institute under award numbers P30CA016672 and R25CA057730. The authors would like to acknowledge the contributions and support provided, in part, by the Assessment, Intervention and Measurement (AIM) Shared Resource through a Cancer Center Support Grant (CA16672, PI: Peter Pisters, MD, MD Anderson Cancer Center) from the National Cancer Institute, National Institutes of Health.

The contents of this work are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health or the National Cancer Institute.

Conflicts of Interest

The authors created and maintained The University of Texas MD Anderson Cancer Center @TheTable electronic cookbook. They declare no other conflicts of interest.

References

1. United States of America: Cancer Country Profile 2020. World Health Organization. URL: https://www.who.int/cancer/country-profiles/USA_2020.pdf?ua=1 [accessed 2020-02-01]
2. Schadler KL, Kleinerman ES, Chandra J. Diet and exercise interventions for pediatric cancer patients during therapy: tipping the scales for better outcomes. Pediatr Res 2018 Jan 25;83(1-1):50-56 [FREE Full text] [doi: 10.1038/pr.2017.225] [Medline: 29068433]
3. Smith MA, Seibel NL, Altekruse SF, Ries LA, Melbert DL, O'Leary M, et al. Outcomes for Children and Adolescents With Cancer: Challenges for the Twenty-First Century. JCO 2010 May 20;28(15):2625-2634. [doi: 10.1200/jco.2009.27.0421]
4. American Cancer Society. URL: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2018/cancer-facts-and-figures-2018.pdf [accessed 2019-05-20]
5. Zhang FF, Parsons SK. Obesity in Childhood Cancer Survivors: Call for Early Weight Management. Adv Nutr 2015 Sep;6(5):611-619 [FREE Full text] [doi: 10.3945/an.115.008946] [Medline: 26374183]
6. Banfield EC, Liu Y, Davis JS, Chang S, Frazier-Wood AC. Poor Adherence to US Dietary Guidelines for Children and Adolescents in the National Health and Nutrition Examination Survey Population. J Acad Nutr Diet 2016 Jan;116(1):21-27 [FREE Full text] [doi: 10.1016/j.jand.2015.08.010] [Medline: 26391469]
7. Berdan CA, Tangney CC, Scala C, Stolley M. Childhood cancer survivors and adherence to the American Cancer Society Guidelines on Nutrition and Physical Activity. J Cancer Surviv 2014 Dec 5;8(4):671-679. [doi: 10.1007/s11764-014-0376-0] [Medline: 24996982]
8. Bérard S, Morel S, Teasdale E, Shivappa N, Hebert JR, Laverdière C, et al. Diet Quality Is Associated with Cardiometabolic Outcomes in Survivors of Childhood Leukemia. Nutrients 2020 Jul 18;12(7):2137 [FREE Full text] [doi: 10.3390/nu12072137] [Medline: 32708379]
9. Belle F, Wengenroth L, Weiss A, Sommer G, Beck Popovic M, Ansari M, Swiss Paediatric Oncology Group (SPOG). Low adherence to dietary recommendations in adult childhood cancer survivors. Clin Nutr 2017 Oct;36(5):1266-1274. [doi: 10.1016/j.clnu.2016.08.012] [Medline: 28277304]
10. Robien K, Ness K, Klesges L, Baker K, Gurney J. Poor adherence to dietary guidelines among adult survivors of childhood acute lymphoblastic leukemia. J Pediatr Hematol Oncol 2008 Nov;30(11):815-822 [FREE Full text] [doi: 10.1097/MPH.0b013e31817e4ad9] [Medline: 18989158]
11. Zhang F, Ojha R, Krull K, Gibson TM, Lu L, Lanctot J, et al. Adult Survivors of Childhood Cancer Have Poor Adherence to Dietary Guidelines. J Nutr 2016 Dec;146(12):2497-2505 [FREE Full text] [doi: 10.3945/jn.116.238261] [Medline: 27798341]
12. Stolley MR, Restrepo J, Sharp LK. Diet and physical activity in childhood cancer survivors: a review of the literature. Ann Behav Med 2010 Jun 18;39(3):232-249 [FREE Full text] [doi: 10.1007/s12160-010-9192-6] [Medline: 20559768]
13. Dehghan M, Akhtar-Danesh N, Merchant AT. Childhood obesity, prevalence and prevention. Nutr J 2005 Sep 02;4(1):24 [FREE Full text] [doi: 10.1186/1475-2891-4-24] [Medline: 16138930]
14. Himes J, Dietz W. Guidelines for overweight in adolescent preventive services: recommendations from an expert committee. The Expert Committee on Clinical Guidelines for Overweight in Adolescent Preventive Services. Am J Clin Nutr 1994 Feb;59(2):307-316. [doi: 10.1093/ajcn/59.2.307] [Medline: 8310979]
15. Barr RD, Gomez-Almaguer D, Jaime-Perez JC, Ruiz-Argüelles GJ. Importance of Nutrition in the Treatment of Leukemia in Children and Adolescents. Arch Med Res 2016 Nov;47(8):585-592. [doi: 10.1016/j.arcmed.2016.11.013] [Medline: 28476186]

16. Lange BJ, Gerbing RB, Feusner J, Skolnik J, Sacks N, Smith FO, et al. Mortality in overweight and underweight children with acute myeloid leukemia. JAMA 2005 Jan 12;293(2):203-211. [doi: 10.1001/jama.293.2.203] [Medline: 15644547]
17. Co-Reyes E, Li R, Huh W, Chandra J. Malnutrition and obesity in pediatric oncology patients: causes, consequences, and interventions. Pediatr Blood Cancer 2012 Dec 15;59(7):1160-1167 [FREE Full text] [doi: 10.1002/pbc.24272] [Medline: 22948929]
18. Goddard E, Cohen J, Bramley L, Wakefield C, Beck E. Dietary intake and diet quality in children receiving treatment for cancer. Nutr Rev 2019 May 01;77(5):267-277. [doi: 10.1093/nutrit/nuy069] [Medline: 30715472]
19. Zhang FF, Saltzman E, Kelly MJ, Liu S, Must A, Parsons SK, et al. Comparison of childhood cancer survivors' nutritional intake with US dietary guidelines. Pediatr Blood Cancer 2015 Aug 24;62(8):1461-1467 [FREE Full text] [doi: 10.1002/pbc.25521] [Medline: 25808589]
20. Moody K, Meyer M, Mancuso CA, Charlson M, Robbins L. Exploring concerns of children with cancer. Support Care Cancer 2006 Sep 26;14(9):960-966. [doi: 10.1007/s00520-006-0024-y] [Medline: 16639553]
21. Birch LL. Development of food acceptance patterns in the first years of life. Proc Nutr Soc 1998 Nov;57(4):617-624. [doi: 10.1079/pns19980090] [Medline: 10096125]
22. Raber M, Crawford K, Baranowski T, Sharma SV, Schick V, Markham C, et al. Exploring food preparation practices in families with and without school-aged childhood cancer survivors. Public Health Nutr 2019 Sep 20;23(3):410-415. [doi: 10.1017/s1368980019002672]
23. Li R, Raber M, Chandra J. Developing a healthy web-based cookbook for pediatric cancer patients and survivors: rationale and methods. JMIR Res Protoc 2015 Mar 31;4(1):e37 [FREE Full text] [doi: 10.2196/resprot.3777] [Medline: 25840596]
24. At the Table - Cooking Up Healthy Lifestyles. MD Anderson Cancer Center. URL: https://atthetable.mdanderson.org/ [accessed 2021-08-06]
25. AICR RECIPES - Fuel a Healthy Diet With Our Delicious and Healthy Recipes. American Institute for Cancer Research. URL: https://www.aicr.org/cancer-prevention/recipes/ [accessed 2021-08-06]
26. Building a Healthier Community. CHEF Recipe for Life. URL: https://www.chefsa.org/ [accessed 2021-08-06]
27. Healthy Cooking for Cancer Patients. Fred Hutch Cook for Your Life. URL: https://www.cookforyourlife.org/ [accessed 2021-08-06]
28. Smith JD, Berkel C, Jordan N, Atkins DC, Narayanan SS, Gallo C, et al. An individually tailored family-centered intervention for pediatric obesity in primary care: study protocol of a randomized type II hybrid effectiveness-implementation trial (Raising Healthy Children study). Implement Sci 2018 Jan 15;13(1):11 [FREE Full text] [doi: 10.1186/s13012-017-0697-2] [Medline: 29334983]
29. Sung-Chan P, Sung YW, Zhao X, Brownson RC. Family-based models for childhood-obesity intervention: a systematic review of randomized controlled trials. Obes Rev 2013 Apr 09;14(4):265-278. [doi: 10.1111/obr.12000] [Medline: 23136914]
30. Raber M, Crawford K, Chandra J. Healthy cooking classes at a children's cancer hospital and patient/survivor summer camps: initial reactions and feasibility. Public Health Nutr 2017 May 02;20(9):1650-1656. [doi: 10.1017/s136898001700060x]
31. Raber M, Huynh TN, Crawford K, Kim S, Chandra J. Development and Feasibility of a Community-Based, Culturally Flexible Colorectal Cancer Prevention Program. J Community Health 2018 Oct 12;43(5):882-885 [FREE Full text] [doi: 10.1007/s10900-018-0497-x] [Medline: 29532214]
32. Arpaci T, Toruner E, Altay N. Assessment of Nutritional Problems in Pediatric Patients with Cancer and the Information Needs of Their Parents: A Parental Perspective. Asia Pac J Oncol Nurs 2018;5(2):231-236 [FREE Full text] [doi: 10.4103/apjon.apjon_78_17] [Medline: 29607385]
33. Bauer J, Jürgens H, Frühwald MC. Important aspects of nutrition in children with cancer. Adv Nutr 2011 Mar;2(2):67-77 [FREE Full text] [doi: 10.3945/an.110.000141] [Medline: 22332035]
34. Ruland CM, Hamilton GA, Schjødt-Osmo B. The complexity of symptoms and problems experienced in children with cancer: a review of the literature. J Pain Symptom Manage 2009 Mar;37(3):403-418 [FREE Full text] [doi: 10.1016/j.jpainsymman.2008.03.009] [Medline: 18694633]
35. Barr RD, Stevens MC. The influence of nutrition on clinical outcomes in children with cancer. Pediatr Blood Cancer 2020 Jun 05;67 Suppl 3(S3):e28117. [doi: 10.1002/pbc.28117] [Medline: 32134218]
36. Łoboś P, Januszewicz A. Food neophobia in children. Pediatr Endocrinol Diabetes Metab 2019;25(3):150-154 [FREE Full text] [doi: 10.5114/pedm.2019.87711] [Medline: 31769273]
37. Barr RD, Ladas EJ. The role of nutrition in pediatric oncology. Expert Rev Anticancer Ther 2020 Feb 03;20(2):109-116. [doi: 10.1080/14737140.2020.1719834] [Medline: 31964179]
38. Ladas EJ, Sacks N, Brophy P, Rogers PC. Standards of nutritional care in pediatric oncology: results from a nationwide survey on the standards of practice in pediatric oncology. A Children's Oncology Group study. Pediatr Blood Cancer 2006 Mar;46(3):339-344. [doi: 10.1002/pbc.20435] [Medline: 15926168]
39. Rogers PC, Barr RD. The relevance of nutrition to pediatric oncology: A cancer control perspective. Pediatr Blood Cancer 2020 Jun 25;67 Suppl 3(S3):e28213. [doi: 10.1002/pbc.28213] [Medline: 32096351]
40. Li R, Donnella H, Knouse P, Raber M, Crawford K, Swartz MC, et al. A randomized nutrition counseling intervention in pediatric leukemia patients receiving steroids results in reduced caloric intake. Pediatr Blood Cancer 2017 Feb 12;64(2):374-380. [doi: 10.1002/pbc.26231] [Medline: 27615542]

41. Beaulieu-Gagnon S, Bélanger V, Meloche C, Curnier D, Sultan S, Laverdière C, et al. Nutrition education and cooking workshops for families of children with cancer: a feasibility study. BMC Nutr 2019 Nov 19;5(1):52 [FREE Full text] [doi: 10.1186/s40795-019-0319-2] [Medline: 32153965]
42. Bhatia S, Gibson TM, Ness KK, Liu Q, Oeffinger KC, Krull KR, et al. Childhood cancer survivorship research in minority populations: A position paper from the Childhood Cancer Survivor Study. Cancer 2016 Aug 01;122(15):2426-2439 [FREE Full text] [doi: 10.1002/cncr.30072] [Medline: 27253866]
43. Trujillo E, Claghorn K, Dixon S, Hill EB, Braun A, Lipinski E, et al. Inadequate Nutrition Coverage in Outpatient Cancer Centers: Results of a National Survey. J Oncol 2019;2019:7462940 [FREE Full text] [doi: 10.1155/2019/7462940] [Medline: 31885583]

Abbreviations

CCS: childhood cancer survivor
CHEF: Culinary Health Education for Families
MDACC: The University of Texas MD Anderson Cancer Center

Edited by R Kukafka; submitted 08.10.20; peer-reviewed by V Marcil, P Becker; comments to author 31.10.20; revised version received 12.11.20; accepted 04.03.21; published 13.09.21.

Please cite as:
Wartenberg L, Raber M, Chandra J
Unique Features of a Web-Based Nutrition Website for Childhood Cancer Populations: Descriptive Study
J Med Internet Res 2021;23(9):e24515
URL: https://www.jmir.org/2021/9/e24515
doi:10.2196/24515
PMID:34515643

Viewpoint

Closing the Virtual Gap in Health Care: A Series of Case Studies Illustrating the Impact of Embedding Evaluation Alongside System Initiatives

Laura Desveaux[1,2], PT, PhD; Suman Budhwani[1], PhD; Vess Stamenova[1], PhD; Onil Bhattacharyya[1], MD, PhD; James Shaw[1,2,3], PT, PhD; R Sacha Bhatia[1], MD, MBA

[1]Institute for Health System Solutions and Virtual Care, Women's College Hospital, Toronto, ON, Canada

[2]Institute of Health Policy, Management, and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[3]Joint Centre for Bioethics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Laura Desveaux, PT, PhD
Institute for Health System Solutions and Virtual Care
Women's College Hospital
76 Grenville St
Toronto, ON, M5S 1B2
Canada
Phone: 1 416 323 6400 ext 4302
Email: laura.desveaux@thp.ca

Abstract

Early decisions relating to the implementation of virtual care relied on necessity and clinical judgement, but there is a growing need for the generation of evidence to inform policy and practice designs. The need for stronger partnerships between researchers and decision-makers is well recognized, but how these partnerships can be structured and how research can be embedded alongside existing virtual care initiatives remain unclear. We present a series of case studies that illustrate how embedded research can inform policy decisions related to the implementation of virtual care, where decisions are either to (1) discontinue (red light), (2) redesign (yellow light), or (3) scale up existing initiatives (green light). Data were collected through document review and informal interviews with key study personnel. Case 1 involved an evaluation of a mobile diabetes platform that demonstrated a mismatch between the setting and the technology (decision outcome: discontinue). Case 2 involved an evaluation of a mental health support platform that suggested evidence-based modifications to the delivery model (decision outcome: redesign). Case 3 involved an evaluation of video visits that generated evidence to inform the ideal model of implementation at scale (decision outcome: scale up). In this paper, we highlight the characteristics of the partnership and the process that enabled success and use the cases to illustrate how these characteristics were operationalized. Structured communication included monthly check-ins and iterative report development. We also outline key characteristics of the partnership (ie, trust and shared purpose) and the process (ie, timeliness, tailored reporting, and adaptability) that drove the uptake of evidence in decision-making. Across each case, the evaluation was designed to address policy questions articulated by our partners. Furthermore, structured communication provided opportunities for knowledge mobilization. Structured communication was operationalized through monthly meetings as well as the delivery of interim and final reports. These case studies demonstrate the importance of partnering with health system decision-makers to generate and mobilize scientific evidence. Embedded research partnerships founded on a shared purpose of system service provided an effective strategy to bridge the oft-cited gap between science and policy. Structured communication provided a mechanism for collaborative problem-solving and real-time feedback, and it helped contextualize emerging insights.

(J Med Internet Res 2021;23(9):e25797) doi:10.2196/25797

KEYWORDS

virtual care; primary care; embedded research; implementation; knowledge exchange; health policy

Background

Technologies that enable virtual care are a central focus of health care transformation efforts worldwide [1-4], yet their uptake in practice fell considerably short of their potential prior to the COVID-19 pandemic [5-8]. The necessary reorganization of care in response to the pandemic has introduced a plethora of large-scale virtual initiatives [9,10], overcoming many of the oft-cited policy barriers to introduction [11,12]. Nevertheless, the production of scientific knowledge lags behind the proliferation of virtual solutions and focuses largely on efficacy [13-15] and experience [16-18]. Such evidence is critical but does not provide actionable insight into how to best operationalize virtual care, including patients and situations for which it is most appropriate. As a result, system administrators and policy makers are left to make decisions based on incomplete evidence and anecdotal experience.

The COVID-19 pandemic has highlighted the challenges of making decisions in the absence of evidence [19], driven largely by the speed and severity of its spread [20]. Although early decisions during the pandemic relied on necessity and clinical judgement, there is a growing recognition of the need to apply scientific principles with greater vigilance and flexibility than we have traditionally applied [20]. Simply put, we need to generate contextually relevant scientific research informed by policy needs and timelines to support decision-making. There is an increasing understanding of the underlying social and relational processes that drive the use of evidence [21], underscoring the need to evolve beyond the traditional model of *knowledge translation* to focus on *knowledge mobilization*. In this sense, knowledge mobilization refers to the "the reciprocal and complementary flow and uptake of research between researchers, knowledge brokers and knowledge users" [22]. The key distinction is that the presence or production of knowledge alone is insufficient to drive change. Instead, personal contact and interaction between knowledge producers (ie, researchers) and knowledge users (ie, decision-makers) are required to ensure evidence-based transformation [23].

The need for stronger partnerships between researchers and decision-makers is well recognized [24]; however, we were unable to find examples of successful partnerships relating to virtual care in the literature. This paper outlines three novel case examples from an embedded research partnership that adopted the model of *engaged scholarship*—a demand-driven approach that seeks partnership with health system decision-makers by focusing specifically on their needs and the context for decision-making [25]. In a demand-driven approach, key political actors ask for evidence to be made available and are, in turn, receptive to the findings [26]. We present these three cases to illustrate how the core components of such engaged scholarship (ie, prioritizing the needs of decision-makers and responsiveness to decision-maker context) were operationalized in order to embed evaluation alongside system initiatives in order to understand whether or not their aims are being achieved. We describe several underlying characteristics that enabled success to support the tactical realization of engaged scholarship and its replication; we also identify directions to build on these characteristics in the future. Although understanding how and why policy decisions are made is beyond the scope of this paper, we highlight how and why the evaluations were constructed in a way to support this decision-making process.

Operationalizing a Model of Engaged Scholarship in Virtual Care

Barrett et al [27] suggest that acquiring relevant knowledge and skills to produce effective research for decision-making involves researchers spending time in a decision-making environment. Our model of partnership approximates this and builds on recommended knowledge mobilization principles of cross-sector and interagency learning [21] by creating an organizational partnership founded on a mutual interest of driving digital transformation within the health system. It represents a "collaborative form of inquiry in which academics and [clinicians] leverage their different perspectives and competencies to coproduce knowledge about a complex problem" [28].

In 2016, the Women's College Hospital Institute for Health System Solutions and Virtual Care (WIHV) partnered with the Ontario Telemedicine Network (OTN)—a nonprofit, government-funded organization and the largest provider of publicly funded telemedicine services in the province of Ontario, Canada. This partnership was initially founded as a product of funding stipulations set by Canada Health Infoway (CHI) that required an independent evaluation alongside a series of virtual care initiatives implemented by the OTN and funded by the CHI. Although the CHI had a direct relationship with the OTN, they were not involved in, nor did they influence the evaluation. These initial evaluations (including cases 1 and 2 described below) created a platform to invest the necessary time, effort, and resources in developing the partnership and establishing a proof of concept that would serve as a foundation for subsequent evaluations (eg, case 3, among others).

The series of case studies described below represent various system-level initiatives independently led by the OTN, who was responsible for selecting and implementing virtual care technologies through the funding received from the CHI. As policy decisions often relate to the question of whether the system should (1) discontinue (red light), (2) redesign (yellow light), or (3) scale up (green light) a given program, we have selected case studies that are illustrative of how evaluation can support each of these situations. Each case involves the evaluation of a patient-facing technology aiming to enable some aspect of health care delivery (eg, interactions with health care providers, remote monitoring, and self-management).

The OTN's overarching interest in each evaluation was to understand the impact of the technology in question. With this in mind, a team of individuals from each organization engaged in early discussions to clarify specific research questions, methods, and the conditions required for successful execution of the project (eg, access to third-party data from a project partner). The WIHV then prepared evaluation proposals for review and approval by the OTN and identified the interdisciplinary expertise required to execute the evaluation (ie, researchers with expertise in quasi-experimental methods;

researchers with expertise in applied qualitative methods; clinician scientists with relevant content expertise; and advisors with expertise in health system governance, regulation, and policy). Following approval from the OTN, the evaluations were conducted independently by the WIHV according to timelines set by the OTN. For each project, interim and final reports were provided to partners to inform decision-making (see Table 1 for project timelines). Initial versions of the reports were submitted to partners for their comments and consideration. The report was then sent back to our team for revisions, including addressing outstanding questions (where possible within the limits of the data), and to provide clarifications where the report was unclear, or the tacit knowledge of the research team was not made explicit. Since this process could continue ad infinitum, the partnership included an agreed upon limit of two rounds of revisions for all project reports. Monthly check-ins also provided an opportunity for emerging findings to inform ongoing planning and decision-making frequently, and in real time.

Table 1. Project timelines.

	Project launch	Interim report	Final report	Total duration
Case 1	July 2016	August 2017	January 2018	18 months
Case 2	July 2016	August 2017	January 2018	18 months
Case 3	September 2017	April 2018	April 2019	19 months

Case 1: A Mobile Self-management Platform for Type 2 Diabetes (Red Light)

The OTN piloted WellDoc BlueStar as an adjunct to the standard diabetes self-management education provided through provincial diabetes education centers [29,30]. BlueStar provides a web-based mobile coaching app for patients with type 2 diabetes mellitus, with tailored messaging based on user-inputted clinical data and the option to share data with patients' health care providers (eg, physicians, nurses).

The WIHV independently conducted a multicenter, pragmatic, randomized, wait-list controlled trial with blinded outcome assessment, inclusive of a qualitative realist evaluation [29]. This study design was selected to ensure we could answer the OTN's key question, "What is the impact of this technology?" while also accounting for the evaluation team's question of *whether* the technology had an impact compared to usual care and, if so, *who* it appeared to have an impact for. Results showed no difference in blood glucose levels, self-care behaviors, general health status, and self-reported health care utilization between the control and intervention groups [31]. Usage of the mobile app varied considerably by site, suggesting that contextual factors play a central role in achieving impact [31]. A secondary analysis of trial data revealed a potential dose-response relationship. The embedded realist evaluation indicated that patient characteristics were associated with positive outcomes, suggesting that individuals with moderate self-efficacy, no competing priorities, evidence of previous behavior change, and beliefs about the value of technology to support health may be more likely to engage and realize benefits [32].

Findings from this evaluation underscored the importance of implementing the technology for the right patients in the right settings. These insights were first brought to the attention of the OTN in a preliminary report, whereby timely delivery of trusted information enabled internal conversations about future directions. The OTN decided to revisit the clinical model for supporting type 2 diabetes mellitus and, therefore, did not proceed with procurement. Success in this partnership was realized by the informed decision to avoid a significant investment, thereby conserving financial resources for investment in areas with the promise of higher return.

Case 2: A Web-Based Self-management Platform for Mental Health (Yellow Light)

To enhance mental health capacity in Ontario, the OTN piloted the Big White Wall (BWW, now known TogetherAll [33]) as a virtual strategy to support the self-management of mental health for individuals requiring specialized mental health care [34,35]. At the time of the study, the BWW was a subscription-based web intervention that offered users anonymous access to peer support and self-directed learning courses [36,37].

The WIHV independently conducted a multicenter, parallel-arm, pragmatic randomized controlled trial evaluating the effectiveness of the BWW pilot implementation within the Ontario context [34,35,38]. This study design was selected to ensure we could answer the OTN's key question "What is the impact of this technology?" while also accounting for the evaluation team's question of *whether* the technology had an impact compared to usual care and *how much* engagement was required to achieve impact. Participants were recruited from outpatient mental health and addictions programs at three hospital sites in Ontario, and they were randomized to receive immediate or delayed access to BWW for a 3-month period (ie, the primary trial) [34,35]. Those who received immediate access and expressed a desire to continue use were randomized to an additional 3-month extension of use of the BWW, for a total use of 6 months, reflecting the default licensing model offered by the company or a control group comprising those who discontinued use during the extension trial [38].

Small but statistically significant benefits in mental health recovery, anxiety, and depression symptoms were observed among participants who received immediate access [34]. Engagement was variable, with approximately 80% of total logins being accounted for by 20% of all participants. A secondary analysis suggested a dose-response relationship; however, most participants did not engage in the platform in an

ongoing way [34]. This suggested differences in how this intervention might be utilized by participants outside of the trial, and how such use might eventually benefit them. This trend was further evaluated through the extension trial, with 51% (119/233) of the participants from the primary trial indicating an interest in extending their access to the BWW [38]. These participants had significantly higher anxiety levels at baseline, providing evidence of the clinical characteristics of patients who are more likely to derive benefit. Of those who received extended access, only 38% (21/55) engaged during the 3-month extension period, with no significant changes observed [38]. Findings from this study directly informed procurement conversations and policy decisions. Despite the company's default subscription model of 6 months, evaluation findings suggested that patient benefit plateaued at 3 months. Furthermore, most participants who were given access logged on only once or not at all, suggesting that a shared risk model whereby payment is triggered after a second login would be more cost-effective. This credible evidence provided decision-makers with the insights and confidence needed to modify their procurement approach, under the conditions outlined above. The nature of the partnership ensured that findings were delivered in a timely manner, with preliminary findings presented at monthly meetings as they emerged, and a final report delivered as the pilot wrapped up. These characteristics of the partnership allowed the OTN to consider emerging evidence as part of their strategic planning, leading to a province-wide roll-out of the BWW in October 2018 [39].

Case 3: A Platform to Support Video Visits and Asynchronous Messaging in Primary Care (Green Light)

The Enhanced Access to Primary Care (EAPC) proof-of-concept pilot was conducted by the OTN across five primary care sites in the province of Ontario as a model to improve primary care access. It introduced two technology platforms through which virtual visits could be delivered, and a renumeration structure through which primary care providers could be reimbursed.

An embedded cohort study using a mixed methods approach was utilized, including patient and provider interviews, patient and provider surveys, as well as data on technology use from the vendor. This study design was selected to ensure we could support the OTN in understanding "How do we scale this model," while also accounting for the evaluation team's question of *which* components of the model should be scaled (and which should not) and *why*. Findings indicated interest from patients [40,41] and providers [42] in conducting primary care visits through virtual modalities, with secure messaging being utilized for 94% of visits [41]. Providers were generally satisfied with the renumeration structure, although it was unclear whether they knew that it was slightly less than the remuneration offered for in-person visits. Lastly, about 81% of all virtual visits were delivered asynchronously, and they did not require additional follow-ups, suggesting that virtual visits could substitute for in-person primary care visits [41] and that they did not seem to create additional work.

Findings from this evaluation demonstrated the value of virtual care, and secure messaging in particular, in improving access to primary care and the patient experience. Of particular interest to our policy partners was the tailored messaging that helped answer the questions "Does this create additional work?" and "What is the appropriate compensate model?" Our findings highlighted the level of acceptance among providers relating to both the workload and pilot renumeration structure. Based on these findings, a province-wide expansion of EAPC is underway [43], and the Ontario Ministry of Health and Long-Term Care has committed CA $3 million (US $ 2.4 million) in new funding to compensate physicians for these visits [44].

Discussion

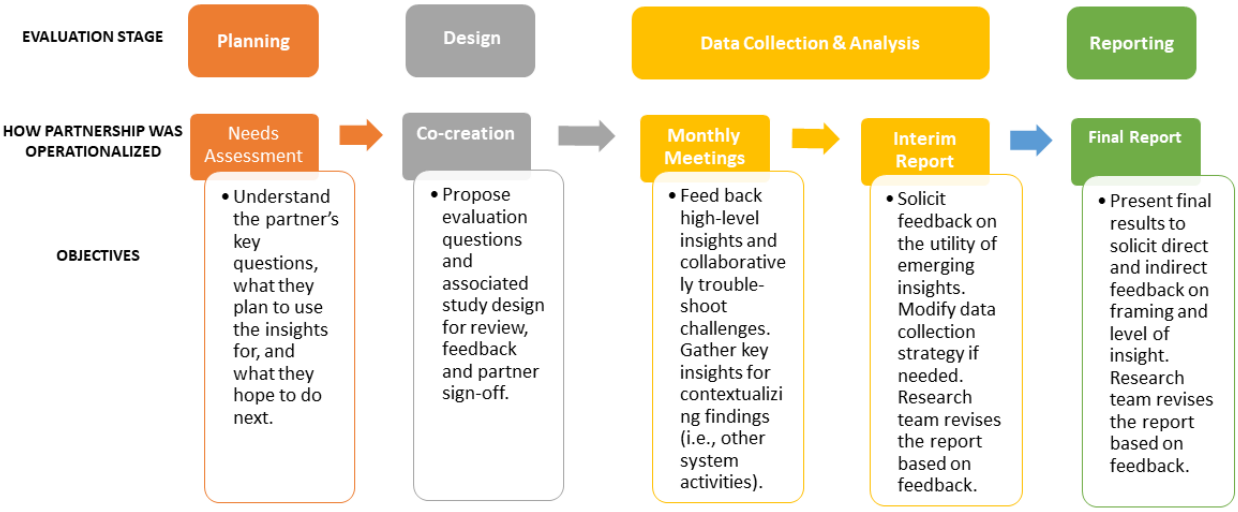
Prior work has characterized the involvement of decision-makers according to different stages of the research process [45], but explicit identification of the relationship-based and practical factors that underpin successful partnerships [46] and how they are realized is poorly described. Furthermore, researchers often cite challenges understanding project context, whereas stakeholders are concerned that results will be delivered well after the agreed deadline [46]. We build upon previous work by identifying the explicit mechanisms through which this partnership avoided the common pitfalls whereby stakeholders do not receive the information or evidence they require [46]. Recognizing that ongoing engagement is key [46], we have explicitly outlined the characteristics of the partnership (Figure 1) and how and why these characteristics were operationalized. This embedded research partnership provides a model to appropriately scale up virtual care initiatives [7] by establishing effective partnerships between policy makers and researchers and enhancing the accessibility of evidence.

Figure 1. Characteristics of embedded scholarship.

The success of this partnership was driven by its underlying characteristics (*the why*) as well as the characteristics of the process (*the how*; see Figure 1). This partnership is founded on a *shared purpose*, which in this instance meant the partnership existed to advance the interests of the health care system and the public it serves. This shared purpose was a function of the OTN's accountabilities to both the funder and the system and the research team's interest in using their expertise for applied impact. Stemming from this, scientific activities were therefore undertaken with the primary goal of providing applied, objective evidence for the health system decision-maker. Although scientific rigor is central to the core values [47], short timelines, limited funding, and the need for local relevance may trump methodological concerns. In this way, this partnership model builds on the existing knowledge base by blending aspects of integrated knowledge translation (ie, an explicit focus on increasing knowledge use and impact) [48] with engaged scholarship (ie, the research process gathers the perspectives of key stakeholders) [25]. Further, this work explicitly describes the tactical elements of the partnership that enabled the core components of engaged scholarship. Built on a fundamentally interdisciplinary approach [47], health system partners are systematically engaged in the discovery, development, and mobilization of scientific knowledge generated through this model [49,50]. In this regard, the process is adaptable but generally aligned with the knowledge-to-action process [51], whereby the tactical elements (ie, regular check-ins with partners) provide an opportunity to solicit feedback with regard to the knowledge inquiry, synthesis, and eventual products, providing a mechanism to inform adaptations to the evaluation to ensure it meets the policy needs. It is important to note that adaptations are made in a way that does not threaten scientific integrity. For example, in case 1, the initial focus of the evaluation was to understand the impact of the technology. While discussing preliminary results, it became clear that the current clinical model was suboptimal, and the evaluation needed to shift to identify the circumstances under which the technology might work and for which types of patients.

Our experience highlights that it is not just the above fundamental characteristics that will ensure success, but how these characteristics were operationalized that will support replication and scale of this model (Figure 2). First, the partnership is driven by *trust* and a *shared purpose*. Trust is established on the strength of the scientific methods, the credibility of the team, and the commitment to delivering insights that meet the needs of decision-makers, and over time, the degree to which the partnership produces such insights. The partnership terms required that evaluation activities be independently conducted to maintain integrity, credibility, and objectivity. This builds trust by ensuring that the resulting evidence is unbiased, rigorous, and respectful of patient, provider, and policy perspectives. Establishing shared purpose requires effective communication and a commitment on behalf of the researchers to describe findings in the context of immediate key challenges faced by system partners [27]. These characteristics inform the processes underlying the partnership, including *timeliness*, *tailored reporting*, and *adaptability*. Across our cases, these characteristics were embodied through an iterative knowledge mobilization strategy. Structured communication activities operationalized timeliness and adaptability in a deliberate attempt to avoid the production of findings that are not available when decisions need to be made or are not aligned to the practical requirements of policy [52,53]. Similarly, providing these empirical insights to decision-makers allowed them to reflect on their assumptions (ie, there is a fit between the problem and the technology) and pivot their questions accordingly to fill gaps in their understanding. Operationalizing the characteristics of trust, shared purpose, timeliness, tailored reporting, and adaptability were central in demonstrating value, maintaining relevance, and achieving mutual benefit [47].

Figure 2. How and why characteristics of the partnership were operationalized.



Embedded research provides the opportunity for the observation and identification of the processes playing a role in the adoption of technology and/or the (in)ability to achieve successful implementation. For example, the findings from the EAPC study [41] not only informed policy but also provided key insights relevant to the provincial response to the COVID-19 pandemic and facilitated the ability to recommend billing for virtual care quickly and confidently, given the evidence on file. Thus, the embedded partnership provides a mechanism to facilitate evidence translation [54] that occurs at the intersection of the knowledge-to-action process (or the transition from the inner to the outer circle) through the production of evidence that is immediately relevant to health system operations. In parallel, the partnership helps stakeholders build credibility and profile through their contribution to evidence generation and operational proof-of-concept [55].

Although this paper demonstrates the value of embedded research alongside policy initiatives and how it can be structured to address policy questions, it was beyond the scope of this work to evaluate the process through which the resulting evidence informed decision-making. Insights into this process would support the design of policy-oriented research as well as provide illustrative examples to decision-makers on how to embed evidence into decision-making. In addition, although these evaluations supported the uptake of evidence-based solutions at scale, no further evaluations were conducted. A longitudinal evaluation of the trajectory of implementation and impact would provide useful insight into the long-term value of this model and the sustainability of its impact.

Conclusions

Involving decision-makers in the formulation of proposals [56] and engaging them throughout the conduct of research [57] increases the likelihood that the resulting evidence will be used. Engaging them throughout the research process also increases the likelihood that resulting evidence is implemented in a sustainable way that can be valuable to policy. Embedded research partnerships founded on a shared purpose of system service provided an effective strategy to bridge the oft-cited gap between science and policy [58-60]. Trust, timeliness, tailored reporting, and adaptability were found to be key characteristics contributing to success that can be operationalized through co-creating evaluation questions, monthly check-ins, and iterative report development when scaling this model to new environments.

Acknowledgments

We would like to thank the Ontario Telemedicine Network for their commitment to embedded evaluation and their investment in a productive partnership. The case studies referred to in this manuscript were funded by Canada Health Infoway through investment projects with the Ontario Telemedicine Network. The publication costs for this article have been covered by Canada Health Infoway Inc., a non-profit corporation funded by the Government of Canada. The funders had no role in study design, data collection, and analysis, or decision to publish the manuscript.

Authors' Contributions

LD, VS, OB, JS, and RSB participated in the original studies. LD, OB, and RSB conceived the current manuscript. SB summarized the case studies. All authors helped interpret the data. LD drafted the manuscript, and all authors critically reviewed it. All authors read and approved the final manuscript

Conflicts of Interest

None declared.

References

1. Notice: Health Canada's Approach to Digital Health Technologies. Government of Canada. 2018. URL: https://www.canada.ca/en/health-canada/services/drugs-health-products/medical-devices/activities/announcements/notice-digital-health-technologies.html [accessed 2020-01-21]
2. Delivering 21st Century IT Support for the NHS - National Strategic Programme. United Kingdom Department of Health. URL: http://www.pacsgroup.org.uk/forum/messages/195/21st_National_Strategic_Programme-412.pdf [accessed 2020-01-21]
3. WHO guideline: recommendations on digital interventions for health system strengthening. World Health Organization. 2019. URL: https://www.who.int/reproductivehealth/publications/digital-interventions-health-system-strengthening/en/ [accessed 2020-08-26]
4. Walsh M, Chipperfield A. Australian National Telehealth Think Tank. J Telemed Telecare 2000;6(6):353. [doi: 10.1258/1357633001935978] [Medline: 11265107]
5. de Bont A, Bal R. Telemedicine in interdisciplinary work practices: on an IT system that met the criteria for success set out by its sponsors, yet failed to become part of every-day clinical routines. BMC Med Inform Decis Mak 2008 Oct 27;8:47 [FREE Full text] [doi: 10.1186/1472-6947-8-47] [Medline: 18954428]
6. Kuipers P, Humphreys JS, Wakerman J, Wells R, Jones J, Entwistle P. Collaborative review of pilot projects to inform policy: A methodological remedy for pilotitis? Aust N Z Health Policy 2008 Jul 19;5(1):1-7. [doi: 10.1186/1743-8462-5-17]
7. Sundin P, Callan J, Mehta K. Why do entrepreneurial mHealth ventures in the developing world fail to scale? J Med Eng Technol 2016;40(7-8):444-457. [doi: 10.1080/03091902.2016.1213901] [Medline: 27686003]
8. Wyatt JC, Sullivan F. eHealth and the future: promise or peril? BMJ 2005 Dec 10;331(7529):1391-1393 [FREE Full text] [doi: 10.1136/bmj.331.7529.1391] [Medline: 16339252]
9. Leite H, Hodgkinson IR, Gruber T. New development: 'Healing at a distance'—telemedicine and COVID-19. Public Money & Management 2020 Apr 09;40(6):483-485. [doi: 10.1080/09540962.2020.1748855]
10. Khairat S, Meng C, Xu Y, Edson B, Gianforcaro R. Interpreting COVID-19 and virtual care trends: cohort study. JMIR Public Health Surveill 2020 Apr 15;6(2):e18811 [FREE Full text] [doi: 10.2196/18811] [Medline: 32252023]
11. Desveaux L, Soobiah C, Bhatia RS, Shaw J. Identifying and overcoming policy-level barriers to the implementation of digital health innovation: qualitative study. J Med Internet Res 2019 Dec 20;21(12):e14994 [FREE Full text] [doi: 10.2196/14994] [Medline: 31859679]
12. Vimarlund V, Le Rouge C. Barriers and opportunities to the widespread adoption of telemedicine: a bi-country evaluation. Stud Health Technol Inform 2013;192:933. [Medline: 23920707]
13. Greenwood DA, Gee PM, Fatkin KJ, Peeples M. A systematic review of reviews evaluating technology-enabled diabetes self-management education and support. J Diabetes Sci Technol 2017 Sep;11(5):1015-1027 [FREE Full text] [doi: 10.1177/1932296817713506] [Medline: 28560898]
14. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual Research Review: Digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
15. Vegesna A, Tran M, Angelaccio M, Arcona S. Remote patient monitoring via non-invasive digital technologies: a systematic review. Telemed J E Health 2017 Jan;23(1):3-17 [FREE Full text] [doi: 10.1089/tmj.2016.0051] [Medline: 27116181]
16. Donelan K, Barreto EA, Sossong S, Michael C, Estrada JJ, Cohen AB, et al. Patient and clinician experiences with telehealth for patient follow-up care. Am J Manag Care 2019 Jan;25(1):40-44 [FREE Full text] [Medline: 30667610]
17. Jones G, Brennan V, Jacques R, Wood H, Dixon S, Radley S. Evaluating the impact of a 'virtual clinic' on patient experience, personal and provider costs of care in urinary incontinence: A randomised controlled trial. PLoS One 2018;13(1):e0189174 [FREE Full text] [doi: 10.1371/journal.pone.0189174] [Medline: 29346378]
18. Mammen JR, Elson MJ, Java JJ, Beck CA, Beran DB, Biglan KM, et al. Patient and physician perceptions of virtual visits for Parkinson's disease: a qualitative study. Telemed J E Health 2018 Apr;24(4):255-267. [doi: 10.1089/tmj.2017.0119] [Medline: 28787250]
19. Pacheco-Barrios K, Fregni F. Evidence-based decision making during COVID-19 pandemic. Princ Pract Clin Res 2020;6(1):1-2 [FREE Full text] [doi: 10.21801/ppcrj.2020.61.1] [Medline: 32766450]
20. Carley S, Horner D, Body R, Mackway-Jones K. Evidence-based medicine and COVID-19: what to believe and when to change. Emerg Med J 2020 Sep;37(9):572-575. [doi: 10.1136/emermed-2020-210098] [Medline: 32651176]
21. Davies HT, Powell AE, Nutley SM. Mobilising knowledge to improve UK health care: learning from other countries and other sectors – a multimethod mapping study. Southampton (UK): NIHR Journals Library 2015 Jun. [Medline: 26110190]
22. Social Sciences and Humanities Research Council – Definitions of Terms. Government of Canada. URL: https://www.sshrc-crsh.gc.ca/funding-financement/programs-programmes/definitions-eng.aspx#km-mc [accessed 2020-08-26]
23. Ward V, House A, Hamer S. Knowledge brokering: the missing link in the evidence to action chain? Evidence & Policy: A Journal of Research, Debate and Practice 2009 Aug 01;5(3):267-279. [doi: 10.1332/174426409x463811]
24. Bannister J, Hardill I. Knowledge mobilisation and the social sciences: dancing with new partners in an age of austerity. Contemporary Social Science 2013 Nov;8(3):167-175. [doi: 10.1080/21582041.2013.770910]
25. Van de Ven AH. Engaged scholarship: A guide for organizational and social research. Oxford, England: Oxford University Press on Demand; 2007.

26. Witter S, Kardan A, Scott M, Moore L, Shaxson L. Generating demand for and use of evaluation evidence in government health ministries: lessons from a pilot programme in Uganda and Zambia. Health Res Policy Syst 2017 Oct 02;15(1):86 [FREE Full text] [doi: 10.1186/s12961-017-0250-4] [Medline: 28969713]
27. Barratt H, Shaw J, Simpson L, Bhatia S, Fulop N. Health services research: building capacity to meet the needs of the health care system. J Health Serv Res Policy 2017 Oct;22(4):243-249 [FREE Full text] [doi: 10.1177/1355819617714815] [Medline: 28786700]
28. Boyer EL. The Scholarship of Engagement. Bulletin of the American Academy of Arts and Sciences 1996 Apr;49(7):18. [doi: 10.2307/3824459]
29. Desveaux L, Agarwal P, Shaw J, Hensel JM, Mukerji G, Onabajo N, et al. A randomized wait-list control trial to evaluate the impact of a mobile application to improve self-management of individuals with type 2 diabetes: a study protocol. BMC Med Inform Decis Mak 2016 Nov 15;16(1):144-111 [FREE Full text] [doi: 10.1186/s12911-016-0381-5] [Medline: 27842539]
30. Diabetes Education Program. Government of Ontario. 2019. URL: https://www.ontario.ca/page/diabetes-education-program [accessed 2020-08-26]
31. Agarwal P, Mukerji G, Desveaux L, Ivers NM, Bhattacharyya O, Hensel JM, et al. Mobile app for improved self-management of type 2 diabetes: multicenter pragmatic randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 10;7(1):e10321 [FREE Full text] [doi: 10.2196/10321] [Medline: 30632972]
32. Desveaux L, Shaw J, Saragosa M, Soobiah C, Marani H, Hensel J, et al. A mobile app to improve self-management of individuals with type 2 diabetes: qualitative realist evaluation. J Med Internet Res 2018 Mar 16;20(3):e81 [FREE Full text] [doi: 10.2196/jmir.8712] [Medline: 29549070]
33. Big White Wall relaunches as Togetherall. Canadian Health Technology. 2020. URL: https://www.canhealth.com/2020/08/12/big-white-wall-relaunches-as-togetherall/ [accessed 2020-08-26]
34. Hensel JM, Shaw J, Ivers NM, Desveaux L, Vigod SN, Cohen A, et al. A web-based mental health platform for individuals seeking specialized mental health care services: multicenter pragmatic randomized controlled trial. J Med Internet Res 2019 Jun 04;21(6):e10838 [FREE Full text] [doi: 10.2196/10838] [Medline: 31165710]
35. Hensel JM, Shaw J, Jeffs L, Ivers NM, Desveaux L, Cohen A, et al. A pragmatic randomized control trial and realist evaluation on the implementation and effectiveness of an internet application to support self-management among individuals seeking specialized mental health care: a study protocol. BMC Psychiatry 2016 Oct 18;16(1):350 [FREE Full text] [doi: 10.1186/s12888-016-1057-5] [Medline: 27756281]
36. Harding C, Chung H. Behavioral health support and online peer communities: international experiences. Mhealth 2016;2:43 [FREE Full text] [doi: 10.21037/mhealth.2016.10.04] [Medline: 28293613]
37. Our Services. Big White Wall. URL: https://www.bigwhitewall.co.uk/our-services [accessed 2020-01-22]
38. Hensel JM, Shaw J, Ivers NM, Desveaux L, Vigod SN, Bouck Z, et al. Extending access to a web-based mental health intervention: who wants more, what happens to use over time, and is it helpful? Results of a concealed, randomized controlled extension study. BMC Psychiatry 2019 Jan 24;19(1):39 [FREE Full text] [doi: 10.1186/s12888-019-2030-x] [Medline: 30678676]
39. Patton J, Ree M. Ontario goverenment-funded program offers mental health support 24/7. Global News. 2018 Oct 02. URL: https://globalnews.ca/news/4509381/ontario-government-big-white-wall-mental-health/ [accessed 2020-01-21]
40. Kelley L, Phung M, Stamenova V, Fujioka J, Agarwal P, Onabajo N, et al. Exploring how virtual primary care visits affect patient burden of treatment. Int J Med Inform 2020 Sep;141:104228. [doi: 10.1016/j.ijmedinf.2020.104228] [Medline: 32683311]
41. Stamenova V, Agarwal P, Kelley L, Fujioka J, Nguyen M, Phung M, et al. Uptake and patient and provider communication modality preferences of virtual visits in primary care: a retrospective cohort study in Canada. BMJ Open 2020 Jul 06;10(7):e037064 [FREE Full text] [doi: 10.1136/bmjopen-2020-037064] [Medline: 32636284]
42. Fujioka J, Nguyen M, Phung M. Re-designing primary care using virtual care: provider perspectives on the clinical utility of virtual visits. Canadian Family Physician (forthcoming) 2021:1-7 (forthcoming).
43. Canada Health Infoway. Expanding Access to Virtual Care in Ontario. Canada Health Infoway. 2020. URL: https://infoway-inforoute.ca/en/solutions/rapid-response-to-covid-19/5266-solutions/rapid-response-to-covid-19/8537-expanding-access-to-virtual-care-services-in-ontario [accessed 2020-08-11]
44. Government of Ontario. News Release: Government Expanding Digital and Virtual Health Care. Government of Ontario. 2019. URL: https://news.ontario.ca/mohltc/en/2019/11/ontario-expanding-digital-and-virtual-health-care.html [accessed 2020-01-22]
45. Ross S, Lavis J, Rodriguez C, Woodside J, Denis J. Partnership experiences: Involving decision-makers in the research process. J Health Serv Res Policy 2016 Dec 04;8(2_suppl):26-34. [doi: 10.1258/135581903322405144]
46. Williamson A, Tait H, El Jardali F, Wolfenden L, Thackway S, Stewart J, et al. How are evidence generation partnerships between researchers and policy-makers enacted in practice? A qualitative interview study. Health Res Policy Sys 2019 Apr 15;17(1):1-11. [doi: 10.1186/s12961-019-0441-2]
47. Beaulieu M, Breton M, Brousselle A. Conceptualizing 20 years of engaged scholarship: A scoping review. PLoS One 2018;13(2):e0193201 [FREE Full text] [doi: 10.1371/journal.pone.0193201] [Medline: 29489870]

48. Graham I, Tetroe J. How to translate health research knowledge into effective healthcare action. Healthc Q 2007;10(3):20-22 [FREE Full text] [doi: 10.12927/hcq..18919] [Medline: 17632905]
49. DeLugan R, Roussos S, Skram G. Linking academic and community guidelines for community-engaged scholarship. Journal of Higher Education Outreach and Engagement 2014;18(1):155-168 [FREE Full text]
50. Barker D. The scholarship of engagement: A taxonomy of five emerging practices. Journal of Higher Education Outreach and Engagement 2004;9(2):1-15 [FREE Full text]
51. Graham I, Logan J, Harrison M, Straus SE, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? J Contin Educ Health Prof 2006;26(1):13-24. [doi: 10.1002/chp.47] [Medline: 16557505]
52. Oliver K, Lorenc T, Innvær S. New directions in evidence-based policy research: a critical analysis of the literature. Health Res Policy Syst 2014 Jul 14;12:34 [FREE Full text] [doi: 10.1186/1478-4505-12-34] [Medline: 25023520]
53. World Bank Group impact evaluations : relevance and effectiveness. Washington, DC: World Bank Group; 2013. URL: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/206961468154467890/world-bank-group-impact-evaluations-relevance-and-effectiveness [accessed 2020-01-21]
54. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. J R Soc Med 2011 Dec;104(12):510-520 [FREE Full text] [doi: 10.1258/jrsm.2011.110180] [Medline: 22179294]
55. Video visits are now everyday health care. Ontario Telemedicine Network. 2019. URL: https://otn.ca/blog/video-visits-everyday-health-care/ [accessed 2020-08-12]
56. Kok MO, Gyapong JO, Wolffers I, Ofori-Adjei D, Ruitenberg J. Which health research gets used and why? An empirical analysis of 30 cases. Health Res Policy Syst 2016 May 17;14(1):36-11 [FREE Full text] [doi: 10.1186/s12961-016-0107-2] [Medline: 27188305]
57. Beyer JM, Trice HM. The utilization process: a conceptual framework and synthesis of empirical findings. Administrative Science Quarterly 1982 Dec;27(4):591. [doi: 10.2307/2392533]
58. Contandriopoulos D, Brousselle A. Reliable in their failure: an analysis of healthcare reform policies in public systems. Health Policy 2010 May;95(2-3):144-152. [doi: 10.1016/j.healthpol.2009.11.018] [Medline: 20022132]
59. Browman G, Snider A, Ellis P. Negotiating for change. The healthcare manager as catalyst for evidence-based practice: changing the healthcare environment and sharing experience. Healthc Pap 2003 Jan 15;3(3):10-22. [doi: 10.12927/hcpap..17125] [Medline: 12811083]
60. Haynes A, Rowbotham SJ, Redman S, Brennan S, Williamson A, Moore G. What can we learn from interventions that aim to increase policy-makers' capacity to use research? A realist scoping review. Health Res Policy Syst 2018 Apr 10;16(1):31 [FREE Full text] [doi: 10.1186/s12961-018-0277-1] [Medline: 29631606]

Abbreviations

BWW: Big White Wall
CHI: Canada Health Infoway
EAPC: Enhanced Access to Primary Care
OTN: Ontario Telemedicine Network
WIHV: Women's College Hospital Institute for Health System Solutions and Virtual Care

Edited by G Eysenbach, S Hagens; submitted 16.11.20; peer-reviewed by K Ho, S Lin, A Weinberg; comments to author 01.03.21; revised version received 20.05.21; accepted 29.07.21; published 03.09.21.

Please cite as:
Desveaux L, Budhwani S, Stamenova V, Bhattacharyya O, Shaw J, Bhatia RS
Closing the Virtual Gap in Health Care: A Series of Case Studies Illustrating the Impact of Embedding Evaluation Alongside System Initiatives
J Med Internet Res 2021;23(9):e25797
URL: https://www.jmir.org/2021/9/e25797
doi:10.2196/25797
PMID:34477560

Original Paper

Mechanism Design of Health Care Blockchain System Token Economy: Development Study Based on Simulated Real-World Scenarios

Se Young Jung[1*], MD, MPH; Taehyun Kim[2*], MSc; Hyung Ju Hwang[2], PhD; Kyungpyo Hong[1], MSc

[1]Office of eHealth Research and Business, Seoul National University Bundang Hospital, Seongnam-si, Republic of Korea

[2]Department of Mathematics, Pohang University of Science and Technology, Pohang-si, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Hyung Ju Hwang, PhD
Department of Mathematics
Pohang University of Science and Technology
77, Cheongam-ro, Nam-gu
Pohang-si, 37673
Republic of Korea
Phone: 82 054 279 2056
Email: hjhwang@postech.ac.kr

Abstract

Background: Despite the fact that the adoption rate of electronic health records has increased dramatically among high-income nations, it is still difficult to properly disseminate personal health records. Token economy, through blockchain smart contracts, can better distribute personal health records by providing incentives to patients. However, there have been very few studies regarding the particular factors that should be considered when designing incentive mechanisms in blockchain.

Objective: The aim of this paper is to provide 2 new mathematical models of token economy in real-world scenarios on health care blockchain platforms.

Methods: First, roles were set for the health care blockchain platform and its token flow. Second, 2 scenarios were introduced: collecting life-log data for an incentive program at a life insurance company to motivate customers to exercise more and recruiting participants for clinical trials of anticancer drugs. In our 2 scenarios, we assumed that there were 3 stakeholders: participants, data recipients (companies), and data providers (health care organizations). We also assumed that the incentives are initially paid out to participants by data recipients, who are focused on minimizing economic and time costs by adapting mechanism design. This concept can be seen as a part of game theory, since the willingness-to-pay of data recipients is important in maintaining the blockchain token economy. In both scenarios, the recruiting company can change the expected recruitment time and number of participants. Suppose a company considers the recruitment time to be more important than the number of participants and rewards. In that case, the company can increase the time weight and adjust cost. When the reward parameter is fixed, the corresponding expected recruitment time can be obtained. Among the reward and time pairs, the pair that minimizes the company's cost was chosen. Finally, the optimized results were compared with the simulations and analyzed accordingly.

Results: To minimize the company's costs, reward–time pairs were first collected. It was observed that the expected recruitment time decreased as rewards grew, while the rewards decreased as time cost grew. Therefore, the cost was represented by a convex curve, which made it possible to obtain a minimum—an optimal point—for both scenarios. Through sensitivity analysis, we observed that, as the time weight increased, the optimized reward increased, while the optimized time decreased. Moreover, as the number of participants increased, the optimization reward and time also increased.

Conclusions: In this study, we were able to model the incentive mechanism of blockchain based on a mechanism design that recruits participants through a health care blockchain platform. This study presents a basic approach to incentive modeling in personal health records, demonstrating how health care organizations and funding companies can motivate one another to join the platform.

(J Med Internet Res 2021;23(9):e26802) doi:10.2196/26802

KEYWORDS

mechanism design; optimization; blockchain; token economy; eHealth; electronic health records; healthcare; economy; health records

Introduction

Precision medicine aims to define diseases at a higher resolution using genomic data, electronic health records, and life-log data by providing new therapies to each targeted subgroup [1,2]. Electronic health records and life-log data from personal health records are crucial to capture phenotypic information in hospitals and in everyday life in order to deliver precision medicine to health care consumers. This is one of the major reasons why even high-income nations have been struggling over the past couple decades to properly disseminate electronic health records. For instance, the United States implemented the Health Information Technology for Economic and Clinical Health Act in 2009 to provide electronic health records throughout the country [3]. By 2017, 96% of general medical and surgical hospitals, 87% of children's hospitals, and 59% of acute long-term care hospitals in the United States had adopted certified electronic health records [4,5]. South Korea has also tried to disseminate electronic health records since early 2000. The Health Insurance Review and Assessment Service announced that 93.6% of hospitals and 91.6% of private clinics in South Korea utilized electronic health records as of 2017 [6]. Even though the adoption rate of electronic health records has increased dramatically throughout many countries, they are still struggling to find solutions to properly disseminating personal health records. Compared to the implementation and use of electronic health records in health care organizations, the implementation of personal health records, for the collection of life-log data through patient participation, still lags due to challenges related to security, privacy, interoperability, and data quality [7-9]. The implementation of personal health records faces other issues as well, including a lack of auditability, legal risk, health care policies, and data accuracy.

As such, health care blockchain has been implemented to solve these various problems. Blockchain is known as a distributed ledger technology—it records given information into small chunks of data sets called *blocks*, and if recorded data are valid, the blocks are, in turn, chained with a consensus protocol. The data are stored in a peer-to-peer system-based distributed storage environment, which does not permit anyone to arbitrarily modify it, because anyone can determine when a data change occurs [10]. Therefore, applying blockchain technology to personal health records strengthens the integrity and security of the clinical data stored within [11]. In addition, it is expected that automation features, such as smart contracts, will reduce both cost and time in managing patient participation through the dynamic consensus system [12,13]. Such an appropriately designed blockchain token economy can help in devising strategies to find the benefits of participating in clinical data sharing and to ensure their fair distribution among multiple stakeholders [14]. Furthermore, blockchain can improve the auditability of transferred health care records with secure privacy, help authenticate participants in health information exchange networks with distributed identification, and boost patient participation in the platform by providing incentives according to active participation and adherence to the system [15].

For the 2 main advantages of blockchain—namely auditability and identification—many proof-of-concept studies have been conducted [11,16-18], yet few studies have explored the incentive mechanism of blockchain with respect to real-world cases, in which incentives are crucial for recruitment and attention [19].

To examine the effectiveness of incentivization, one study [20] designed a token economy to encourage adherence to activities of daily living—bathing, physical activity, and oral hygiene—to reduce the risk of bloodstream infections, oral complications, and deconditioning in hematopoietic stem cell transplant for pediatric patients. Activities of daily living adherence rate increased from 0.51 to 2.5 after implementing the token economy [20]. In other studies [21-23], the preferred incentive methods were free parking, modest financial compensation, food coupons, guitar lessons, transportation, and donations to charity. Though £100 (approximately US $137.51) as an incentive was effective when recruiting older adult patients in a clinical trial [24], an amount of US $2.00 as an incentive was not effective [25]. This indicates that the amount of monetary compensation is important. Incentivizing also proved to be effective for patient adherence to smoking cessation, diet, and digital therapeutic programs [26-29]. Some questions have arisen—How about giving incentives multiple times? How much should we compensate for participants' effort? How long can we collect patients' information on the blockchain platform? Is it possible that the shorter the collecting time, the bigger the incentives become? In digital therapeutics, one main strategy in behavior modification is to make a contingency plan to address poor adherence, and monetary reward seems to be the most effective way for active engagement [30].

Mathematical studies have been performed to model incentivization for blockchain. For decentralization, designing protocols that make it difficult to tamper with transactions is essential, and a mathematical approach to designing blockchain protocol that achieves secure information exchange has been examined [31,32]. In addition, the analysis of players' mining strategies was conducted on blockchain using game theory [32,33]. Classical mechanism design theories have been examined on some apps, such as Auction [34]. However, there has been no research linking apps in the blockchain system to mechanism design, which is considered to be representative token design theory for blockchain [35-38].

Why is a new theoretical foundation needed for health care blockchain cryptoeconomics? The purpose of cryptoeconomics is to create internet services, but why do we need new theories and methodologies? In short, it is because the design of decentralized networks is completely different from that of traditional business/service planning. Existing business planning is given certain rules (market conditions) and aims to ensure

that companies make the best choices to maximize profits; however, the design of a decentralized network aims for the opposite. Assuming that each entity acts strategically and selfishly, the design of a decentralized network makes rules to achieve desired results.

Despite the pressing need for further research on the cryptoeconomics of health care blockchain, there has been a lack of research on token economy design to induce patient engagement using health care blockchain automated incentive programs. Thus, we aimed to explore parameters to consider when designing incentive programs that can be embedded in the smart contract of blockchain by using experiments based on 2 representative virtual scenarios with mechanism design.

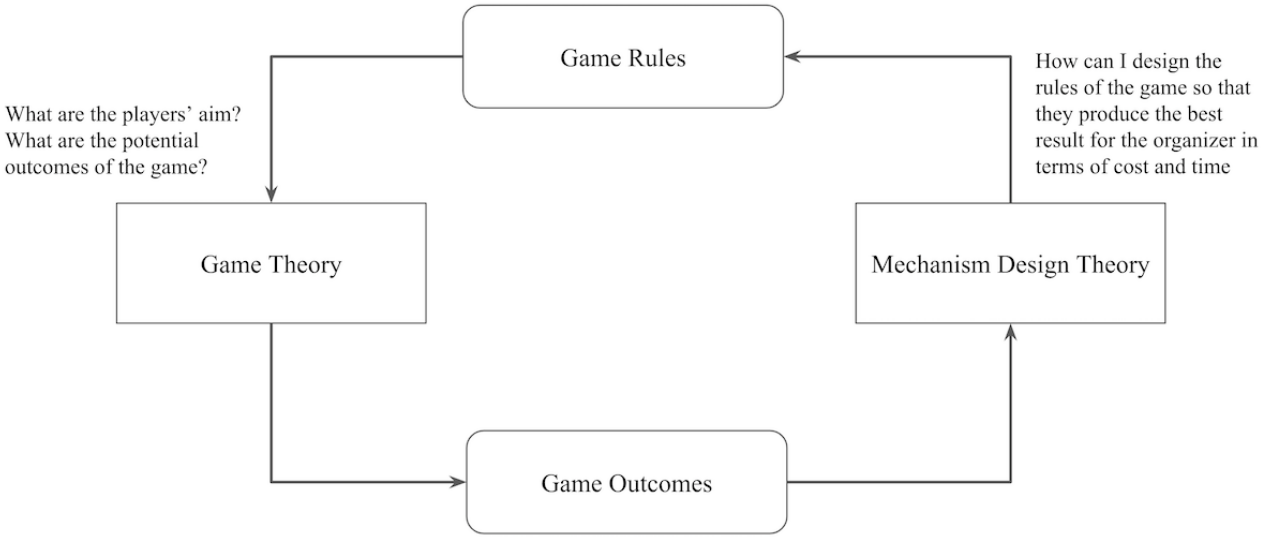
Methods

Methodological Background

Blockchain-based networks have no central principal. Therefore, we need a system that uses tokens (as the medium) and market principles to help individuals grow the network even if they act in pursuit of their own interests. This is called the *Token Model*, which serves as the invisible hand in decentralized networks [39].

To design the token economy, we considered 2 well-known economic theories: game theory and mechanism design. In game theory, existing business planning is given rules (market conditions) and aims to ensure that companies make the best choices to maximize their profits. The theory examines the best strategy in a given game. It explains how to draw conclusions in situations in which several rational decision-makers compete to win over one another. It is called a *game* because it is reminiscent of an actual game with competitors using strategies to win [40]. The relationship between game theory and mechanism design is as described in Figure 1. The organizer of recruitment is able to set the game rules in order to achieve intended goals in terms of cost and time in our scenarios (Figure 1 [41]).

Figure 1. Relationship between game theory and mechanism design theory.



However, the design of a decentralized network is the reverse. Mechanism design uses an engineering approach in which designers act rationally toward desired goals in a strategic environment by applying economic mechanisms and incentives to design strategies. It is also called *reverse game theory* because it starts at the end of the game and moves backward [34].

With the principle of mechanism design, we can reduce cost caused by trial and error and provide a platform for solving real-life problems as a theoretical model. Combined with blockchain, mechanism design has been adopted in various types of research. The auction is a representative example of mechanism design in a blockchain environment with randomness and information disclosure [36]. Based on this possibility, a previous study [35] provided the framework that analyzes the blockchain protocol using mechanism design and game theory. These studies [35,36] suggested that mechanism design can be applied to the blockchain system, especially to the health care blockchain economy, thereby resulting in the development of a basic token economy. We wanted to make ground truth models for the token economy of patient participation and data provision in personal health records. First, we assumed a basic scenario with essential variables—recruiting healthy participants for a vitality program that aims to reduce weight. The vitality program is a technology-based wellness program that is included with most life insurance policies, to support and reward healthy habits [42,43]. Second, we validated model robustness with another hypothetical scenario for validation—recruiting participants for clinical trials of anticancer drugs.

Assumption of Roles on a Blockchain Platform

When recruiting participants to a vitality program in the real world, a life insurance company provides incentives to those involved in the program. This mechanism is easily applied to the blockchain using smart contracts. In addition, the blockchain protects personal health information and is able to conduct universal recruitment through the app; therefore, efficient and safe recruitment of participants can be carried out through health care blockchain.

Meanwhile, the compensation that companies provide to participants on the blockchain is cryptocurrency. Therefore, it is essential to discuss how to calculate the value of this cryptocurrency, which can be performed in many ways. The value of bitcoin is determined by the free market. Suppose a vitality program participant is recruited on a blockchain; the life insurance company must have a certain amount of cryptocurrency. A participant who received the cryptocurrency must be able to exchange it for their benefit. In a conventional blockchain, operators obtain cryptocurrency in exchange for maintaining a blockchain system. However, unless the life insurance company participates in the blockchain operation, it must purchase the cryptocurrency from another party to secure a certain amount of it. In addition, assuming that the value of this cryptocurrency changes, there could be some concern that the total amount of money in the blockchain system would become negative. To solve these problems, we define a concept called *a currency exchange*, which assumes that a certain amount of cryptocurrency can be purchased or exchanged in a certain amount of fiat currency. The advantage of this method is that the total amount of cryptocurrency in the entire blockchain system, and the fiat value of cryptocurrency does not fall below zero, which helps the health care blockchain operate in a stable manner.

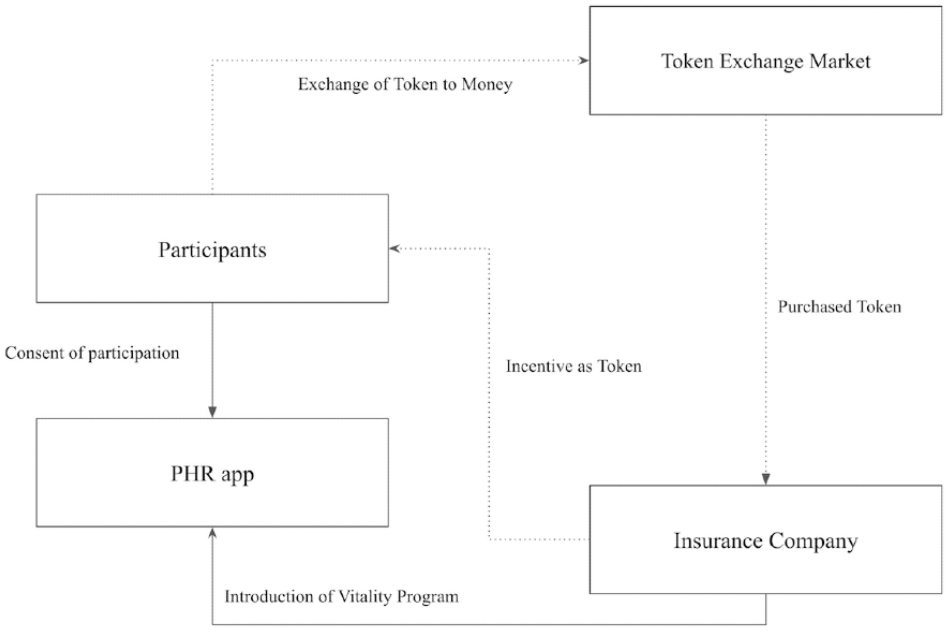
Several members exist in a blockchain system with a currency exchange. First, there is a blockchain operation party that operates the blockchain and acts as a currency exchange. In this study, it is assumed that the blockchain operating party does not affect the token economy because the party only generates cryptocurrency, and the value of cryptocurrency is fixed via the exchange. Second, there is a data provider that has patients' health information. Members who store personal health information include entities such as hospitals or genetic companies. They receive cryptocurrency from members who request information when there is a transaction of the information they store. The reason for receiving the cryptocurrency is the cost of storing the information. The third party is a user on the medical information platform. Users can be patients requesting their genomic information, healthy people uploading their life-log data, or life insurance companies wanting to recruit participants. In this study, we assume a life insurance company prepares a vitality program from the perspective of mechanism design by recruiting participants and finding out how long the recruitment period will be, depending on the amount of the incentive.

Scenarios

Scenario for the Development of the Model

Figure 2 shows a brief process of the basic scenario. The insurance company uploads the following program information through mobile phone apps regarding the collection of life-log data: (1) the institutions that conducts a vitality program; (2) duration of program operation; and (3) token rewards for participating in the program. Based on this information, participants will decide whether to take part in the program. If they decide to participate, they send requests through the app. The vitality program organizer will track participation status until a certain number of participants are obtained. The number of participants obtained here is different from the number of *N* desired by the organizer because the organizer needs to select participants based on their demographic information so that the sample is as unbiased as possible. If participants with certain socioeconomic statuses are enrolled more, the result of the intervention and analysis on effects of the program can be biased. The number *N* is set to greater than the necessary number of participants for the program because some enrollees could be dropped after adjustment for socioeconomic status. The vitality program organizer will pay participants who agree after screening tokens through the app.

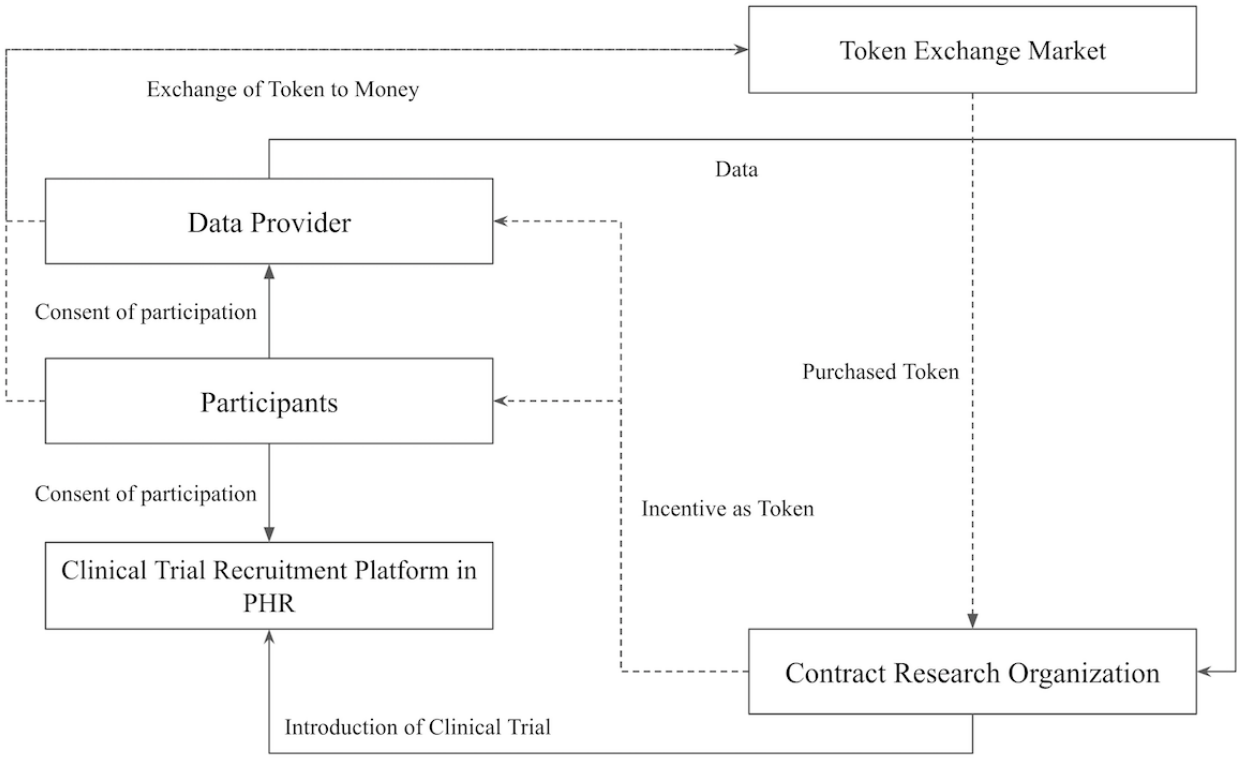
Figure 2. Basic scenario. PHR: personal health record.



Scenario for the Validation of the Model

We established the validation scenario by modifying the basic scenario (Figure 3). In the validation scenario, the research organization searches for participants who have specific gene mutations on the blockchain network for clinical trials of anticancer drugs. For convenience, participants who satisfy all the information required by the researcher are described as group A. Potential participants who do not have necessary clinical information are described as group B. The researcher sends a request for consent to both group A and group B.

Figure 3. Validation scenario. PHR: personal health record.



The participants will decide to participate in the research based on the opinion of doctors who are treating their cancer. In our study, we assumed that the smaller the amount of clinical information that participants are required to send, the higher the probability of consent. This is based on the general assumption that the more information participants have to provide, the more careful they are because normally people care about privacy and security.

For example, if participant X in group A agrees to consent, the institution storing the data of participant X shares the data with the research company and the institution obtains cryptocurrency compensation. In the validation scenario, the companies transfer tokens to data providers that have clinical information for candidate patients in exchange for compensation for storing the data, which is different from the basic scenario in that data providers are added in the token flow

Mathematical Modeling of the Scenario

Mathematical models for both scenarios were constructed; a methodology that minimizes the cost of the organizers was modeled. The essential members were participants and organizers. In both scenarios, the reward was the main parameter that affects participation—we assumed that the higher the reward, the greater the probability of consent. In the case of an organizer, an assumption is needed about which information can be obtained from the participants and the organizers' cost policies. The cost was divided into 2 categories: reward-related parameters and time cost–related parameters.

Cost Function of a Life Insurance Company

We assume that the organizer's cost function is divided into 2 parts—reward and cost. Rewards are provided to participants; therefore, the reward portion is multiplied by the total number of participants. The time is also included in the time cost, and because the rewards and time are different in units, it is necessary to balance both parts by giving one part a weight. In our model, we give a weight to the time cost part (the time weight). The time weight may vary in different recruitment programs. A higher time weight means that an organizer values recruitment time more and wants to recruit participants quickly. We express the cost function as $cost = reward \times N + t \times \alpha$.

When minimizing the cost, if the time weight α is small, cost will be largely dependent on the reward. The unit of time t, where t is a natural number, is assumed to be day for convenience.

In addition, the recruiting time calculation method should be defined. The organizer sends consent requests to those who meet the requirements, and it is assumed that each participant has a probability of consent. Under the assumption, the time taken for recruitment is calculated as follows. On day 1, the organizer sends consent requests to the people in the participant

pool. Some participants agree based on their probability, and others disagree. If the number of people who agree does not satisfy the required number of recruits, on day 2, the organizer will send the consent request again to those who did not agree. This process is repeated to calculate the recruiting time. Recall that the probability of consent to participate in research is dependent upon the reward, and the higher the reward, the greater the probability of consent. Increasing the probability of consent means it will be easier to recruit participants, and the time *t* for recruitment completion will be reduced.

Modeling the Basic Scenario

We assumed that participants can access information about how long the program lasts and the reliability of the insurance company. They also consider inconvenience that may be caused by joining the program.

The welfare of a participant is separated into 2 parts: (1) monetary value and (2) labor. In order to have unbiased results, the socioeconomic status of participants should be considered. The lower the status, the more significant impact on the monetary part, which results in biased recruitment; therefore, we introduced a concave function to represent monetary welfare. Participants have information about their socioeconomic positions, the degree of annoyance for recording tasks, and the reliability of life insurance companies, which are not disclosed to the research companies. Let the socioeconomic status of the *i*th participant be SP_i, let the degree of annoyance be b_i, and let the reliability of the company be R_i. Life insurance companies can set the program period *T* and compensation rewards for the programs, which are communicated to participants through apps.

The participant's welfare is defined using the square root (which is a concave function), which means the higher SP_i, the lower the incremental increase in total welfare of participants. The welfare function of the *i*th participant is

The former term is the weight given to the welfare according to the socioeconomic status, and the latter term is the fatigue of continuously uploading the log by multiplying the basic quadratic function for *T* by weight. Therefore, for participants, a sufficient condition for participation is that the welfare function is positive.

Socioeconomic status is likely to be biased because it provides compensation for participation. Therefore, at initial recruitment, a certain number of the desired population is selected to ensure that socioeconomic status is as evenly matched as possible. In labor-related welfare, fatigue builds up faster over time, and each person's degree of annoyance is different. Hence, the labor-related cost is proportional to the degree of annoyance and the square of the duration.

In addition to consideration of financial benefits, the probability of participating in the program was introduced (to consider the psychology of real-world participants). The probability will be large when the reward is large, the reliability of the organizer is high, and the socioeconomic status is low.

Assuming that the welfare function of the *i*th participant is positive, the probability that the *i*th participant will consent is calculated based on the following assumptions: (1) the higher the rewards, the higher the confidence in the research company; and (2) the lower the socioeconomic level of the participant, the higher the probability of participation. Thus, the probability is calculated as

The normalizer is a buffer weight to allow the rewards to grow. If the right term of equation 2 is greater than 1 for most of *i*, the simulation becomes meaningless; thus, the right term is changed to a slightly more meaningful simulation by multiplying the normalizer by less than 1. In the basic scenario, SP_i=5-10, b_i=0.05-0.1, R_i=0.5-0.7, and *normalizer*=200.

Modeling the Validation Scenario

We modified the previous assumptions and added an intermediate data provider in the validation scenario to ensure that the model is robust after we change variables.

For data providers, assumptions are not necessary because there is no separate strategy available; however, there is a higher chance of making a profit in proportion to the amount of information held by the data providers. A research company can define its cost function, which sets the compensation and time weight per participant. Participants must decide whether to agree or not for the consent request they received. Theoretically, participants decide to agree when their economic gains are greater than 0. In real life, however, other factors play a role in determining whether a patient will be able to participate in clinical trials. Thus, we assume that the agreement probability is based on the expert advice which is provided through the blockchain platform. Normally, patients get expert opinions from their oncologists. If *k* is the number or range of treatment options available, the smaller the *k*, the higher the probability of consent. Utilizing a sigmoid function, the probability of consent has a value between 0 and 1.

Participants in research must disclose whether their genes have been tested, genotype (if they have been tested), and which data provider is storing their data. Moreover, some information on the medical blockchain is open to companies: (1) whether one's genes have been tested and (2) genotype. A few properties are needed to describe this. First, DP_i is defined to determine whether the genetic data of the *i*th participant are stored by the data provider. A DP_i of 0 indicates that there are no genomic data, and a value greater than *n* means that information is stored by the *n*th data provider. The participant's genotype is defined as $Type_i$. If $Type_i$ is 0, it means that the genotype is not known because it has not examined, and DP_i has a value of 0. If $Type_i$ is *n* with a value greater than 0, the dielectric means it is *n*th type.

In the basic scenario, the probability of consent of participants is the expert advice. The greater the rewards, the smaller the number of protocols available to participants and the greater the expert advice. The probability can be defined as a simple

product in the basic scenario, but in the validation scenarios, the sigmoid function is used,

The sigmoid function converts a real value to a value between 0 and 1. Therefore, this value has the advantage of being used directly as a probability. When x is 0, the probability should be close to 0. To shift the sigmoid function to the right, *middle* should be introduced. In the validation scenario, SP_i=0-3, k=1-5, DP_i=1-2, and *middle*=150.

Optimization Formulation

We chose the convex optimization model. For each scenario, we suppose that the expected recruitment time can be found when the clinical trial is given a reward. For each reward and time of completion of the recruitment pair, we obtain the cost of applying the pair. The reward and time t at the lowest cost point will be the optimization values.

To find the expected recruitment time for a given reward amount, the probability of each participant agreeing to participate in the clinical trial is determined. Then using Bernoulli implementation, we obtain the expected number of participants that agreed until N days. Suppose that the probability that the ith participant agrees is q_i. In that case, whether one will eventually agree by the maximum N days is the same as the Bernoulli implementation. The probability of agreeing on the first day is q_i, the probability of agreeing on the first day and agreeing on the second day is $(1 - q_i)q_i$, and the probability of agreeing on the third day is $(1 - q_i)^2q_i$. When this is implemented by the jth day, the expected number of participants that agreed until j days can be obtained as

Therefore, if these expectations are added for all i, the expected number of participants until n days is obtained. When rewards and n are given, the expected number of participants can be obtained, and if n is obtained, this value n is the expected completion of the recruitment.

Aggregating the expected values, we determine the expected number of participants upon N days. When we find the minimal N that the expected number of participants exceeds the company targeted number of participants, then we obtain the reward and the minimal N pair. We put the pairs into the cost function, and find (reward, N, cost) pair that the cost is minimal.

Results

Optimization

The reward affects the probability of consent of each participant, and this probability of consent affects the expected recruitment period t. Note that the expected recruitment period is a function of the reward. Therefore, rewards and expected recruitment period pairs can be obtained. We can make a tuple by considering the cost value of these pairs applied to the cost function. Thus, the relationship of t with the reward and the cost relationship with the reward can be obtained. Figure 4 shows the relationship when the company's time weight is 1500 in the basic scenario, and Figure 5 shows the relationship in the validation scenario in which the company's time weight is 100. Because the basic scenario assumes a company cannot access some of the participant's information, we should use expected values of distributions when we calculate the expected recruitment time. Therefore, even if participants' information is changed through initialization, the graphs are unified into one. On the other hand, the validation scenario allows the company to access all of the participant's information, which means that we should consider each participant’s information when we calculate the expected recruitment time. Therefore, as the participant's information changes, the graph changes accordingly. Figure 5C and Figure 5D show 10 initializations.

In Figures 4 and 5, the graphs on the left are the relationships between rewards and expected recruitment periods, and the graphs on the right are the relationship between rewards and costs. The smaller the rewards, the larger the expected recruitment period, and vice versa. The right graph that puts the rewards and expected recruitment period in the cost function forms a convex curve. The left side of the shape is high because the time cost has grown, and the right side is large because the reward cost affects the entire cost. Therefore, the left part shows a similar appearance as the time decreases dramatically. The right side becomes a linear graph for the reward because the time cost becomes very small. The middle part of the trade-off between these 2 will be the point where the cost is minimized, and the reward value and the expected recruitment period will be *reward** and *t**.

Figure 4. (A) Time–reward and (B) cost–reward trade-offs in the basic scenario when the time weight is equal to 1500.

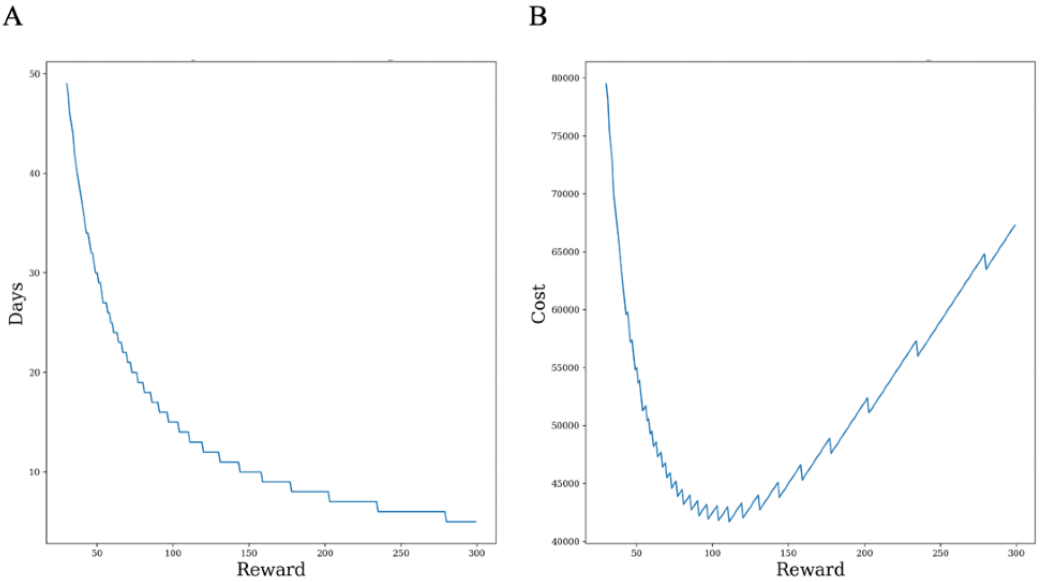
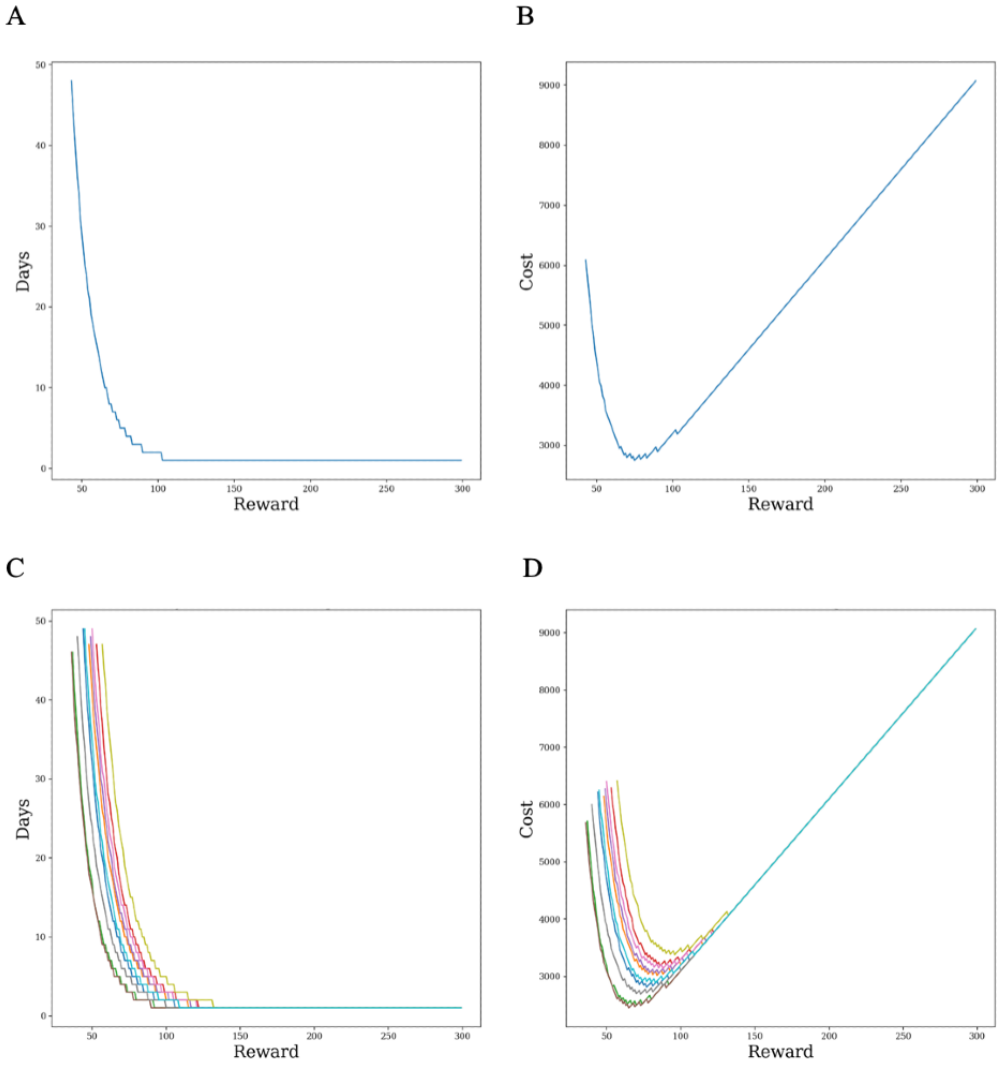


Figure 5. Single-case (A) time–reward and (B) cost–reward and multiple-case (C) time–reward and (D) cost–reward trade-offs in the validation scenario when the time weight is equal to 100.



Comparison With Simulation

We compared the simulation results with optimization results. Since *reward** is an independent variable, it is appropriate to compare dependent variable *t**. In order to obtain robust results, we conducted repetitive experiments by changing the variables of participants randomly. For each initialization, we simulated 100 times to get the experimental results. Then, we calculate the error between 100 experimental results and recruitment time, *t**. Finally, we obtained the mean and standard deviation of those values. In the basic scenario, the simulation values are concentrated at the optimized values (Figure 6, Table 1).

Figure 7 and Table 2 show differences between optimized results up to a maximum of 2 days for the validation scenario.

Figure 6. Basic scenario mean values and standard deviation of errors (expected days=13).

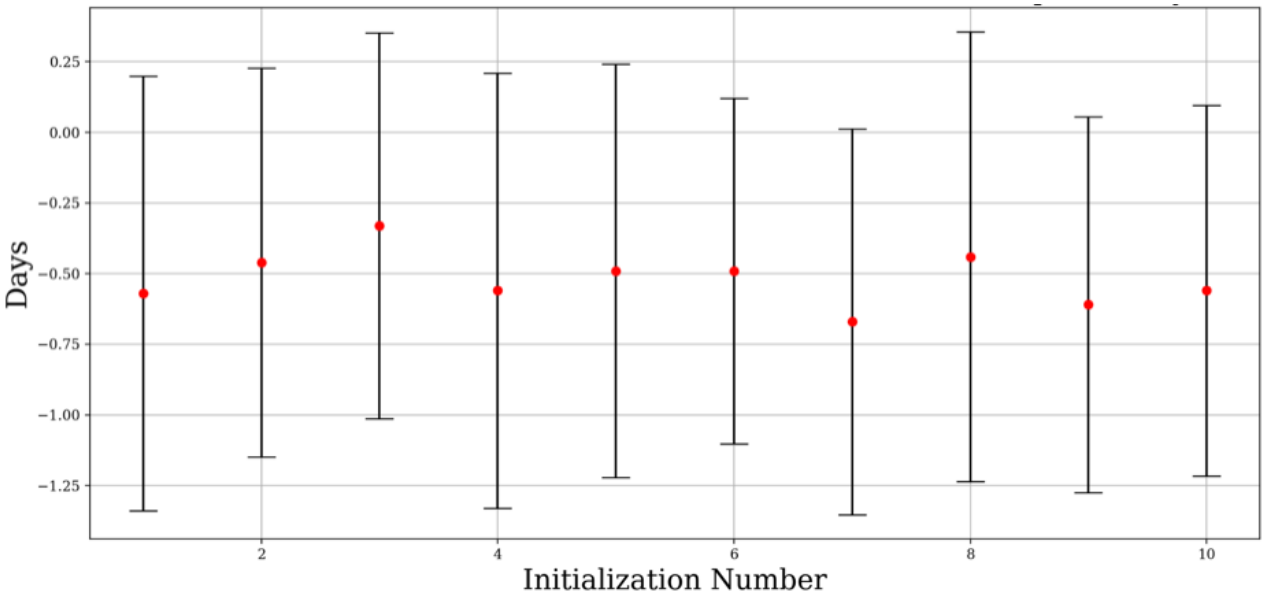


Table 1. Basic scenario.

Experiment	Error, mean (SD)
1	–0.57 (0.77)
2	–0.46 (0.69)
3	–0.33 (0.68)
4	–0.56 (0.77)
5	–0.49 (0.73)
6	–0.49 (0.61)
7	–0.67 (0.68)
8	–0.44 (0.80)
9	–0.61 (0.67)
10	–0.56 (0.66)

Figure 7. Validation scenario mean values and standard deviation of errors.

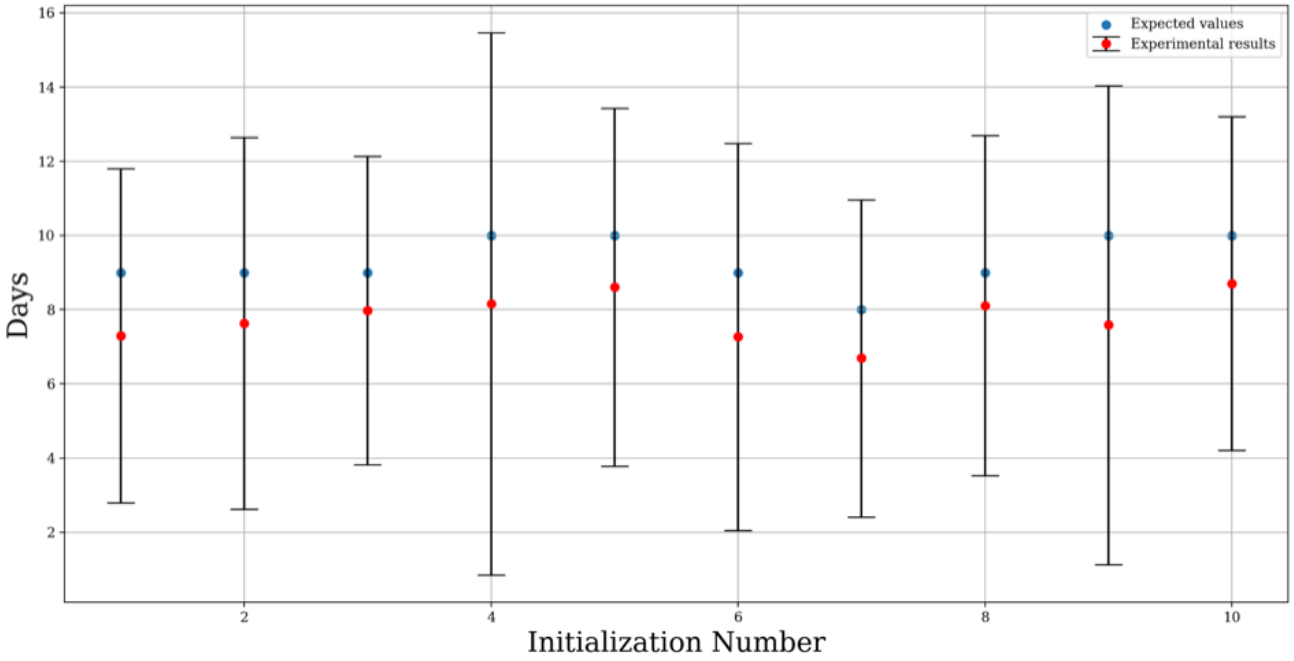


Table 2. Validation scenario.

Experiment	Expected days, n	Error, mean (SD)
1	9	–1.70 (4.51)
2	9	–1.37 (5.01)
3	9	–1.02 (4.16)
4	10	–1.84 (7.31)
5	10	–1.39 (4.82)
6	9	–1.73 (5.22)
7	8	–1.31 (4.28)
8	9	–0.89 (4.59)
9	10	–2.41 (6.45)
10	10	–1.30 (4.50)

Sensitivity Analysis

The company can adjust the number of participants N and time weight α; therefore, we analyzed the sensitivity of t^* as N and α change. For 1% increments of N, recruitment time increases 1.54% and 2.22% in the basic and validation scenarios, respectively. If the time weight is increased by 1%, the recruitment time is decreased by 1.54% and 1.67% in the basic and validation scenarios, respectively, which means this model is in line with our general knowledge and provides a guideline for designing recruitment of participants.

Discussion

Principal Findings

In this paper, we propose 2 token economy scenarios of health care blockchain with mechanism design. We set basic components in each scenario, which were the number of participants and recruitment time, constructed mathematical models to explain the 2 scenarios, and simulated changes in recruitment time and the number of expected participants. Through mechanism design, we demonstrated that the recruiter is able to set a desired and expected number of participants and recruitment time by adjusting the amount of incentive. This study is the first, to the best of our knowledge, to apply mechanism design to health care blockchain for real-world problems.

In classic game theory, designers of the game are not able to determine expected results of games quantitatively. They can expect rational participants to compete with each other for the best results by assuming that participants are reasonable, therefore, act to maximize their profits; however, it is not realistic to apply this assumption directly in the health care blockchain with token economy because it is crucial for a funder of health care blockchain to know expected time, cost, and the number of recruited participants.

In our models that employ mechanism design, the recruiter can set a quantitative outcome of programs by adjusting the relationship of number of participants, recruitment time, and rewards. In addition to the predictability of outcomes, the blockchain system can provide participants more secure environments with immutability of the system, although the system is more open to the public compared to conventional ways of recruiting participants.

Traditionally, the issue of whether direct incentives for health care participants are acceptable has been debated because direct incentives can bias medical research and treatment outcomes [44]. From a medical standpoint, there is less of an ethical problem with direct incentives for healthy people [45], such as those modeled in the basic scenario. In the validation scenario, there may be a concern with direct incentives to participants within the blockchain. However, we assumed a circumstance that allows direct incentives for the participation of clinical trials in the future because patients are selected randomly as participants on a blockchain system, and we assumed that direct incentives can be allowed to increase recruitment and attention rates in clinical trials.

From a mathematical perspective, information about each participant is different depending on whether recruiting companies have access to it or not. While recruiting healthy individuals, we assumed that information about each participant is unobtainable by the company because they usually do not need strict inclusion criteria, which means that they can exclude some of applicants after closing the recruitment. When collecting life-log data, the degree of annoyance for each participant over a long period is unobtainable and we can only see the distribution for the degree of annoyance. In this study, this distribution is assumed to be a uniform distribution. In this case, a clinical trial company uses the expected value of the distribution (ie, the uniform distribution's average value). Meanwhile, in the validation scenario, each participant should provide gene data to the research company. From the gene data, the company can infer the probability of agreement of each participants and can adjust model to fit real-world data.

Recruiting participants is one of the most difficult parts of conducting a clinical research or trial. However, the token economy, through mathematical modeling and simulation techniques proposed in this study, will enable companies to obtain important insight into whether they can recruit participants within the desired period by setting appropriate rewards for recruiting participants. Moreover, due to difficulties in recruiting participants suitable for clinical trials, companies may make a request to hospitals for patients information. However, hospitals do not feel the need and reason to provide data they hold, or even if hospitals are willing to provide data, the consent process for the use of patient data is complicated, and many patients are concerned about privacy and security [46]. Therefore, blockchain token economy can be applied to appropriately compensate entities participating in the blockchain data sharing platform, thereby reducing gaps between the needs of companies, hospitals, and participants (patients) with respect to data utilization, with high security and privacy, enabling efficient recruitment of participants through user-centered participation. This study is meaningful because we were able to make rational models that can be used as a starting point for designing health care blockchain for patient recruitment.

Limitations and Future Research

Our assumptions did not reflect complex circumstances for recruiting participants in the real world. Thus, if the results of our models are different from the expectations of recruiting company, recruitment would not be successful. Models with basic mandatory requirements such as recruitment time, amount of incentives, and number of participants, and without complicated assumptions, can be useful for other researchers as a starting point of their own modeling—they can use our framework as a basic scenario to design their own sophisticated token design in health care blockchain. Another limitation is that comparison and analysis with actual data could not be carried out to provide a more robust framework.

Conclusion

There have been few business models adopting blockchain technology in health care. Token economy of blockchain can be a powerful driver by incentivizing health care consumers with an immutably trackable token transaction system. This research is a starting point of designing the token economy in real-world health care settings; we modeled 2 possible scenarios, optimized the cost of the company, and compared the results with simulation data. This study can be used as a basis for designing an incentive system for companies recruiting clinical trial participants or health care program participants within a set period of time with a blockchain-based patient recruitment platform.

Acknowledgments

This research was supported by the Seoul National University Bundang Hospital Research Fund (grant 02-2020-0031), the Korea Health Industry Development Institute of the Ministry of Health and Welfare (HI19C0152), the National Research Foundation of Korea grant funded by the Korean Ministry of Science and ICT (2017R1E1A1A03070105), and by the Information Technology Research Center support program supervised by the Institute for Information and Communications Technology Promotion (IITP-2018-0-01441). We are especially thankful to Ilha Yune for editing the manuscript.

Authors' Contributions

SYJ and THK drafted the manuscript. KPH contributed to the discussion. As the corresponding author, HJH supervised the process of writing the manuscript.

Conflicts of Interest

None declared.

References

1. Ashley EA. Towards precision medicine. Nat Rev Genet 2016 Aug 16;17(9):507-522. [doi: 10.1038/nrg.2016.86] [Medline: 27528417]
2. All of Us Research Program Investigators. The "All of Us" research program. N Engl J Med 2019 Aug 15;381(7):668-676. [doi: 10.1056/nejmsr1809937]
3. Blumenthal D. Stimulating the adoption of health information technology. N Engl J Med 2009 Apr 09;360(15):1477-1479. [doi: 10.1056/NEJMp0901592] [Medline: 19321856]
4. Adler-Milstein J, Jha AK. HITECH act drove large gains in hospital electronic health record adoption. Health Aff (Millwood) 2017 Aug 01;36(8):1416-1422. [doi: 10.1377/hlthaff.2016.1651] [Medline: 28784734]
5. Percent of specialty hospitals that possess certified health IT. The Office of the National Coordinator for Health Information Technology. URL: https://dashboard.healthit.gov/quickstats/pages/specialty-hospital-ehr-adoption.php [accessed 2020-02-13]
6. Digital health South Korea, market intelligence report. Department for International Trade Report. 2019 Jun 01. URL: https://www.intralinkgroup.com/getmedia/3153c79b-463d-47c7-84e6-56848c98aab7/Intralink-Report_Life-Sciences_June2019 [accessed 2020-01-01]
7. Carrión Señor I, Fernández-Alemán JL, Toval A. Are personal health records safe? a review of free web-accessible personal health record privacy policies. J Med Internet Res 2012 Aug 23;14(4):e114 [FREE Full text] [doi: 10.2196/jmir.1904] [Medline: 22917868]
8. Roehrs A, da Costa CA, Righi RDR, de Oliveira KSF. Personal health records: a systematic literature review. J Med Internet Res 2017 Jan 06;19(1):e13 [FREE Full text] [doi: 10.2196/jmir.5876] [Medline: 28062391]
9. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
10. Kim S. A study on the effect of blockchain on personal information protection. J Multimed Inf Syst 2019 Sep 30;6(3):125-130. [doi: 10.33851/jmis.2019.6.3.125]
11. Choudhury O, Sylla I, Fairoza N. A blockchain framework for ensuring data quality in multi-organizational clinical trials. 2019 Presented at: IEEE International Conference on Healthcare Informatics; June 10-13; Xian, China p. 1-9. [doi: 10.1109/ichi.2019.8904634]
12. Gordon WJ, Catalini C. Blockchain technology for healthcare: Facilitating the transition to patient-driven interoperability. Comput Struct Biotechnol J 2018;16:224-230 [FREE Full text] [doi: 10.1016/j.csbj.2018.06.003] [Medline: 30069284]
13. Buterin V. A next-generation smart contract and decentralized application platform. Cryptorating. URL: https://cryptorating.eu/whitepapers/Ethereum/Ethereum_white_paper.pdf [accessed 2020-01-01]
14. Maslove DM, Klein J, Brohman K, Martin P. Using blockchain technology to manage clinical trials data: a proof-of-concept study. JMIR Med Inform 2018 Dec 21;6(4):e11949 [FREE Full text] [doi: 10.2196/11949] [Medline: 30578196]
15. Li Y, Shan B, Li B, Liu X, Pu Y. Literature review on the applications of machine learning and blockchain technology in smart healthcare industry: a bibliometric analysis. J Healthc Eng 2021;2021:9739219 [FREE Full text] [doi: 10.1155/2021/9739219] [Medline: 34426765]
16. Hussien HM, Yasin SM, Udzir SNI, Zaidan AA, Zaidan BB. A systematic review for enabling of develop a blockchain technology in healthcare application: taxonomy, substantially analysis, motivations, challenges, recommendations and future direction. J Med Syst 2019 Sep 14;43(10):320. [doi: 10.1007/s10916-019-1445-8] [Medline: 31522262]
17. Liang X, Zhao J, Shetty S. Integrating blockchain for data sharing and collaboration in mobile healthcare applications. 2017 Presented at: IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC); October 8-13; Montreal, Canada. [doi: 10.1109/pimrc.2017.8292361]
18. Zhang P, Schmidt D, White J. Blockchain technology use cases in healthcare. In: Advances in Computers. England: Elsevier; Jan 01, 2018.
19. Parkinson B, Meacock R, Sutton M, Fichera E, Mills N, Shorter GW, et al. Designing and using incentives to support recruitment and retention in clinical trials: a scoping review and a checklist for design. Trials 2019 Nov 09;20(1):624-621 [FREE Full text] [doi: 10.1186/s13063-019-3710-z] [Medline: 31706324]
20. Hickey V, Flesch L, Lane A, Pai AL, Huber J, Badia P, et al. Token economy to improve adherence to activities of daily living. Pediatr Blood Cancer 2018 Nov 26;65(11):e27387 [FREE Full text] [doi: 10.1002/pbc.27387] [Medline: 30051581]
21. Sood A, Prasad K, Chhatwani L, Shinozaki E, Cha SS, Loehrer LL, et al. Patients' attitudes and preferences about participation and recruitment strategies in clinical trials. Mayo Clinic Proceedings 2009 Mar;84(3):243-247. [doi: 10.4065/84.3.243] [Medline: 19252111]
22. Carroll JM. Co-production scenarios for mobile time banking. In: End-User Development. Berlin Heidelberg: Springer; Jan 01, 2013:137-152.

23. Han K, Shih P, Bellotti V. It's time there was an app for that too: a usability study of mobile timebanking. In: Multigenerational Online Behavior and Media Use: Concepts, Methodologies, Tools, and Applications. Hershey, Pennsylvania: IGI Global; Jan 01, 2019:242-265.
24. Jennings CG, MacDonald TM, Wei L, Brown MJ, McConnachie L, Mackenzie IS. Does offering an incentive payment improve recruitment to clinical trials and increase the proportion of socially deprived and elderly participants? Trials 2015 Mar 07;16(1):80-81 [FREE Full text] [doi: 10.1186/s13063-015-0582-8] [Medline: 25888477]
25. Ruffin MT, Nease DE. Using patient monetary incentives and electronically derived patient lists to recruit patients to a clinical trial. J Am Board Fam Med 2011 Sep 07;24(5):569-575 [FREE Full text] [doi: 10.3122/jabfm.2011.05.100169] [Medline: 21900440]
26. Radley A, Ballard P, Eadie D, MacAskill S, Donnelly L, Tappin D. Give It Up For Baby: outcomes and factors influencing uptake of a pilot smoking cessation incentive scheme for pregnant women. BMC Public Health 2013 Apr 15;13(1):343-341 [FREE Full text] [doi: 10.1186/1471-2458-13-343] [Medline: 23587161]
27. Volpp KG, Troxel AB, Pauly MV, Glick HA, Puig A, Asch DA, et al. A randomized, controlled trial of financial incentives for smoking cessation. N Engl J Med 2009 Feb 12;360(7):699-709. [doi: 10.1056/nejmsa0806819] [Medline: 19213683]
28. Vilponen J. Digital therapeutics solution and business model creation: multiple-case study through VISOR lens. Åbo Akademi University. URL: https://www.doria.fi/handle/10024/173205 [accessed 2020-01-01]
29. Afra P, Bruggers CS, Sweney M, Fagatele L, Alavi F, Greenwald M, et al. Mobile software as a medical device (SaMD) for the treatment of epilepsy: development of digital therapeutics comprising behavioral and music-based interventions for neurological disorders. Front Hum Neurosci 2018 May 1;12:171-171 [FREE Full text] [doi: 10.3389/fnhum.2018.00171] [Medline: 29780310]
30. Khirasaria R, Singh V, Batta A. Exploring digital therapeutics: the next paradigm of modern health-care industry. Perspect Clin Res 2020;11(2):54-58 [FREE Full text] [doi: 10.4103/picr.PICR_89_19] [Medline: 32670828]
31. Babaioff M, Dobzinski S, Oren S. On bitcoin and red balloons. 2012 Presented at: 13th ACM Conference on Electronic Commerce; June 4-8; Valencia, Spain p. 56-73. [doi: 10.1145/2229012.2229022]
32. Lewenberg Y, Bachrach Y, Sompolinsky Y. Bitcoin mining pools: a cooperative game theoretic analysis. 2015 Presented at: International Conference on Autonomous Agents and Multiagent Systems; May 4-8; Istanbul, Turkey p. 919-927.
33. Kiayias A, Koutsoupias E, Kyropoulou M. Blockchain mining games. 2016 Presented at: ACM Conference on Economics and Computation; July 24-28; Maastricht The Netherlands p. 365-382. [doi: 10.1145/2940716.2940773]
34. Jackson M. Mechanism Theory. SSRN Journal 2009 Feb 01;6(1):5-20. [doi: 10.2139/ssrn.2542983]
35. Ray A, Ventresca M, Wan H. A mechanism design approach to blockchain protocols. 2018 Presented at: IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData); July 2-7; San Francisco, California p. 1603-1608. [doi: 10.1109/cybermatics_2018.2018.00268]
36. Mamageishvili A, Schlegel J. Mechanism design and blockchains. arXiv. Preprint posted online on May 5, 2020. [FREE Full text]
37. Toyoda K, Zhao J, Zhang ANS, Mathiopoulos PT. Blockchain-enabled federated learning with mechanism design. IEEE Access 2020;8:219744-219756. [doi: 10.1109/access.2020.3043037]
38. Toyoda K, Zhang A. Mechanism design for an incentive-aware blockchain-enabled federated learning platform. 2019 Presented at: IEEE International Conference on Big Data; December 9-12; Los Angeles, California p. 395-403. [doi: 10.1109/bigdata47090.2019.9006344]
39. Teichmann F, Falker MC. Blockchain: implications of the impending token economy. In: Scientific and Technical Revolution: Yesterday, Today and Tomorrow. New York City: Springer International Publishing; Jan 01, 2020:1551.
40. Game theory. Wikipedia. URL: https://en.wikipedia.org/w/index.php?title=Game_theory&oldid=1017158618 [accessed 2021-04-24]
41. Moura J, Marinheiro R, Silva J. Game theory for collaboration in future networks. In: Mobile Computing and Wireless Networks: Concepts, Methodologies, Tools, and Applications. Hershey, Pennsylvania: IGI Global; 2016:2061-2062.
42. Patel DN, Lambert EV, da Silva R, Greyling M, Nossel C, Noach A, et al. The association between medical costs and participation in the vitality health promotion program among 948,974 members of a South African health insurance company. Am J Health Promot 2010;24(3):199-204. [doi: 10.4278/090217-QUAN-68R2.1] [Medline: 20073387]
43. Swencionis C, Wylie-Rosett J, Lent MR, Ginsberg M, Cimino C, Wassertheil-Smoller S, et al. Weight change, psychological well-being, and vitality in adults participating in a cognitive-behavioral weight loss program. Health Psychol 2013 Apr;32(4):439-446 [FREE Full text] [doi: 10.1037/a0029186] [Medline: 22888821]
44. Hsieh G, Kocielnik R. You get who you pay for: the impact of incentives on participation bias. 2012 Presented at: 19th ACM Computer Supported Cooperative Work and Social Computing; Feb 27-Mar 2; San Francisco, California. [doi: 10.1145/2818048.2819936]
45. London AJ, Borasky DA, Bhan A, Ethics Working Group of the HIV Prevention Trials Network. Improving ethical review of research involving incentives for health promotion. PLoS Med 2012;9(3):e1001193 [FREE Full text] [doi: 10.1371/journal.pmed.1001193] [Medline: 22479154]

46. Esmaeilzadeh P, Sambasivan M. Patients' support for health information exchange: a literature review and classification of key factors. BMC Med Inform Decis Mak 2017 Apr 04;17(1):33 [FREE Full text] [doi: 10.1186/s12911-017-0436-2] [Medline: 28376785]

Edited by G Eysenbach; submitted 28.12.20; peer-reviewed by L Cilliers, Y Chu; comments to author 20.01.21; revised version received 12.03.21; accepted 07.08.21; published 13.09.21.

Please cite as:
Jung SY, Kim T, Hwang HJ, Hong K
Mechanism Design of Health Care Blockchain System Token Economy: Development Study Based on Simulated Real-World Scenarios
J Med Internet Res 2021;23(9):e26802
URL: https://www.jmir.org/2021/9/e26802
doi:10.2196/26802
PMID:34515640

Original Paper

Scholarly Productivity Evaluation of KL2 Scholars Using Bibliometrics and Federal Follow-on Funding: Cross-Institution Study

Kelli Qua[1*], PhD; Fei Yu[2,3*], MPS, PhD; Tanha Patel[4*], MPH; Gaurav Dave[4,5*], MPH, MD, DrPH; Katherine Cornelius[6*], MPH; Clara M Pelfrey[1*], PhD

[1]Clinical and Translational Science Collaborative, School of Medicine, Case Western Reserve University, Cleveland, OH, United States

[2]Health Sciences Library, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[3]School of Information and Library Science, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[4]North Carolina Translational and Clinical Sciences Institute, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[5]Division of General Medicine and Clinical Epidemiology, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[6]Center for Clinical and Translational Science, Mayo Clinic, Rochester, MN, United States

[*]all authors contributed equally

Corresponding Author:
Fei Yu, MPS, PhD
Health Sciences Library
University of North Carolina at Chapel Hill
335 S. Columbia Street CB 7585
Chapel Hill, NC, 27599
United States
Phone: 1 9199622219
Email: feifei@unc.edu

Abstract

Background: Evaluating outcomes of the clinical and translational research (CTR) training of a Clinical and Translational Science Award (CTSA) hub (eg, the KL2 program) requires the selection of reliable, accessible, and standardized measures. As measures of scholarly success usually focus on publication output and extramural funding, CTSA hubs have started to use bibliometrics to evaluate the impact of their supported scholarly activities. However, the evaluation of KL2 programs across CTSAs is limited, and the use of bibliometrics and follow-on funding is minimal.

Objective: This study seeks to evaluate scholarly productivity, impact, and collaboration using bibliometrics and federal follow-on funding of KL2 scholars from 3 CTSA hubs and to define and assess CTR training success indicators.

Methods: The sample included KL2 scholars from 3 CTSA institutions (A-C). Bibliometric data for each scholar in the sample were collected from both SciVal and iCite, including scholarly productivity, citation impact, and research collaboration. Three federal follow-on funding measures (at the 5-year, 8-year, and overall time points) were collected internally and confirmed by examining a federal funding database. Both descriptive and inferential statistical analyses were computed using SPSS to assess the bibliometric and federal follow-on funding results.

Results: A total of 143 KL2 scholars were included in the sample with relatively equal groups across the 3 CTSA institutions. The included KL2 scholars produced more publications and citation counts per year on average at the 8-year time point (3.75 publications and 26.44 citation counts) than the 5-year time point (3.4 publications vs 26.16 citation counts). Overall, the KL2 publications from all 3 institutions were cited twice as much as others in their fields based on the relative citation ratio. KL2 scholars published work with researchers from other US institutions over 2 times (5-year time point) or 3.5 times (8-year time point) more than others in their research fields. Within 5 years and 8 years postmatriculation, 44.1% (63/143) and 51.7% (74/143) of KL2 scholars achieved federal funding, respectively. The KL2-scholars of Institution C had a significantly higher citation rate per publication than the other institutions (P<.001). Institution A had a significantly lower rate of nationally field-weighted collaboration than did the other institutions (P<.001). Institution B scholars were more likely to have received federal funding than scholars at Institution A or C (P<.001).

Conclusions: Multi-institutional data showed a high level of scholarly productivity, impact, collaboration, and federal follow-on funding achieved by KL2 scholars. This study provides insights on the use of bibliometric and federal follow-on funding data to evaluate CTR training success across institutions. CTSA KL2 programs and other CTR career training programs can benefit from these findings in terms of understanding metrics of career success and using that knowledge to develop highly targeted strategies to support early-stage career development of CTR investigators.

(J Med Internet Res 2021;23(9):e29239) doi:10.2196/29239

KEYWORDS

bibliometrics; Clinical and Translational Science Award; KL2; translational research; career development

Introduction

Evaluating outcomes of a Clinical and Translational Science Award (CTSA) hub's clinical and translational research (CTR) training, such as the KL2 program, requires the selection of reliable, accessible, and standardized measures. The KL2 program is a multiyear mentored training award focusing on early-stage career development of investigators in CTR. The National Center for Advancing Translational Science (NCATS) at the National Institutes of Health (NIH) funds over 60 KL2 programs across CTSA hubs. All CTSA hubs offer a KL2 program, which is a formal, mentored training experience for scholars with doctoral degrees. Each CTSA hub selects KL2 candidates from a variety of fields (eg, medicine, nursing, and biostatistics) to participate in translational research training in a multidisciplinary setting with up to 5 years of career development support.

One of the strategic goals of NCATS for the CTSA program is to "develop and foster innovative translational training and a highly skilled, creative, and diverse translational science workforce" [1]. The KL2 program is one mechanism that offers such training. Each CTSA hub is responsible for evaluating its KL2 program, typically using methods such as surveys, focus groups, exit interviews, and alumni follow-up. Measures of scholarly success are predominantly scholarship products—publication output and extramural funding, such as an R01-equivalent award. The NCATS Common Metrics Initiative also created measures for all KL2 programs to report on the number and percentage of total graduates: women, underrepresented minorities, and KL2 scholars who sustain their translational research engagement [2].

All NCATS-funded KL2 programs follow a set of established requirements. For example, 75% of the time of enrolled KL2 scholars is funded by the program, with the exception of 50% for surgeons [3]. All KL2 programs must provide training in rigorous research methodologies aligned with CTR competencies. KL2 programs must provide opportunities that allow scholars to effectively communicate and collaborate across multidisciplinary teams [3,4]. Beyond these requirements, CTSA hubs are encouraged to innovate and tailor their KL2 programs to their specific needs, leading to some variability in KL2 programs. For example, although programs provide 2 years of KL2 funding, some programs provide funding for 3 or more years [5]. Such differences in hub-level KL2 program characteristics present challenges for defining indicators of success and standardizing the evaluation of KL2 program outcomes across CTSA hubs.

CTSA hubs sometimes use bibliometric analysis to evaluate their impact on moving translational research forward. Bibliometrics refers to the use of quantitative and statistical methods to analyze a chosen group of publications [6,7]. Bibliometrics complement other evaluation methods, such as surveys and interviews, by allowing programs to measure tangible and intangible outcomes such as publications and scholarly impact, which otherwise cannot be objectively measured via self-report methods such as surveys or interviews. One study [8] used bibliometrics to analyze publications citing all CTSA hub grants from 2006 to 2016. This study highlighted bibliometric values to assess the impact of research across all CTSAs as measured by interdisciplinary collaboration, influence on other publications, and breadth of scientific fields. Another study [9] used bibliometrics to compare publications citing one of 6 CTSA hubs with measures from three sources: NIH iCite (publicly available), Thomson Reuters (now Clarivate Analytics), and Elsevier. This study identified relevant data sources and standardized analyses for cross-CTSA comparisons. A more recent CTSA hub-specific study [10] implemented advanced bibliometric measures to assess scholarly productivity, citation impact, the scope of research collaboration, and clusters of research topics of publications supported by their CTSA. Furthermore, some CTSA hubs have evaluated the feasibility of different bibliometric approaches to assessing KL2 scholarly productivity and influence compared with other NIH-funded K-awardees [11,12]. These studies illustrated the importance of standardizing the (1) collection of publication data, (2) definition of success and measures for KL2 training outcomes, and (3) use of bibliometric methods to evaluate the scholarly productivity and impact of KL2 scholars across the CTSA consortium.

To our knowledge, the evaluation of KL2 programs across CTSAs is limited, and the use of bibliometrics and follow-on funding is minimal. This study evaluates scholarly productivity using bibliometrics and federal follow-on funding of KL2 scholars across 3 different CTSA hubs to define and assess CTR training success indicators. Such indicators, coupled with evidence of the feasibility of their collection and use, would allow CTR training programs to demonstrate effectiveness on a wide variety of outcomes.

Methods

Participants

Our sample included KL2 scholars from 3 CTSA institutions, described as institutions A-C throughout the paper. A description of program-level characteristics is provided in the Multimedia

Appendix 1. The 3 institutions were selected as convenience samples based on previous collaborations between the CTSAs. We only included scholars in our analysis if they had at least 5 years of bibliometric data, starting at matriculation (ie, year of entry into the KL2 program). We chose the date range of 2005-2013 to ensure that all scholars met the minimum 5-year data requirement. In other words, scholars beginning their KL2 program in 2013 have 5 years of bibliometric data as of December 2018.

Measures

Overview

We selected bibliometrics and federal follow-on funding because they are objective measures with evidence of reliability and validity and are verifiable. In the following sections we describe various bibliometrics along a continuum of objectivity. Although some metrics, such as number of publications, are objective, other metrics, such as citation impact and collaboration, are quantified representations of construct and are open to interpretation of scale, scope, and reliability.

Bibliometrics

When we selected bibliometric measures and sources for this study, we considered the (1) validity, relevancy, and feasibility of measures, especially the metrics and sources that have been verified and adopted in previous CTSA evaluation studies; (2) coverage of publications of our KL2 scholars in three CTSAs; and (3) availability of bibliometric data sources such as institutional subscriptions and free accessibility. Accordingly, we chose Elsevier SciVal and NIH iCite for the bibliometric measurements. SciVal is a research analytics tool for various bibliometrics gathered from the Elsevier Scopus citation database. Scopus has much broader coverage in biomedical and life sciences than Web of Science [13,14] and has been adopted in multiple bibliometric studies concerning CTSA evaluations [9,10,15]. The 3 CTSA institutions in this study have subscriptions to Scopus or SciVal and have experience in using Elsevier metrics to assist research performance evaluation. In addition, we added the relative citation ratio (RCR) generated by NIH iCite as an additional citation measure. Being freely accessible and increasingly used by CTSA evaluators, RCR is a field-independent metric, measuring the citation impact of an article relative to other NIH-supported research papers produced in the same field during the same timeframe. However, iCite is limited to analyzing only articles indexed in PubMed, whereas SciVal can provide aggregated data from the more comprehensive Scopus database. Extracting data from both sources can help reduce potential bias and provide a more reliable estimate of the quantity and impact of scholarly work. Therefore, we exported bibliometric data for each of the included scholars from both SciVal and iCite to provide evidence of validity and reliability within this sample [16,17].

Data exported from SciVal provide evidence for three metrics of research performance: productivity, impact, and collaboration. Productivity metrics provide an overview of total scholarly output, that is, the number of publications a scholar produces within a specified period. Impact metrics focus on citation counts through raw and calculated variables, accounting for scientific field-weighted or ratio values. We excluded self-citations from the analysis. Collaboration metrics consider all authors in a publication with attention to the affiliated institutions of coauthors. We used iCite data to understand the impact of the publications compared with the average of NIH-funded publications published in the same year and field [18]. Table 1 shows a detailed list of metrics used per domain.

Table 1. Summary of bibliometrics used.

Source, domain, and metric	Definition
SciVal	
Productivity	
Scholarly output	Number of publications indexed in Scopus
Impact	
Citations per publication	Average number of citations received per publication
PPTPJ[a]	Number of publications in the world's top journals
FWCI[b]	Ratio of citations received relative to the expected average for the field, type, and year
Collaboration	
NFWC[c]	Collaboration ratio is computed based on the expected collaboration for that document type, publication year grouping, and subject area assignment
iCite	
Impact	
Average RCR[d]	Cites per year of each paper, normalized to the citations per year received by NIH[e]-funded papers in the same field and year
Average citations per year	Citations per year received since publication; this is the numerator for the RCR.
Average field citation rate	Intrinsic citation rate of this paper's field is estimated using its co-citation network.
Average NIH percentile	Percentile rank of this paper's RCR compared with all NIH publications.

[a]PPTPJ: percentage of publications in the top 10th percentile of journals.

[b]FWCI: field-weighted citation impact.

[c]NFWC: National Field-weighted Collaboration.

[d]RCR: relative citation ratio.

[e]NIH: National Institutes of Health.

Federal Follow-on Funding

Three federal follow-on grant funding measures were also collected: NIH funding at the 5-year, 8-year, and overall time points. These follow-on funding measures include only NIH federal funding received by a scholar as the principal investigator, coprincipal investigator, or coinvestigator. This information can be independently confirmed for all 3 institutions by examining NIH RePORTER (Research Portfolio Online Reporting Tools Expenditures and Results) for NIH funding records [19].

Procedures

This study received institutional review board exemption from 2 institutions, with a third institution determining that this was not human subjects research. Each participating institution developed a list of KL2-scholars awarded since 2005, including gender, race, ethnicity, and highest degree earned for each scholar. We dichotomized race due to sample skewness; however, dichotomization still resulted in unequal group sizes. We excluded ethnicity analysis because of the homogeneity of the sample. We collapsed the categories of degrees to MD and non-MD terminal degrees. Clinical training was the distinction between MD and non-MD degrees. Scholars with both MD and PhD were considered MDs.

Cohorts of scholars were created in SciVal using the year the scholar began their KL2-grant (ie, all 2005 KL2's were in one cohort) according to the SciVal cohort-instructions. To obtain iCite data, a list of publications associated with each KL2 scholar was retrieved from PubMed and then manually verified. The PubMed identification numbers (PMIDs) of the verified publications were filtered to the 2005-2013 date ranges for this analysis. Next, the validated PMIDs were imported into iCite to download the data for each scholar. All metrics from SciVal and iCite were extracted and entered into an SPSS (IBM; version 25) file within 6 months. Grant data were collected using internal records of federal funding and confirmed through NIH RePORTER [19].

Data Analysis

All statistical analyses were computed using SPSS [20]. Nonparametric tests for independent samples (Kruskal-Wallis and Mann-Whitney U) were used when data were not normally distributed. We conducted descriptive and inferential statistical analyses according to demographic subgroups and institutions to assess the bibliometric results of KL2 scholars. Federal funding data were analyzed using logistic regression to evaluate the relationship between categorical variables (eg, gender, race, degree, and institution) and a series of dichotomized variables representing the presence or absence of federal funding at three time points, as described below.

Data were time-bound to include the year the scholar started their K-award (ie, matriculation) through the completion of 2018. Three time points determined analysis: 5 years and 8

years after matriculation and overall (from the start year of the KL2-award through the end of 2018). These time points are based on established precedents from K08 or K23 evaluations and K to R-award funding trajectories, which peak at 8 years [21,22]. iCite data were not included for the 5-and 8-year time points because, at the time of analysis, no time point–specific data could be produced by iCite and only the overall citation metrics were available.

Results

Sample Characteristics

The total sample consisted of 143 KL2 scholars with relatively equal groups across institutions (Table 2). The overall sample at each institution had a slightly greater number of male scholars. The majority of KL2 scholars in this sample were non-Hispanic White people. Most scholars held terminal MD degrees, with PhDs being the second most common degree.

Table 2. Summary of scholar demographics[a].

Demographics	Sample value (N=143)	Institution A value (n=50)	Institution B value (n=48)	Institution C value (n=45)
Gender, n (%)				
Male	*77 (54)*[b]	*27 (54)*	*25 (52)*	*25 (55)*
Female	66 (46)	23 (46)	23 (48)	20 (44)
Ethnicity, n (%)				
Hispanic or Latino	6 (4)	4 (8)	1 (2)	1 (2)
Not Hispanic or Latino	*136 (95)*	*45 (90)*	*47 (98)*	*44 (98)*
Not reported	1 (1)	1 (2)	0 (0)	0 (0)
Race, n (%)				
American Indian	0 (0)	0 (0)	0 (0)	0 (0)
Asian	19 (13)	9 (18)	1 (2)	9 (20)
African American	10 (7)	3 (6)	2 (4)	5 (11)
White	*112 (78)*	*36 (72)*	*45 (94)*	*25 (55)*
Not reported	1 (1)	2 (4)	0 (0)	0 (0)
Degree, n (%)				
MD	*74 (52)*	*20 (40)*	*29 (60)*	*25 (55)*
PhD	47 (33)	17 (34)	15 (31)	15 (33.3)
MD and PhD	15 (11)	7 (14)	4 (8)	4 (8.9)
Other	7 (5)	6 (12)	0 (0)	1 (2.2)

[a]Values were rounded up using the tenths to the nearest whole number.

[b]Italicized values indicate the largest value.

Bibliometrics

SciVal

Data from SciVal for all 143 scholars revealed a median of 17 publications or 3.4 publications per year on average at the 5-year time point. Impact metrics showed a midpoint of 26.16 citations per year or an average of 130.8 citations over 5 years. KL2 scholars were cited approximately 7 times more than other researchers in their field. In addition, 44.01% (1558/3540) of the publications were in the top 10th percentile of the journals. In terms of collaboration, KL2 scholars published work with researchers from other US institutions over 2 times more than others in their research fields. Contrary to the overall trends, investigating scholars at the institution level presented some differences (Table 3). The KL2 scholars of Institution C had a significantly higher citation rate per publication than the other institutions (H_2=12.35; P=.002). Institution A had a significantly lower rate of nationally field-weighted collaboration than did the other institutions (H_2=70.49; P<.001). No other 5-year significant differences were reported.

Data from SciVal for 112 scholars showed an increase in publication rate at the 8-year time point, with scholars from all institutions publishing an average of 3.75 publications per year, up from 3.4 at the 5-year time point. The mean citation rate also slightly increased from 26.16 at the 5-year time point to 26.44 at the 8--year time point. The percentage of publications in the top 10th percentile of journals decreased by less than 1% to 43.34% (2658/6132). Rates of collaboration increased, with KL2 scholars at 8 years publishing with researchers across the nation 3.5 times more than other researchers in their field. Results at the institutional level for the 8-year time point were similar to those at the 5-year time point (Table 3). Institution C scholars had a significantly higher rate of citations per publication (H_2=10.12; P=.006), and scholars of Institution A had a significantly lower rate of field-weighted national

collaboration (H_2=65.08; *P*<.001; n=112; Institution A: n=37; Institution B: n=37; and Institution C: n=38).

Regarding demographics, outcomes from SciVal metrics reported that male scholars published significantly more work at the 5-year time point than female scholars (U=1114.50; *P*=.008; Table 4). There were no differences between the scholar race and any reported bibliometric outcomes. Scholars with an MD degree had significantly higher field-weighted citation indices than scholars without an MD degree (U=1202.50; *P*=.04).

Table 3. Postmatriculation SciVal bibliometric summary (medians reported).

Bibliometric	Value, median							
	All institutions		Institution A		Institution B		Institution C	
	5-year value (n=143)	8-year value (n=112)	5-year value (n=143)	8-year value (n=112)	5-year value (n=143)	8-year value (n=112)	5-year value (n=143)	8-year value (n=112)
FWCI[a]	7.16	12.4	6.79	10.2	6.8	12.6	*8.4*[b]	*14.2*
Scholarly output	17	30	15	27	15	26.5	*22*	*40.5*
Citations per publication	130.8	211.5	93.9	161.3	109.9	174.4	*174.1*	*288*
PPTPJ[c] (%)	44	43.3	40.9	43.2	39	41.9	*48.5*	*45.8*
NFWC[d]	2.07	3.6	0.78	65	*3.4*	*5.7*	2.9	4.9

[a]FWCI: field-weighted citation impact.

[b]Italicized values indicate the largest values.

[c]PPTPJ: percentage of publications in the top 10th percentile of journals.

[d]NFWC: National Field-Weighted Collaboration.

Table 4. SciVal bibliometric outcomes by demographic groups (medians reported).

Bibliometric	Value, median											
	Gender				Race				Degree			
	5-year value		8-year value		5-year value		8-year value		5-year value		8-year value	
	Male	Female	Male	Female	White	People of color	White	People of color	MD	Non-MD	MD	Non-MD
FWCI[a]	*8.1*[b]	6.3	*14*	10	7	*8*	11	*13*	*8*	6	*12*	9
Scholarly output	*23*	15	*35*	26	*18*	12	*30*	29	*20*	15	*35*	25
Citations per publication	*144*	115	*200*	185	125	*167*	185	*204*	*144*	109	192	*193*
PPTPJ[c]	*45*	39	*43*	41	42	*49*	40	45	*47*	39	*44*	36
NFWC[d]	*2.2*	1.8	3.6	3.6	*2.1*	1.6	*3.9*	2.2	*2.3*	1.3	*4*	2.5

[a]FWCI: field-weighted citation impact.

[b]Italicized values indicate the largest value.

[c]PPTPJ: percentage of publications in the top 10th percentile of journals.

[d]NFWC: National Field-Weighted Collaboration.

iCite

Data from iCite only provided a comprehensive report from all the years included in this study, as illustrated in Table 5. KL2 publications from all institutions were cited twice as much as other researchers in their fields, earning an average of 4.5 citations per year. The average NIH percentile was 53%, indicating that KL2 publications had an RCR higher than 53% of all NIH-funded publications [17]. At the institutional level, Institution A reported significantly lower results for the average field-weighted citation rate (H_2=8.96; *P*=.01). No other significant results were reported.

Table 5. iCite bibliometric outcomes by institution (medians reported).

Bibliometric	Value, median			
	All institutions	Institution A	Institution B	Institution C
Average RCR[a]	1.67	1.51	1.72	*1.78*[b]
Average citations per year	3.47	2.76	*3.84*	3.82
Average field citation rate	4.05	3.79	*4.27*	4.09
Average NIH[c] percentile	52.91	51.86	*54.17*	52.69

[a]RCR: relative citation ratio.

[b]Italicized values indicate the largest value.

[c]NIH: National Institutes of Health.

Federal Follow-on Funding

Grant analysis of KL2 scholars at all 3 institutions indicated that 44.1% (63/143) of KL2 scholars received federal funding within 5 years postmatriculation. At the 8-year time point, 51.7% (74/143) of scholars had achieved federal funding as a principal investigator, coprincipal investigator, or coinvestigator. Significant differences between institutions were found at the 5-year, 8-year, and overall funding rates using the chi-square test (Table 6). Scholars from Institution B were more likely to have received federal funding than scholars at Institution A or C at the 5-year (χ^2_2=15.28; P<.001) and 8-year (χ^2_2=7.07; P=.03) time points. Investigating federal grant funding by scholar characteristics showed no significant differences between gender, race, or degree at the 5-year or 8-year time points. Male scholars in our sample appeared to receive federal funding at higher rates than females, but this did not reach significance (χ^2_2=3.56; P=.06).

Table 6. Scholars with extramural federal funding by institution.

Time point	All institutions (N=143), n (%)	Institution A (n=50), n (%)	Institution B (n=48), n (%)	Institution C (n=45), n (%)
5-year	63 (44.1)	16 (32)	*31 (64.5[a])*	13 (28.9)
8-year	74 (51.7)	24 (48)	*32 (66.7)*	18 (40)
Overall	74 (51.7)	24 (48)	*32 (66.7)*	19 (42.2)

[a]Italicized values indicate the largest value.

Discussion

Principal Findings

Overall, this study emphasizes the high level of scholarly productivity, impact, collaboration, and funding achieved by KL2 scholars. This study also provides insights into the utility of both bibliometric and federal follow-on funding to evaluate CTR training success, especially for measuring the scholarly work of training participants. Bibliometric data provide a better understanding of the impact of research publications produced by KL2 scholars. Federal funding data demonstrate the extent to which KL2 scholars are receiving subsequent federal funding and therefore successfully sustaining their research.

Bibliometrics

On the basis of previous experience of applying bibliometrics to CTSA performance assessment [8-12], this study adopted a proprietary research analytics tool, SciVal, and a publicly available federal government–developed bibliometric tool, iCite, to investigate the productivity and citation impact of KL2 scholars across 3 institutions. Both SciVal and iCite metrics showed that the examined research publications of KL2 scholars had a more significant citation influence than those published simultaneously by other researchers in the same fields. For instance, SciVal metrics (eg, field-weighted citation impact [FWCI] and percentage of publications in the top 10th percentile of journals) showed that, on average, at 5 years, KL2 scholars were cited almost 7 times more than other researchers in their field. Approximately half of their articles were published in the top 10th percentile of the world's journals, which indicates the distinguishing influence of CTSA-supported KL2 scholars. Similarly, the RCRs generated by iCite disclosed that the KL2 scholars at 3 institutions received almost twice as many citations per year as other researchers in their fields, which is consistent with results reported in a previous study [12]. The difference of citation impact results produced by SciVal (FWCI) versus iCite (RCR) is associated with varied time ranges and citation tracking scopes that these two tools examined. In this study, FWCI results were generated at the 5- and 8-year time point, whereas the RCR of an institution was the average of all included publications from 2005 to 2019. In addition, the Scopus database, from which SciVal data are derived, is one of the largest citation databases in the world and covers the MEDLINE database, which is the primary component of PubMed in addition to thousands of international publishers. Consequently, the citation counts and FWCIs provided by SciVal or Scopus are often higher than those generated by iCite, which only tracks citation counts within the NIH Open Citation Collection, including MEDLINE, PubMed Central, and CrossRef [23].

Furthermore, KL2-scholars demonstrated significant collaborations with researchers across the United States, with coauthorship occurring 2 times more than others in their research

fields at the 5-year time point and 3.6 times more at the 8-year time point, as measured by the National Field-Weighted Collaboration score generated by SciVal. Therefore, our results confirm the feasibility of applying bibliometrics to assess scholarly work supported by CTSA programs and corroborate the effectiveness of three CTSA programs in supporting KL2 scholars for translational research [8,9].

Nevertheless, a subanalysis highlighted a few critical demographic differences between scholars across institutions. Male scholars published significantly more than female scholars did at 5 years postmatriculation; however, this difference was not observed at the 8-year time point. Previous research confirms the difference in publications by males, but limited research has investigated any long-term differences [24]. Our results suggest that differences over time may be a critical point where female scholars reduce this gap, which may arise from a variety of personal commitments that impact the pace of the career of a female scientist. At 8 years postmatriculation, scholars with an MD degree had a significantly higher FWCI than those without an MD degree. These demographic differences should be further studied to better understand the role the scientific field plays in supporting scholarly work across different scientific areas.

Similarly, bibliometrics brought institutional differences to our attention (Multimedia Appendix 1). We found that KL2 scholars at Institution C had a significantly higher rate of citations per publication. In comparison, Institution A had a significantly lower rate of National Field-Weighted Collaboration than the other 2 institutions. These cross-institutional differences emphasize the need to invest in the standardization of measures (eg, bibliometrics), improving the ability to evaluate scholarly success across the nation. In previous research, CTSA hubs reported various techniques for reporting and tracking publications [8,9,25]. In addition, differences in publication characteristics have been shown to impact bibliometric outcomes (eg, reviews are usually cited more than original articles; open-accessible articles are cited more than nonopen accessible ones), such as in journal coverage of bibliographic databases, and the size and establishing time of CTSA programs [9,10,25]. Therefore, although it is feasible to use bibliometrics to analyze the scholarly output and influence of multiple CTSA programs, there are complications in applying citation metrics, interpreting results, and benchmarking the performance of multi-institutional programs. Future research should consider the role of the program-level characteristics outlined in the Multimedia Appendix 1.

Federal Follow-on Funding

Our analysis reveals that federal funding award rates for KL2 scholars are higher than the national average of 20.2% [26]. Previous research reported a 34% R01 funding rate for KL2 scholars 6 years postmatriculation [27]. Our study included R01 equivalent awards, but nonetheless suggests that the federal funding rate for KL2 scholars in this sample (63/143, 44.1% at 5 years) may be slightly higher than other institutions. The data also show that male KL2 scholars are more likely to receive federal funding than their female counterparts; however, this was not statistically significant. Previous studies examining gender differences in obtaining grant funding have been mixed. There were mitigating effects by both the type of training and the length of time following training [21,22,28]. In our analysis, we investigated the attainment of any NIH funding, and thus, we might have picked up additional funding mechanisms not examined in previous research. To our knowledge, none of the studies, including ours, has examined funding from foundations or other sources. This exclusion is a limitation and potential area of further research, especially to better understand the underlying factors associated with male KL2 scholars having a higher likelihood of receiving follow-on funding.

Limitations

The limitations of this study include programmatic differences between the 3 institutions, differences in scholarly scientific fields, and grant data. We intended to investigate the scholarly success of a multi-institutional sample of KL2 scholars and to identify appropriate and feasible methods that could be used across institutions. However, the analysis of these measures revealed that performance on certain metrics could be linked to institutional characteristics. Perhaps the most considerable difference between these three KL2 programs is the length of KL2 funded training time. Each institution had a different length of funding, ranging from 2 to 4 years, provided to scholars. Other programmatic distinctions, such as variations in grant writing support, also exist. Future research should identify which program characteristics are related to the outcomes studied here. Other possible analyses include investigating the impact of scholarly field and degree on outcomes given the differences between MDs and PhDs that we highlighted in our analyses.

Conclusions

This study emphasized the use of bibliometrics and federal follow-on funding as necessary evaluation measures for assessing scholarly productivity, impact, collaboration, and funding achieved by KL2 scholars. We have shown that evaluators can use these metrics to evaluate CTR training programs that focus on scholarly productivity as a critical outcome. These methods can be used to complement existing evaluation strategies to demonstrate program performance. The findings of this study highlight the need to better understand barriers and facilitators of scholarly productivity. Institutions should consider similar subanalyses within their evaluations to explore equity within their programs. In addition, there is a need to investigate the impact of programmatic components and best practices that yield high follow-on funding rates. Program-level goals within and among institutions influence funding outcomes, scholarly productivity, and collaboration. Identifying these differences will enhance the specificity of KL2 program evaluations. CTR training programs, such as CTSA KL2 programs, can benefit from the findings of this and future analysis as they continuously adapt their program strategy to support the early-stage career development of CTR investigators.

Acknowledgments

This project was supported by the Clinical and Translational Science Collaborative at Case Western Reserve University, the North Carolina Translational and Clinical Sciences Institute at the University of North Carolina—Chapel Hill, and the Center for Clinical and Translational Science at the Mayo Clinic, which is funded by the National Institutes of Health National Center for Advancing Translational Science Clinical and Translational Science Award (UL1TR002548; UL1TR002489; KL2 TR002379). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1
KL2 program comparison across 3 Clinical and Translational Science Award hubs.
[XLSX File (Microsoft Excel File), 75 KB - jmir_v23i9e29239_app1.xlsx]

References

1. Strategic goal 3: develop and foster innovative translational training and a highly skilled, creative and diverse translational science workforce internet. National Center for Advancing Translational Sciences. URL: https://ncats.nih.gov/strategicplan/goal3 [accessed 2021-08-18]
2. Common metrics initiative. Center for Leading Innovation & Collaboration. URL: https://clic-ctsa.org/common-metrics-initiative/ [accessed 2021-08-18]
3. Clinical and translational science award (U54 Clinical Trial Optional). National Institutes of Health. URL: https://grants.nih.gov/grants/guide/pa-files/par-18-940.html [accessed 2021-08-18]
4. Calvin-Naylor NA, Jones CT, Wartak MM, Blackwell K, Davis JM, Divecha R, et al. Education and training of clinical and translational study investigators and research coordinators: a competency-based approach. J Clin Transl Sci 2017 Feb;1(1):16-25 [FREE Full text] [doi: 10.1017/cts.2016.2] [Medline: 28480054]
5. Sorkness CA, Scholl L, Fair AM, Umans JG. KL2 mentored career development programs at clinical and translational science award hubs: practices and outcomes. J Clin Transl Sci 2019 Dec 26;4(1):43-52 [FREE Full text] [doi: 10.1017/cts.2019.424] [Medline: 32257410]
6. Pritchard A. Statistical bibliography or bibliometrics? J Doc 1969 Jan;25(4):348-349.
7. Ellegaard O, Wallin JA. The bibliometric analysis of scholarly production: how great is the impact? Scientometrics 2015;105(3):1809-1831. [doi: 10.1007/s11192-015-1645-z] [Medline: 26594073]
8. Llewellyn N, Carter DR, DiazGranados D, Pelfrey C, Rollins L, Nehl EJ. Scope, influence, and interdisciplinary collaboration: the publication portfolio of the NIH Clinical and Translational Science Awards (CTSA) program from 2006 through 2017. Eval Health Prof 2020 Sep;43(3):169-179. [doi: 10.1177/0163278719839435] [Medline: 30917690]
9. Schneider M, Kane CM, Rainwater J, Guerrero L, Tong G, Desai SR, et al. Feasibility of common bibliometrics in evaluating translational science. J Clin Transl Sci 2017 Feb;1(1):45-52 [FREE Full text] [doi: 10.1017/cts.2016.8] [Medline: 28480055]
10. Yu F, Van AA, Patel T, Mani N, Carnegie A, Corbie-Smith GM, et al. Bibliometrics approach to evaluating the research impact of CTSAs: a pilot study. J Clin Transl Sci 2020 Apr 02;4(4):336-344 [FREE Full text] [doi: 10.1017/cts.2020.29] [Medline: 33244415]
11. Amory JK, Louden DK, McKinney C, Rich J, Long-Genovese S, Disis ML. Scholarly productivity and professional advancement of junior researchers receiving KL2, K23, or K08 awards at a large public research institution. J Clin Transl Sci 2017 Apr;1(2):140-143 [FREE Full text] [doi: 10.1017/cts.2016.22] [Medline: 28649457]
12. Sayavedra N, Hogle JA, Moberg DP. Using publication data to evaluate a Clinical and Translational Science Award (CTSA) career development program: early outcomes from KL2 scholars. J Clin Transl Sci 2017 Dec;1(6):352-360 [FREE Full text] [doi: 10.1017/cts.2018.1] [Medline: 29707257]
13. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. FASEB J 2008 Feb;22(2):338-342. [doi: 10.1096/fj.07-9492LSF] [Medline: 17884971]
14. Pranckutė R. Web of Science (WoS) and Scopus: the titans of bibliographic information in today's academic world. Publications 2021 Mar 12;9(1):12. [doi: 10.3390/publications9010012]
15. Luke DA, Carothers BJ, Dhand A, Bell RA, Moreland-Russell S, Sarli CC, et al. Breaking down silos: mapping growth of cross-disciplinary collaboration in a translational science initiative. Clin Transl Sci 2015 Apr;8(2):143-149 [FREE Full text] [doi: 10.1111/cts.12248] [Medline: 25472908]
16. Dresbeck R. SciVal. J Med Libr Assoc 2015 Jul;103(3):164-166. [doi: 10.3163/1536-5050.103.3.018]
17. New analysis. National Institutes of Health. URL: https://icite.od.nih.gov/analysis [accessed 2021-08-18]

18. Hutchins B, Yuan X, Anderson JM, Santangelo GM. Relative Citation Ratio (RCR): a new metric that uses citation rates to measure influence at the article level. PLoS Biol 2016 Sep 6;14(9):e1002541 [FREE Full text] [doi: 10.1371/journal.pbio.1002541] [Medline: 27599104]
19. RePORTER. National Institutes of Health. URL: https://reporter.nih.gov/ [accessed 2021-08-18]
20. IBM SPSS statistics. IBM. URL: https://www.ibm.com/products/spss-statistics [accessed 2021-08-18]
21. Jagsi R, Motomura AR, Griffith KA, Rangarajan S, Ubel PA. Sex differences in attainment of independent funding by career development awardees. Ann Intern Med 2009 Dec 01;151(11):804-811. [doi: 10.7326/0003-4819-151-11-200912010-00009] [Medline: 19949146]
22. Kalyani RR, Yeh H, Clark JM, Weisfeldt ML, Choi T, MacDonald SM. Sex differences among career development awardees in the attainment of independent research funding in a department of medicine. J Womens Health (Larchmt) 2015 Nov;24(11):933-939 [FREE Full text] [doi: 10.1089/jwh.2015.5331] [Medline: 26291588]
23. Hutchins BI, Baker KL, Davis MT, Diwersy MA, Haque E, Harriman RM, et al. The NIH open citation collection: a public access, broad coverage resource. PLoS Biol 2019 Oct 10;17(10):e3000385 [FREE Full text] [doi: 10.1371/journal.pbio.3000385] [Medline: 31600197]
24. Posen M, Templer DI, Forward V, Stokes S, Stephens J. Publication rates of male and female academic clinical psychologists in California. Psychol Rep 2005 Dec;97(3):898-902. [doi: 10.2466/pr0.97.3.898-902] [Medline: 16512310]
25. Llewellyn N, Carter DR, Rollins L, Nehl EJ. Charting the publication and citation impact of the NIH Clinical and Translational Science Awards (CTSA) Program From 2006 Through 2016. Acad Med 2018 Aug;93(8):1162-1170 [FREE Full text] [doi: 10.1097/ACM.0000000000002119] [Medline: 29298181]
26. Research project success rates by NIH Institute for 2018. NIH Research Portfolio Online Reporting Tools (RePORT). URL: https://report.nih.gov/success_rates/success_byic.cfm [accessed 2021-08-24]
27. Schneider M, Guerrero L, Jones LB, Tong G, Ireland C, Dumbauld J, et al. Developing the Translational Research Workforce: A Pilot Study of Common Metrics for Evaluating the Clinical and Translational Award KL2 Program. Clin Transl Sci 2015 Dec;8(6):662-667 [FREE Full text] [doi: 10.1111/cts.12353] [Medline: 26602332]
28. Pohlhaus JR, Jiang H, Wagner RM, Schaffer WT, Pinn VW. Sex differences in application, success, and funding rates for NIH extramural programs. Acad Med 2011 Jun;86(6):759-767 [FREE Full text] [doi: 10.1097/ACM.0b013e31821836ff] [Medline: 21512358]

Abbreviations

CTR: clinical and translational research
CTSA: Clinical and Translational Science Award
FWCI: field-weighted citation impact
NCATS: National Center for Advancing Translational Science
NIH: National Institutes of Health
RCR: relative citation ratio
RePORTER: Research Portfolio Online Reporting Tools Expenditures and Results

Edited by R Kukafka; submitted 30.03.21; peer-reviewed by N Llewellyn, PKH Mo, J Du; comments to author 10.05.21; revised version received 04.07.21; accepted 27.07.21; published 29.09.21.

Please cite as:
Qua K, Yu F, Patel T, Dave G, Cornelius K, Pelfrey CM
Scholarly Productivity Evaluation of KL2 Scholars Using Bibliometrics and Federal Follow-on Funding: Cross-Institution Study
J Med Internet Res 2021;23(9):e29239
URL: https://www.jmir.org/2021/9/e29239
doi:10.2196/29239
PMID:34586077

Original Paper

Effectiveness of a Multistrategy Behavioral Intervention to Increase the Nutritional Quality of Primary School Students' Web-Based Canteen Lunch Orders (Click & Crunch): Cluster Randomized Controlled Trial

Rebecca Wyse[1], PhD; Tessa Delaney[2], BND; Fiona Stacey[2], PhD; Rachel Zoetemeyer[2], BND; Christophe Lecathelinais[2], DESS; Hannah Lamont[2], BND; Kylie Ball[3], PhD; Karen Campbell[3], PhD; Chris Rissel[4], PhD; John Attia[2], PhD; John Wiggers[2], PhD; Sze Lin Yoong[2], PhD; Christopher Oldmeadow[2], PhD; Rachel Sutherland[2], PhD; Nicole Nathan[2], PhD; Kathryn Reilly[2], PhD; Luke Wolfenden[2], PhD

[1]School of Medicine and Public Health, University of Newcastle, Callaghan, Australia

[2]Hunter New England Population Health, Wallsend, New South Wales, Australia

[3]School of Exercise and Nutrition Sciences, Deakin University, Burwood, Victoria, Australia

[4]School of Public Health, University of Sydney, Sydney, Australia

Corresponding Author:
Rebecca Wyse, PhD
School of Medicine and Public Health
University of Newcastle
University Drive
Callaghan, NSW 2308
Australia
Phone: 61 49246310
Email: rebecca.wyse@health.nsw.gov.au

Abstract

Background: School food outlets represent a key setting for public health nutrition intervention. The recent proliferation of web-based food ordering systems provides a unique opportunity to support healthy purchasing from schools. Embedding evidence-based choice architecture strategies within these routinely used systems provides the opportunity to impact the purchasing decisions of many users simultaneously and warrants investigation.

Objective: This study aims to assess the effectiveness of a multistrategy behavioral intervention implemented via a web-based school canteen lunch ordering system in reducing the energy, saturated fat, sugar, and sodium content of primary students' web-based lunch orders.

Methods: The study used a parallel-group, cohort, cluster randomized controlled trial design with 2207 students from 17 Australian primary schools. Schools with a web-based canteen lunch ordering system were randomly assigned to receive either a multistrategy behavioral intervention that included choice architecture strategies embedded in the web-based system (n=9 schools) or the standard web-based ordering system only (n=8 control schools). Automatically collected student purchasing data at baseline (term 2, 2018) and 12 months later (term 2, 2019) were used to assess trial outcomes. Primary trial outcomes included the mean energy (kJ), saturated fat (g), sugar (g), and sodium (mg) content of student lunch orders. Secondary outcomes included the proportion of all web-based lunch order items classified as *everyday*, *occasional*, and *caution* (based on the New South Wales Healthy School Canteen Strategy) and canteen revenue.

Results: From baseline to follow-up, the intervention lunch orders had significantly lower energy content (−69.4 kJ, 95% CI −119.6 to −19.1; *P*=.01) and saturated fat content (−0.6 g, 95% CI −0.9 to −0.4; *P*<.001) than the control lunch orders, but they did not have significantly lower sugar or sodium content. There was also a small significant between-group difference in the percentage of energy from saturated fat (−0.9%, 95% CI −1.4% to −0.5%; *P*<.001) but not in the percentage of energy from sugar (+1.1%, 95% CI 0.2% to 1.9%; *P*=.02). Relative to control schools, intervention schools had significantly greater odds of having *everyday* items purchased (odds ratio [OR] 1.7, 95% CI 1.5-2.0; *P*<.001), corresponding to a 9.8% increase in *everyday* items, and lower odds of having *occasional* items purchased (OR 0.7, 95% CI 0.6-0.8; *P*<.001), corresponding to a 7.7% decrease in

occasional items); however, there was no change in the odds of having *caution* (least healthy) items purchased (OR 0.8, 95% CI 0.7-1.0; *P*=.05). Furthermore, there was no change in schools' revenue between groups.

Conclusions: Given the evidence of small statistically significant improvements in the energy and saturated fat content, acceptability, and wide reach, this intervention has the potential to influence dietary choices at a population level, and further research is warranted to determine its impact when implemented at scale.

Trial Registration: Australian New Zealand Clinical Trials Registry (ANZCTR) ACTRN12618000855224; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=375075.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2019-030538

(J Med Internet Res 2021;23(9):e26054) doi:10.2196/26054

KEYWORDS

nudge; choice architecture; intervention; online canteen; online ordering systems; digital interventions; school children; school food service; canteens; menu labeling

Introduction

Background

Dietary risk factors are a leading cause of disease [1]. Schools are a key setting for public health nutrition [2], and school food outlets can play an important role in children's nutrition. Studies of Australian school canteens show energy-dense, nutrient-poor foods are the most commonly purchased items [3,4], despite the existence of interventions to encourage healthier choices [5-7]. Such interventions typically include supply-based approaches that target the relative availability of healthier foods. For example, the New South Wales (NSW) Healthy School Canteen Strategy (the NSW Strategy) requires that canteen menus have at least 75% *everyday* items (healthy foods consistent with the Australian Dietary Guidelines) and no *should not be sold* items (unhealthy items high in saturated fat, sugars, or salt) [8]. The NSW Strategy is mandatory in government schools and is strongly encouraged in nongovernment schools [8]. However, the impact of such supply-based interventions is highly variable due to less than optimal implementation [5-7,9,10].

The rise of web-based technology represents a unique opportunity to provide broader support to consumers to encourage healthy canteen purchases in schools using demand-based or consumer-focused approaches. In particular, choice architecture strategies [11], including menu labeling [12-16], prompts [17-19], and changing the position of food [20], have been shown to influence food choices in schools and other settings. Embedding these strategies within routinely used systems is appealing from a public health perspective as they provide the opportunity to reach many people simultaneously at low cost, require only minimal engagement, and are not reliant on the education or skills of the consumers [11].

We previously evaluated a choice architecture intervention delivered via an existing, routinely used web-based canteen ordering system in Australian schools. The pilot study, conducted in 10 NSW government primary schools, established that choice architecture strategies could be successfully embedded within a web-based canteen ordering system [21]. Web-based ordering allows parents and carers (hereafter *parents*) and students to view the school canteen menu and prepurchase items on the web. This cluster randomized controlled pilot trial found that a 2-month intervention, which incorporated menu labeling, positioning, prompting, and availability strategies, significantly increased healthy food purchases [21]. At follow-up, the lunch orders from the intervention schools contained less energy, saturated fat, and sodium than lunch orders from control schools (*P*<.001) [21]. However, the pilot study tested only a limited range of strategies within a small number of government schools for a short period. Furthermore, since the pilot, the NSW canteen guidelines have changed [22]. In 2017, the Fresh Tastes at School Strategy [22], which adopted a traffic light system to classify foods as *green*, *amber*, or *red* from most to least healthy, was replaced by the current NSW Healthy School Canteen Strategy [8], which was revised to align with the Australian Dietary Guidelines. The current NSW Strategy uses the national Health Star Rating front-of-pack labeling system and specific portion limits to classify foods as *everyday*, *occasional*, or *should not be sold* (*caution*). Schools were provided with support to help implement the guidelines, with the goal of complete implementation in NSW schools by 2019.

Objectives

This study seeks to extend the impact of the pilot by testing additional strategies in a changed policy context with schools from additional sectors (ie, Catholic and independent schools) for a longer period (approximately 12 months) to determine the broader utility of this novel approach in improving public health nutrition. This trial aims to assess the effectiveness of a multistrategy behavioral intervention embedded within an existing web-based canteen ordering system in reducing the energy, saturated fat, sugar, and sodium content of primary school students' web-based lunch orders.

Methods

Study Design

This study used a parallel-group, cohort, cluster randomized controlled trial (RCT) design and is reported according to the CONSORT (Consolidated Standards of Reporting Trials) extension for clustered RCTs. This trial was prospectively registered (ACTRN12618000855224) and approved by the University of Newcastle Human Research Ethics Committee (H-2017-0402) and the Catholic Education Office Dioceses of

Sydney, Parramatta, Lismore, Maitland-Newcastle, Bathurst, Canberra-Goulburn, Wagga Wagga, Wollongong, and Wilcannia-Forbes. Schools (clusters) with a web-based canteen ordering system were randomly assigned to either an intervention group (receiving choice architecture strategies embedded in the web-based systems, plus audit and feedback) or a control group (receiving the standard web-based canteen only), as it was not possible to randomize individuals. A detailed protocol has been published [23], which is summarized below.

Participant Recruitment

Participating Schools

Nongovernment (Catholic and independent) primary schools within NSW, Australia, were approached to participate by mail and telephone. Although it was originally intended that government schools would be included, extensive delays in obtaining ethical approval meant that these schools were excluded, as the timeframe of intervention exposure would have been too short to warrant inclusion. Although the NSW Healthy School Canteen Strategy is mandatory in government schools, all schools are strongly encouraged to adopt this strategy [8].

Recruitment took place from May to September 2018 (17 weeks).

Inclusion and Exclusion Criteria

Textbox 1 describes the inclusion and exclusion criteria.

Textbox 1. School, user, and order inclusion and exclusion criteria.

Inclusion Criteria

- Schools: schools were eligible if they used the Flexischools web-based canteen service and had done so for at least a month before recruitment. Flexischools is the largest provider of web-based canteen services in Australia, servicing more than 1200 schools and processing more than 13 million lunch orders per year [21].
- Users: students from kindergarten to grade 5 who placed a web-based lunch order in the baseline data collection period were included in the study.
- Orders: all new orders placed on a mobile device were included (typically representing approximately 70% of all orders). Orders placed on a desktop or where order modality could not be determined were excluded as users would not have been exposed to any of the intervention strategies.

Exclusion Criteria

- Schools
  - Externally licensed school canteens were excluded as these private operators often service multiple schools, increasing the risk of intervention contamination. However, two recruited canteens that were not initially identified as being externally licensed were retained, as they did not service any other school within the sample.
  - Combined primary and secondary schools were ineligible unless they had separate menus for primary and secondary students, as the New South Wales Strategy is applied differently across these age groups.
- Users: grade 6 and grade 5 and 6 composite students were excluded from the baseline as they would have moved on to secondary school during the follow-up period. School staff members were excluded.
- Orders: the baseline and follow-up periods were intended to run for the 10-week school term. However, orders from 2 weeks of the baseline term were excluded as the order mode (mobile vs desktop) could not be determined because of a software update. The follow-up data collection period was subsequently reduced to 8 weeks to match the baseline. Recurring orders placed before the intervention period and orders from special food days were excluded, as users would not have been exposed to the intervention strategies. Lunch orders with an implausible number of items (ie, 15 or more) were excluded based on consensus from research dietitians with extensive canteen experience.

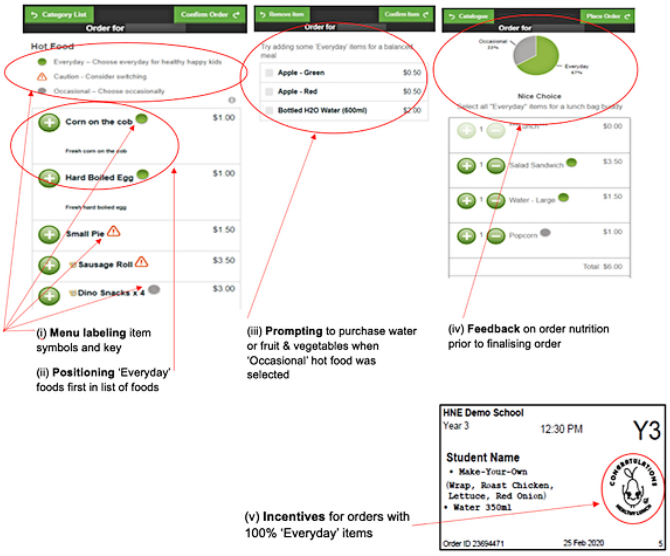
Intervention

Overview

A multistrategy behavioral intervention targeting users of the web-based system was implemented by modifying the interface of the Flexischools web-based ordering system. The intervention (Figure 1) is fully described in the published protocol [23] and incorporated choice architecture strategies [11], including the following:

- Menu labeling: a colored symbol denoting *everyday*, *occasional*, or *caution* (also referred to as *should not be sold*) items based on the NSW Healthy Canteen Strategy was added next to each menu item along with a key explaining each of the symbols.
- Positioning: menu items and categories were arranged to make healthier *everyday* options more prominent (ie, positioned first within the list of menu items), and where there were multiple flavors of an *occasional* or *caution* item, users were required to *click* on the category before the full list of flavors was displayed in a separate pop-up box.
- Prompting: an appealing image and brief text prompt were placed next to all healthy categories. When users selected an *occasional* or *caution* hot food item, they received a prompt to purchase water and a piece of fruit or vegetable (*healthy add-ons*).
- Feedback: before finalizing and paying for their lunch order, the user was shown a pie graph with tailored feedback based on the proportion of *everyday* items in the order.
- Incentives: orders that contained 100% *everyday* items had a cartoon character and congratulatory text printed on the lunch label, which was printed out and stuck on the paper bag in which the lunch order was delivered to students.

Figure 1. Screenshot of the Click & Crunch intervention.



Canteen Supportive Strategies

An audit and feedback report were emailed to canteen managers and principals, classifying each menu item as per the NSW Strategy (ie, *everyday* or *occasional*) and providing general information about pricing items to encourage healthy purchases [6].

Control

Control schools received no change to their web-based canteen menu and no audit or feedback reports. However, NSW Local Health Districts support all schools to use the Menu Check Service—a free service that reviews canteen menus against the NSW Strategy and provides feedback to schools.

Outcomes and Measures

Overview

Student purchasing data automatically collected by the Flexischools system was used as the basis for evaluation. Data were collected simultaneously for all schools, for two 8-week periods, 12 months apart (term 2, 2018, and term 2, 2019). Baseline data were retrospectively collected.

Primary Outcomes

The primary outcomes were the mean lunch order content of energy (kg), saturated fat (g), sugar (g), and sodium (mg). For prepackaged foods, this was based on a blinded dietitian's assessment using a series of sources in the following order: (1) a database containing more than 2000 commonly stocked canteen products developed by canteen researchers for the past 5 years, (2) the *Healthy Food Finder* database (NSW Government) [24], (3) the *FoodSwitch* website (The George Institute for Global Health) [25], and (4) a web search for the nutrient information panel. For canteen-prepared food, the recipe was obtained from the canteen manager and analyzed using *FoodWorks 9 Professional* nutrition analysis software (Xyris Software) [26].

Secondary Outcomes

Healthier Purchasing

The proportion of all web-based lunch order items that were *everyday*, *occasional*, and *caution* was calculated by a dietitian using the criteria underpinning the NSW Strategy; the mean proportion of energy within lunch orders that was derived from saturated fat and sugar was also calculated based on 37 kJ/g of fat and 17 kJ/g of sugar [27].

Canteen Revenue (Adverse Outcome)

Purchasing data that were automatically collected by the web-based system were used to calculate mean weekly canteen revenue.

Canteen Characteristics

At the end of the follow-up period, a telephone interview with canteen managers was conducted to collect information regarding canteen operations (eg, principal operated; parent and citizen operated; or externally licensed—privately operated), type of manager (eg, paid or volunteer manager), the days of operation, and the usual number of web-based orders per week.

Process Measures

Change in Availability

The proportion of schools with canteen menus meeting the NSW Strategy was calculated at baseline and follow-up based on dietitian menu assessment.

Change in Pricing

The average price of *everyday*, *occasional*, and *caution* items was calculated at baseline and follow-up based on each school's web-based menu.

Intervention Acceptability

During the telephone interview, canteen managers were asked to rate the acceptability of the intervention (eg, individual strategies and the intervention overall) using a Likert scale (Multimedia Appendix 1).

Additional Support Received

To determine the potential impact of canteen managers making changes to their menu, they were also asked about any additional support they had received during the 12-month intervention period.

School Characteristics

School characteristics were obtained from a national website [28]. The following information was extracted: the number of student enrolments, the proportion of Aboriginal and Torres Strait Islander enrolment, and the school postcode.

User Characteristics

User characteristics were derived from automatically collected data within the Flexischools system. *Student class* (eg, 3F) contains student grade (eg, Grade 3) and is a required field entered by the parents at system registration, and the frequency of use was based on a count of all orders placed by a student within the baseline data collection period and an average per week was calculated.

Randomization and Blinding

Following the provision of principal (cluster) consent, and after the completion of the baseline menu assessment, an independent statistician used Microsoft Excel to randomize schools to an intervention or control group in a 1:1 ratio, using block randomization with block size between 2 and 4. Randomization was stratified by school sector (ie, independent schools and Catholic schools) and socioeconomic status based on school postcode using Socioeconomic Indexes for Areas (SEIFA) [29]. Due to the difficulty in blinding participants, this was run as an open trial.

Fidelity and Data Quality

During each school term for the duration of the intervention, a research assistant checked the web-based menus to ensure that the menu labels were correctly applied and recorded whether other intervention strategies were present or absent. An exception was the incentive strategy, which, being printed on lunch labels, was not verifiable by checking the web-based menu. Canteen visits were conducted in 6 schools (including 3 intervention schools) to verify the automatically collected purchasing data against the orders received and distributed by the canteens.

Sample Size

Recruitment of 26 schools and 194 students per school (allowing for 86% follow-up and 70% of orders placed using a mobile device) would allow detection of a between-group difference of 195 kJ per average lunch order, assuming an intraclass correlation coefficient of 0.05, with 80% power, and a significance level of $P<.0125$ (Holm-Bonferroni adjusted for the four primary outcomes).

Analysis

An intention-to-treat approach was adopted, whereby all student orders and schools were analyzed based on the groups to which they were originally allocated and included data from all students who had baseline purchasing data. To adjust for multiple outcomes, $P<.0125$ was adopted as the prespecified level of significance.

Primary Outcomes

Each primary trial outcome was assessed using a separate linear mixed model. The nutritional content (ie, energy, saturated fat, sugar, and sodium) of all web-based lunch orders placed by students was compared between the intervention and control groups throughout time by including a group-by-time interaction fixed effect. All models included a random intercept for school (to account for potential school-level clustering), a nested random intercept and random time effect for students (to account for repeated measurements between and within baseline and follow-up), and fixed effects for the school sector and SEIFA.

Secondary Outcomes

Healthier purchasing outcomes (ie, *everyday*, *occasional*, and *caution*) were assessed using separate logistic mixed models. To assess whether there was a significant change in the purchase of *everyday*, *occasional*, and *caution* items between groups, three separate logistic regressions were used (ie, items that are *everyday* vs items that are not *everyday*), including a group-by-time interaction fixed effect. As with the primary outcomes, all models included a random intercept for school (to account for potential school-level clustering), a nested random intercept and random time effect for students (to account for repeated measurements between and within time points), and fixed effects for the school sector and SEIFA. Differences in the proportion of energy derived from saturated fat and sugar within each order and differences in average weekly revenue were assessed according to the primary outcomes.

Per-Protocol Analysis

A per-protocol analysis was conducted to determine the effect on energy content (kJ) and proportion of *everyday* foods when the intervention was applied in full. Schools were included if they had >80% of verifiable strategies correctly applied at follow-up and if the incentive strategy was reported as present in the canteen manager survey. Prespecified subgroup analyses were conducted based on energy content (kJ), student grade (kindergarten-grade 2 vs grade 3-5), school sector (Catholic vs independent), and order frequency (*low* <1 order/week on average vs *high* ≥1 order/week on average) by adding a three-way interaction fixed effect (group-by-time-by-subgroup). The following assumptions underpin the prespecified subgroup analyses: (1) parents may have more control over the lunch orders of younger students and may be more influenced by the intervention, leading to healthier purchasing for younger students. (2) There may be differences in the implementation of the NSW Healthy Schools Canteen Strategy between the school sectors (Catholic vs independent), which may have influenced the menu composition, and therefore the ability of the intervention to have an effect. (3) Users who ordered less frequently may consider canteen lunch orders to be more of a *treat* purchase rather than part of their usual diet and therefore be less influenced by the intervention strategies.

Statistical analyses were performed using the SAS version 9.3 (SAS Institute).

Results

Overview

The CONSORT diagram (Figure 2) shows the number of schools and students participating in the trial. After 17 weeks of recruitment, 40% (17/43) of eligible schools had consented, 33% (14/43) had refused, and 28% (12/43) were undecided. In total, 9 schools were randomized to the intervention group, and 8 schools were randomized to the control group. None of the schools dropped out of the study. Four *combined* schools that enrolled kindergarten to grade 12 students were included in the sample (3 intervention schools and 1 control school).

Only four orders, representing <0.01% of all orders, were excluded because they were implausibly large. The characteristics of participating schools and students are described in Tables 1 and 2. Intervention schools had approximately 30% higher student enrolments than control schools (no significance testing) [30], and as such, the average number of web-based lunch orders per week was higher in intervention schools.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials; extension for clustered randomized controlled trials) diagram.

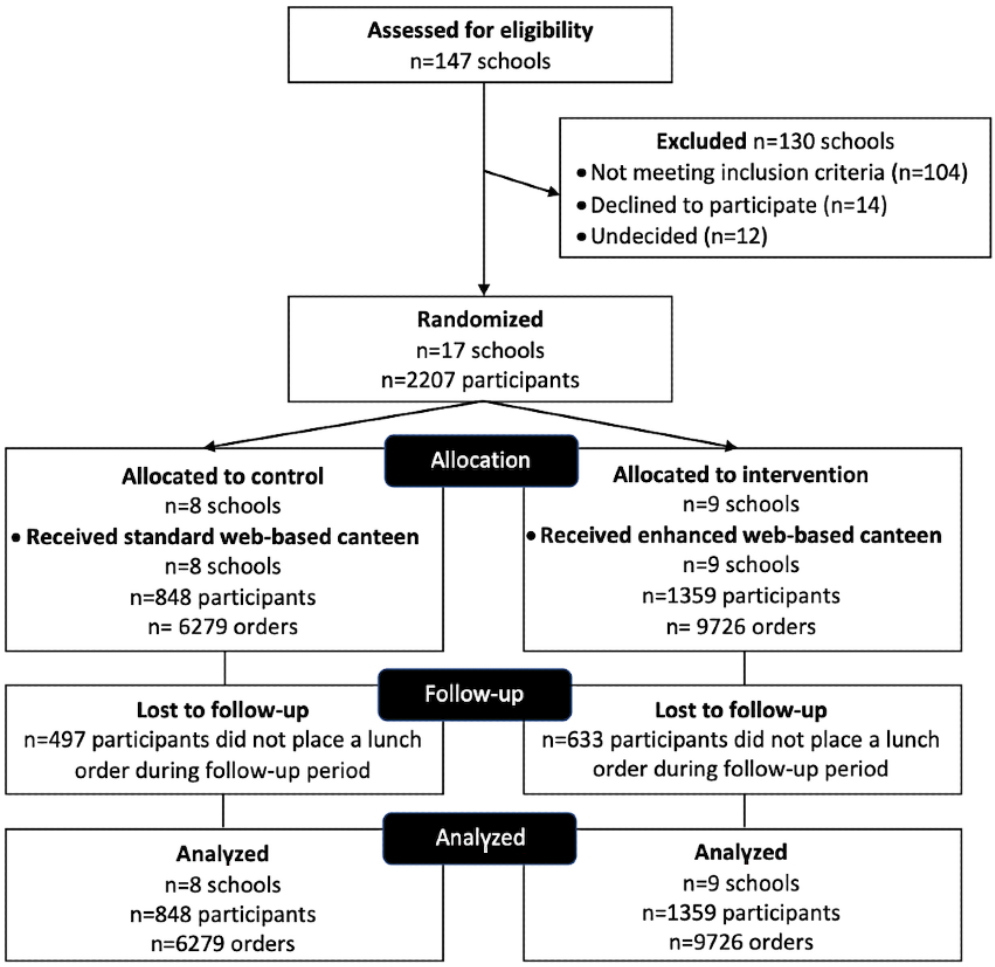


Table 1. Characteristics of the sample of 17 participating schools.

School and canteen characteristics	Intervention (n=9 schools)	Control (n=8 schools)
School sector[a], n (%)		
Independent	4 (44)	3 (38)
Catholic	5 (56)	5 (63)
Number of enrolments[a,b], mean (SD)	501 (208)	386 (134)
Aboriginal or Torres Strait Islander students[a] (%), mean (SD)	6 (9.7)	4 (3.5)
Socioeconomic status[c], n (%)		
Least advantaged	3 (33)	4 (50)
Most advantaged	6 (67)	4 (50)
Type of canteen operation[d], n (%)		
Principal or school run	6 (86)	6 (75)
Parent & Citizen or Parents & Friends run[e]	0 (0)	1 (13)
Contracted food service	1 (14)	1 (13)
Type of canteen manager[d], n (%)		
Paid	7 (100)	8 (100)
Volunteer	0 (0)	0 (0)
Days of canteen operation per week[d], n (%)		
5	6 (86)	5 (63)
3-4	1 (14)	2 (25)
1-2	0 (0)	1 (13)
Number of weekly web-based lunch orders (per school)[f], mean (SD)	136 (80.3)	98 (91.3)

[a]On the basis of publicly available school statistics (MySchool 2018).

[b]Excluding combined schools (as this information was not available on the MySchool website).

[c]Socio-economic Indexes for Areas 2016 data, based on the postcode of school locality, and dichotomized (median split).

[d]On the basis of the canteen manager survey conducted after collecting follow-up data, completed by 7 intervention and 8 control canteen managers.

[e]Parent & Citizen or Parents & Friends (in Catholic schools) run canteens are those that are managed by a governing body or committee consisting of parents and citizens of the school community and the school principal.

[f]On the basis of Flexischools purchasing data.

Table 2. Characteristics of the sample of 2207 participating students.

User characteristics	Intervention (n=1359)	Control (n=848)
Grade of student at baseline, n (%)		
Kindergarten-grade 2	677 (49.82)	446 (52.59)
Grade 3-5	682 (50.18)	402 (47.41)
Frequency of use[a], n (%)		
High users (≥1 order/week on average)[b]	463 (34.07)	312 (36.79)
Low users (<1 order/week on average)	896 (65.93)	536 (63.21)

[a]Frequency of use based on baseline purchasing characteristics.

[b]Orders ≥8 during the 8-week baseline data collection period.

Primary Outcomes

Overview

The linear mixed model analysis indicated that the between-group differences throughout time for the intervention group were as follows: −69.4 kJ energy (95% CI −119.6 to −19.1; *P*=.01), −0.6 g saturated fat (95% CI −0.9 to −0.4; *P*<.001), −32.1 mg sodium (95% CI −56.3 to −7.9; *P*=.013), and +0.4 g sugar (95% CI −0.7 to 1.5; *P*=.47). The differences in energy and saturated fat were statistically significant at the prespecified level of *P*<.0125, and sodium was borderline (Table 3).

Table 3. Primary and secondary outcomes in intervention and control groups from baseline to follow-up for 2207 student participants (linear mixed model analysis).

Variable	Baseline, mean (SD)		Follow-up, mean (SD)		Intervention versus control[a]			
	Intervention (N=1359 children; N=9726 orders; N=23,526 items)	Control (N=848 children; N=6279 orders; N=14,124 items)	Intervention (N=1108 children; N=9434 orders; N=22,061 items)	Control (N=691 children; N=6334 orders; N=14,087 items)	Main analysis		Per-protocol analysis	
					Group-by-time differential effect (95% CI)	*P* value	Group-by-time differential effect (95% CI)	*P* value
Primary outcomes[b]								
Energy (kJ)[b]	1634.4 (704.2)	1632.1 (743.0)	1623.3 (699.2)	1685.6 (838.6)	−69.4 (−119.6 to −19.1)	*.01*[c]	−89.4 (−148.9 to −29.9)	*.007*
Saturated fat (g)[b]	5.2 (3.9)	4.6 (3.2)	4.7 (3.7)	4.9 (3.4)	−0.6 (−0.9 to −0.4)	*<.001*	−0.7 (−1.1 to −0.4)	*<.001*
Sugar (g)[b]	12.9 (14.0)	15.8 (19.1)	13.3 v(14.5)	15.4 (21.1)	0.4 (−0.7 to 1.5)	.47	0.7 (−0.6 to 2.0)	.28
Sodium (mg)[b]	596.1 (343.0)	599.3 (328.9)	580.1 (342.0)	618.1 (350.7)	−32.1 (−56.3 to −7.9)	.013	−29.9 (−58.1 to −1.8)	.04
Secondary outcomes[d]								
Energy from saturated fat[b] (%)	11.0 (5.9)	9.9 (5.1)	10.2 (5.8)	10.4 (5.2)	−0.9 (−1.4 to −0.5)	*<.001*	−1.1 (−1.6 to −0.5)	*<.001*
Energy from sugar[b] (%)	12.0 (11.8)	13.9 (12.7)	12.4 (11.9)	13.1 (12.7)	1.1 (0.2 to 1.9)	.02	1.5 (0.5 to 2.5)	*.006*
Average weekly revenue per school (US $)	668.60 (420.90)	496.10 (442.63)	938.60 (574.07)	700.81 (480.06)	65.28 (−76.02 to 206.58)	.36	119.7 (−20.94 to 260.40)	.10

[a]Data were analyzed using separate linear mixed models adjusted for Socioeconomic Indexes for Areas, the school sector, and clustering at the school and student levels.

[b]Baseline intraclass correlation coefficient values: energy 0.100; saturated fat 0.130; sugar 0.131; sodium 0.111.

[c]Italics indicate statistical significance *P*<.0125.

[d]Baseline intraclass correlation coefficient values: percentage energy from sugar 0.104; percentage of energy from saturated fat 0.117.

Per-Protocol Analysis

In total, 4 schools partially implemented the intervention, with 3 schools not implementing the healthy add-ons strategy (see the section *Fidelity Checks*), and 4 schools where it could not be confirmed that the incentive strategy was delivered. The per-protocol analysis of the 5 intervention schools that implemented the intervention in full, relative to control schools, showed larger significant effects for three of the four primary outcomes (−89.4 kJ energy, *P*=.007; −0.7 g of saturated fat, *P*<.001; Table 3).

Secondary Outcomes

Healthier Purchasing

From baseline to follow-up, relative to control schools, intervention schools had greater odds of having *everyday* items purchased (odds ratio [OR] 1.69, 95% CI 1.46-1.96; *P*<.001) corresponding to a 9.77% increase in *everyday* items, and lower odds of having *occasional* items purchased (OR 0.67, 95% CI 0.57-0.78; *P*<.001) corresponding to a 7.69% decrease in *occasional* items (Table 4). The decrease in *caution* items was not significant at the prespecified significance level (OR 0.82,

95% CI 0.68-1.00; *P*=.05). The per-protocol analysis revealed that the intervention effect and significance remained similar for both *everyday* (OR 1.50, 95% CI 1.27-1.78; *P*<.001) and *occasional* (OR 0.70, 95% CI 0.59-0.84; *P*=.001) items. There were very small between-group differences in the percentage of energy from saturated fat (−0.9%, 95% CI −1.4% to −0.5%; *P*<.001) and percentage of energy from sugar (+1.1%, 95% CI 0.2%-1.9%; *P*=.02) with the result for energy significant in the main analysis, and both energy and sugar significant in the per-protocol analysis (Table 3).

Table 4. Secondary outcomes in intervention and control groups from baseline to follow-up (logistic mixed model analysis; analysis of 73,798 items purchased)[a].

Classification[b]	Baseline, n (%)		Follow-up, n (%)		Main analysis		Per-protocol analysis	
	Intervention (n=23,526)	Control (n=14,124)	Intervention (n=22,061)	Control (n=14,087)	Relative odds ratio (95% CI)	*P* value	Relative odds ratio (95% CI)	*P* value
Everyday[c]	7423 (31.55)	5711 (40.43)	8518 (38.61)	5276 (37.45)	1.69 (1.46-1.96)	*<.001*[d]	1.50 (1.27-1.78)	*<.001*
Occasional[c]	11,261 (47.87)	6185 (43.79)	9943 (45.07)	6821 (48.42)	0.67 (0.57-0.78)	*<.001*	0.70 (0.59-0.84)	*.001*
Caution[c]	4842 (20.58)	2228 (15.77)	3600 (16.32)	1990 (14.13)	0.82 (0.68-1.00)	.05	0.92 (0.74-1.14)	.39

[a]Separate logistic mixed models were used, which included a random intercept for school (to account for potential school-level clustering), a nested random intercept and random time effect for students (to account for repeated measurements between and within time points), and fixed effects for sector and Socioeconomic Indexes for Areas. Variables were dichotomized (eg, *everyday* vs other items).

[b]Baseline intraclass correlation coefficient values: percentage of *everyday* foods 0.07; percentage of *occasional* foods 0.135; percentage of *caution* foods 0.231.

[c]Chicken nuggets are commonly sold in multiple units. Some schools prepackage them (ie, 1 serve=6 nuggets), whereas other schools allow any quantity to be purchased. To account for this difference, this analysis counted the number of nuggets purchased by a single child at 1 purchasing occasion as a single item.

[d]Italics indicate statistical significance *P*<.0125.

Revenue

There was no between-group difference in the average weekly revenue from web-based canteen purchases over time (*P*=.36; Table 3).

Subgroup Analysis

There were no differences in intervention effectiveness with respect to energy content across student grade, school sector, or frequency of order (Table 5).

Table 5. Intervention impact on mean energy content (kJ) of lunch orders from baseline to follow-up: subgroup analysis for 2207 student participants.

Variable	Baseline, mean (SD)		Follow-up, mean (SD)		Intervention versus control[a]			
	Intervention (n=9726 orders)	Control (n=6279 orders)	Intervention (n=9434 orders)	Control (n=6334 orders)	Group-by-time differential effect (95% CI)	*P* value	Group-by-time-by-subgroup differential effect (95% CI)	*P* value
Student grade								
Kindergarten-grade 2	1557.9 (676.4)	1606.8 (758.0)	1551.5 (688.0)	1635.2 (820.1)	−63.9 (−134.4 to 6.6)	.07	Reference	N/A[b]
Grade 3-5	1708.7 (722.5)	1658.3 (726.3)	1699.7 (703.0)	1734.7 (853.5)	−71.6 (−144.2 to 1.0)	.05	−7.7 (−108.9 to 93.5)	.87
Frequency of use								
Low (<1 order/week)	1719.4 (706.1)	1753.2 (768.4)	1697.0 (714.9)	1723.5 (805.0)	−26.5 (−95.8 to 42.8)	.43	Reference	N/A
High (1 or more orders/week)	1592.4 (699.5)	1578.5 (725.1)	1562.3 (679.9)	1659.7 (859.9)	−118.9 (−191.9 to −45.9)	*.003*[c]	−92.4 (−193.1 to 8.2)	.07
School sector								
Independent	1569.9 (672.6)	1579.4 (664.9)	1516.6 (661.9)	1601.3 (638.6)	−96.3 (−181.4 to −11.2)	.03	Reference	N/A
Catholic	1708.6 (731.9)	1654.8 (773.2)	1738.2 (719.8)	1720.2 (905.8)	−33.4 (−97.5 to 30.8)	.28	63.0 (−43.6 to 169.5)	.22

[a]Analysis adjusted for Socioeconomic Indexes for Areas, school sector, and clustering at the school and student levels.

[b]N/A: not applicable; *P* values not applicable for reference values.

[c]Italics indicate statistical significance (*P*<.01).

Process Measures

Change in Availability

At baseline, no intervention schools and 1 control school had menus consistent with the NSW Strategy, and at follow-up, this had changed to 1 intervention school and zero control schools. The proportion of *everyday*, *occasional*, and *caution* menu items was similar between intervention and control menus at both baseline (489/858, 56.9% and 406/694, 58.5% *everyday* items; 138/858, 16.1% and 107/694, 15.4% *occasional* items; 231/858, 26.9% and 181/694, 26.1% *caution* items, respectively) and follow-up (554/876, 63.2% and 424/703, 60.3% *everyday* items; 150/876, 17.1% and 121/703, 17.2% *occasional* items; 172/876, 19.6% and 158/703, 22.5% *caution* items, respectively).

Change in Pricing

There was no between-group difference throughout time in the average price of *everyday* (*P*=.54), *occasional* (*P*=.92), or *caution* (*P*=.66) items.

Intervention Acceptability

All intervention canteen managers who completed the interview (7/9, 78%) were satisfied and would recommend the intervention to other canteen managers. Almost all agreed that the intervention strategies were acceptable (6/7, 86%; range 6/7, 86% to 7/7, 100%; Multimedia Appendix 1).

Additional Support Received

Of the 15 schools (7 intervention and 8 control schools) that completed the interview, 71% (5/7) of intervention schools and 38% (3/8) of control schools reported using the Menu Check Service during the 12-month intervention period. All intervention schools (7/7, 100%) and 75% (6/8) of the control schools reported receiving other menu support.

Fidelity Checks

Menu labeling was the basis for all automated strategies (ie, positioning, provision of tailored feedback, and incentives). As such, verifying that the labels were correctly applied indicated that these other strategies were implemented as intended. The proportions of correct labels across all 9 intervention schools were 93.6% (673/719), 94.3% (666/706), 93.4% (657/703), and 95.3% (696/730) for each of the 4 fidelity checks. There were initial issues with applying the *healthy add-ons* strategy, whereby users were overcharged for add-ons when the chosen menu item was ordered in multiples. This strategy was removed from all intervention menus for items commonly ordered in multiples (ie, chicken nuggets). In addition, this strategy caused changes to the usual layout of the production lists that could be used by managers to make up the orders. As a result, 3 intervention schools requested that the *healthy add-ons* strategy be turned off, and 5 of 9 intervention canteen managers verified that incentives were printed on lunch labels of orders with all *everyday* items.

Data Quality

In 6 schools, the purchasing data from the provider (Flexischools) was validated against observed data collected from within the school canteen during school visits. There was a 95.8% (767/800) agreement between the data sources.

Discussion

Principal Findings

This trial investigated the effect of *Click & Crunch* on the nutritional quality of students' web-based lunch orders. Intervention orders had significantly lower energy and saturated fat content relative to controls, but there was no significant difference in sugar or sodium content. Encouragingly, there was no impact on canteen revenue, suggesting no adverse intervention effect in this regard. Intervention schools had significantly greater odds of having *everyday* items purchased and lower odds of having *occasional* items purchased, corresponding to a 9.8% increase in *everyday* items and a 7.7% decrease in *occasional* items, respectively.

There was no significant difference in the odds of having *caution* items purchased between groups throughout time. The fact that all foods are classified into one of three categories means that the significant increase in *everyday* foods must happen at the cost of a decrease in other foods. In this case, the decrease is split into a reduction in the other two categories. The point estimates reflect this decrease, but only the reduction in *occasional* foods was significant, given it is the more common option, whereas the reduction in *caution* foods was not significant, given that it is the rarer option and hence there is less power to detect this effect.

Very few studies have tested the effect of delivering nudge strategies via web-based food ordering systems. In a cluster RCT conducted with 4th-7th grade students from a US school, nudge strategies delivered via a web-based ordering system led to a higher selection of fruit, vegetables, or low-fat milk in lunch orders compared with controls receiving no nudge [31]. However, the study was conducted in a single school with only a 2-week follow-up period. A nonrandomized trial of traffic light nutrition labels applied to 53 products across five food categories within a web-based supermarket found no difference in sales in a 10-week period relative to the comparison store [32]. However, a trial of web-based nudging by providing default orders resulted in a higher purchase of whole grains, fruits, and vegetables among 50 participants from New York food pantries, relative to those receiving brief nutrition education before ordering [33].

The patterns of results for this trial are similar to those of the pilot trial, which found significant reductions in the average energy, saturated fat, and sodium content of intervention lunch orders, and an increase in the proportion of healthy items ordered in the 2-month period immediately after the intervention strategies were switched on [21]. However, the effect sizes in the pilot (ie, −572 kJ between-group difference; 22% increase in *healthy* purchases) were larger than in this trial. Methodological differences between the pilot and this trial may account for the apparent differences in effect sizes. Specifically, the pilot trial was conducted for a much shorter period (2-month vs 12-month follow-up) within government schools, using different classifications (*Fresh Tastes @ School*) [22] and a different labeling system (traffic light symbols). Nonetheless, although modest, the effect on the primary outcomes may be meaningful at the population level and should be considered in light of the potential intervention reach. In NSW, 95% of primary schoolchildren access a canteen, with 55% ordering at least weekly and 4% ordering 3 to 5 times a week [34]. As such, many students could potentially benefit from intervention if disseminated broadly through the school system.

This study planned to include 26 schools. Although ethical approval was eventually obtained for government school participation, the differences in the time of exposure to the intervention between the government and nongovernment schools would have been too great to include in this 12-month trial. As such, the final sample included only 17 schools, reducing the precision of the detected effect estimates. Nonetheless, the trial was still able to detect small differences in primary trial outcomes as significant. In total, 14 schools did not consent to participate and 12 were undecided. As the principal was required to provide consent for this canteen-based intervention, it is recommended that future studies evaluate intervention acceptability with a range of stakeholders, including school principals.

Limitations

Although the intervention and control schools were broadly similar (no significance testing was conducted) [30], intervention schools had higher enrolments and more lunch orders at baseline. However, factors that are more closely related to the healthiness of student orders (eg, proportion and price of healthy foods available) were similar between groups. The research team did not have any access to individual demographic characteristics beyond student grade, and future research should seek to collect more information regarding user sociodemographic characteristics to allow an analysis of whether the intervention was more effective in certain subgroups. No data were collected from intervention end users regarding intervention acceptability, and findings regarding canteen manager acceptability were based on only 7 participants. The delay in ethics approval meant that the intervention duration varied from 8 to 10.5 months and that government schools were excluded. Previous research demonstrated no differences between government and nongovernment schools in terms of awareness of or current use of web-based canteens [35]. However, there may be differences among school sectors in terms of compliance with the NSW Strategy, which may affect the relative healthiness of canteen menus (ie, the proportion of *caution* foods vs *everyday* foods). Compliance with the strategy may influence the ability of the intervention to have an effect, in that a school compliant with the guidelines at baseline will have less room to show an improvement (ie, a *ceiling effect*). As such, future research with a larger sample of schools should investigate whether there are differences in intervention effectiveness among school sectors and investigate the relationship with strategy compliance. Finally, our analysis is based on purchasing data and not consumption data. However, objectively recorded purchasing data can provide a reasonably accurate estimate of diet quality [36,37].

The strengths of this trial include a rigorous clustered randomized controlled design. The evaluation of the intervention was based on purchasing data automatically collected by the web-based system for more than 2200 students and verified

through in-school observations. The nutritional information extracted from the purchasing data was based on rigorous menu assessment, independently verified by 2 dietitians, and based on established nutrition databases. The 12-month follow-up is also a study strength and addresses the limitations of previous trials involving web-based school lunch ordering systems [21,31,38].

Conclusions

The *Click & Crunch* intervention evaluated in this cluster RCT used existing web-based canteen ordering systems to implement choice architecture strategies to encourage healthier lunch orders for primary school students. Given the evidence of its effectiveness in decreasing the energy content of student lunch orders and increasing the purchase of *everyday* foods, its acceptability among canteen managers, and its wide reach, this intervention has the potential to influence dietary choices at a population level. The intervention may be a useful addition to the suite of strategies available to policymakers to improve child diet, including supply-side interventions encouraging the increased availability of healthier food options and interventions in nonschool settings, including after-school activities, sporting clubs, and the home. Additional research is required to determine if the results are consistent across all school sectors and sustained for a longer period and to determine intervention cost-effectiveness before investigating intervention impact in primary schools when implemented at scale.

Acknowledgments

The authors would like to thank Flexischools, the Research Advisory Group, and the participating schools, canteen managers, and canteen users. The food composition data available on the Healthy Food Finder were provided by SP Health Co Pty Ltd, trading as *Digital Wellness* (Australian Business Number 73 123 248 046) in collaboration with the NSW Ministry of Health. The data sets generated and analyzed during this study are available from the corresponding author on reasonable request and pending ethics approval. This work was supported by the National Health and Medical Research Council (NHMRC; grant number APP1120233). In-kind support was provided by the Hunter New England Population Health and the Hunter Medical Research Institute. RW is supported by a Heart Foundation Postdoctoral fellowship (ID: 102156). LW receives salary support from an NHMRC Career Development fellowship (ID: APP1128348) and Heart Foundation Future Leader fellowship (ID: 101175). Neither the NHMRC nor the Heart Foundation had any role in the design of the study, data collection, analysis or interpretation, or dissemination of findings. The provider (Flexischools) was selected through a competitive tender process. Flexischools is a commercial organization that provided the web-based canteen ordering infrastructure to schools included in the study. Flexischools had no role in the study design, data analysis, data interpretation, or manuscript writing.

Authors' Contributions

RW conceived the study; RW, LW, KC, KB, JW, CR, and TD developed the overall research plan; RW, LW, SY, NN, RS, and KR provided study oversight; RW, TD, RZ, HL, LW, and FS conducted the research; CL, FGS, and TD analyzed data; JA, CO, and CL provided statistical advice; RW led the writing of the paper with all authors contributing to paper revisions. RW has primary responsibility for the final content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Intervention acceptability to canteen managers.
[DOCX File , 17 KB - jmir_v23i9e26054_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 100 KB - jmir_v23i9e26054_app2.pdf]

References

1. Australian Burden of Disease Study 2015: Interactive data on risk factor burden. Australian Institute of Health and Welfare. 2020. URL: https://www.aihw.gov.au/reports/burden-of-disease/interactive-data-risk-factor-burden/contents/dietary-risk-factors [accessed 2021-07-16]
2. World Health Organization. School Policy Framework: Implementation of the WHO Global Strategy on Diet, Physical Activity and Health. Geneva: World Health Organization; 2008:1-48.
3. Finch M, Sutherland R, Harrison M, Collins C. Canteen purchasing practices of year 1-6 primary school children and association with SES and weight status. Aust N Z J Public Health 2006 Jun;30(3):247-251 [FREE Full text] [doi: 10.1111/j.1467-842x.2006.tb00865.x] [Medline: 16800201]

4. Delaney T, Sutherland R, Wyse R, Wolfenden L, Lecathelinais C, Janssen L, et al. A cross-sectional study of the nutritional quality of student canteen purchases from New South Wales primary-school canteens. Public Health Nutr 2019 Jul 25;22(16):3092-3100 [FREE Full text] [doi: 10.1017/s1368980019001903]
5. Wolfenden L, Nathan N, Janssen LM, Wiggers J, Reilly K, Delaney T, et al. Multi-strategic intervention to enhance implementation of healthy canteen policy: a randomised controlled trial. Implement Sci 2017 Jan 11;12(1):6 [FREE Full text] [doi: 10.1186/s13012-016-0537-9] [Medline: 28077151]
6. Nathan N, Yoong SL, Sutherland R, Reilly K, Delaney T, Janssen L, et al. Effectiveness of a multicomponent intervention to enhance implementation of a healthy canteen policy in Australian primary schools: a randomised controlled trial. Int J Behav Nutr Phys Act 2016 Oct 07;13(1):106 [FREE Full text] [doi: 10.1186/s12966-016-0431-5] [Medline: 27717393]
7. Williams CM, Nathan N, Delaney T, Yoong SL, Wiggers J, Preece S, et al. CAFÉ: a multicomponent audit and feedback intervention to improve implementation of healthy food policy in primary school canteens: protocol of a randomised controlled trial. BMJ Open 2015 Jun 24;5(6):e006969 [FREE Full text] [doi: 10.1136/bmjopen-2014-006969] [Medline: 26109111]
8. Healthy school canteens. The NSW Healthy School Canteen Strategy: Food and Drink Criteria (3rd Edition). 2018. URL: https://healthyschoolcanteens.nsw.gov.au/ [accessed 2018-11-26]
9. Hills A, Nathan N, Robinson K, Fox D, Wolfenden L. Improvement in primary school adherence to the NSW Healthy School Canteen Strategy in 2007 and 2010. Health Promot J Austr 2015 May 21;26(2):89-92. [doi: 10.1071/he14098]
10. Reilly K, Yoong SL, Sutherland R, Wiggers JH, Delaney T, Reynolds RC, et al. Secondary school implementation of a healthy eating policy. Health Promot J Austr 2021 Jan 06;32(1):21-25. [doi: 10.1002/hpja.310] [Medline: 31749164]
11. Hollands GJ, Shemilt I, Marteau TM, Jebb SA, Kelly MP, Nakamura R, et al. Altering micro-environments to change population health behaviour: towards an evidence base for choice architecture interventions. BMC Public Health 2013 Dec 21;13(1):1218 [FREE Full text] [doi: 10.1186/1471-2458-13-1218] [Medline: 24359583]
12. Sacco J, Lillico H, Chen E, Hobin E. The influence of menu labelling on food choices among children and adolescents: a systematic review of the literature. Perspect Public Health 2017 May;137(3):173-181 [FREE Full text] [doi: 10.1177/1757913916658498] [Medline: 27436235]
13. Sinclair SE, Cooper M, Mansfield ED. The influence of menu labeling on calories selected or consumed: a systematic review and meta-analysis. J Acad Nutr Diet 2014 Sep;114(9):1375-1388. [doi: 10.1016/j.jand.2014.05.014] [Medline: 25037558]
14. Long M, Tobias D, Cradock A, Batchelder H, Gortmaker S. Systematic review and meta-analysis of the impact of restaurant menu calorie labeling. Am J Public Health 2015 May;105(5):11-24 [FREE Full text] [doi: 10.2105/ajph.2015.302570]
15. Littlewood JA, Lourenço S, Iversen CL, Hansen GL. Menu labelling is effective in reducing energy ordered and consumed: a systematic review and meta-analysis of recent studies. Public Health Nutr 2015 Dec 30;19(12):2106-2121. [doi: 10.1017/s1368980015003468]
16. Cecchini M, Warin L. Impact of food labelling systems on food choices and eating behaviours: a systematic review and meta-analysis of randomized studies. Obes Rev 2016 Mar 23;17(3):201-210. [doi: 10.1111/obr.12364] [Medline: 26693944]
17. Schwartz MB. The influence of a verbal prompt on school lunch fruit consumption: a pilot study. Int J Behav Nutr Phys Act 2007 Mar 05;4(1):6 [FREE Full text] [doi: 10.1186/1479-5868-4-6] [Medline: 17338812]
18. Lai C, List JA, Samek AS. How do informational prompts affect choices in the school lunchroom? SSRN J 2017:1-37. [doi: 10.2139/ssrn.2909407]
19. Reicks M, Redden JP, Mann T, Mykerezi E, Vickers Z. Photographs in lunch tray compartments and vegetable consumption among children in elementary school cafeterias. J Am Med Assoc 2012 Feb 22;307(8):784-785. [doi: 10.1001/jama.2012.170] [Medline: 22302602]
20. Bucher T, Collins C, Rollo ME, McCaffrey TA, De Vlieger N, Van der Bend D, et al. Nudging consumers towards healthier choices: a systematic review of positional influences on food choice. Br J Nutr 2016 Apr 29;115(12):2252-2263. [doi: 10.1017/s0007114516001653]
21. Delaney T, Wyse R, Yoong SL, Sutherland R, Wiggers J, Ball K, et al. Cluster randomized controlled trial of a consumer behavior intervention to improve healthy food purchases from online canteens. Am J Clin Nutr 2017 Nov;106(5):1311-1320. [doi: 10.3945/ajcn.117.158329] [Medline: 28971849]
22. 'Fresh Tastes @ School' NSW Healthy School Canteen Strategy, Canteen Menu Planning Guide. NSW Department of Health, NSW Department of Education. 2006. URL: https://www.goodforkids.nsw.gov.au/media/1049/canteen-menu-planning-guide.pdf [accessed 2021-07-21]
23. Wyse R, Delaney T, Gibbins P, Ball K, Campbell K, Yoong SL, et al. Cluster randomised controlled trial of an online intervention to improve healthy food purchases from primary school canteens: a study protocol of the 'click & crunch' trial. BMJ Open 2019 Sep 05;9(9):e030538 [FREE Full text] [doi: 10.1136/bmjopen-2019-030538] [Medline: 31492788]
24. Healthy food finder. FoodFinder. URL: https://www.foodfinder.health.nsw.gov.au [accessed 2021-07-16]
25. FoodSwitch. The George Institute for Global Health. URL: https://www.foodswitch.com.au/ [accessed 2021-07-16]
26. FoodWorks 9 Professional. Xyris Software. 2018. URL: https://xyris.com.au/whats-new-foodworks-9-professional/ [accessed 2018-09-12]

27. Kraisid T, MacLean W, Warwick P, Food and Agriculture Organization of the United Nations. Food Energy: Methods of Analysis and Conversion Factors: Report of a Technical Workshop, Rome. Rome: Food and Agriculture Organization of the United Nations; 2003.
28. My school. Australian Curriculum, Assessment and Reporting Authority. URL: https://www.myschool.edu.au/ [accessed 2021-07-16]
29. Socio-economic indexes for areas (SEIFA). Commonwealth of Australia, Australian Bureau of Statistics. 2018. URL: https://www.abs.gov.au/ausstats/abs@.nsf/mf/2033.0.55.001 [accessed 2019-09-18]
30. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux P, CONSORT. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surg 2012;10(1):28-55 [FREE Full text] [doi: 10.1016/j.ijsu.2011.10.001] [Medline: 22036893]
31. Miller G, Gupta S, Kropp J, Grogan K, Mathews A. The effects of pre-ordering and behavioral nudges on National School Lunch Program participants' food item selection. J Econ Psychol 2016 Aug;55:4-16 [FREE Full text] [doi: 10.1016/j.joep.2016.02.010]
32. Sacks G, Tikellis K, Millar L, Swinburn B. Impact of 'traffic-light' nutrition information on online food purchases in Australia. Aust N Z J Public Health 2011 Apr;35(2):122-126 [FREE Full text] [doi: 10.1111/j.1753-6405.2011.00684.x] [Medline: 21463406]
33. Coffino J, Udo T, Hormes J. Nudging while online grocery shopping: a randomized feasibility trial to enhance nutrition in individuals with food insecurity. Appetite 2020 Sep 01;152:104714 [FREE Full text] [doi: 10.1016/j.appet.2020.104714] [Medline: 32304731]
34. Centre for Population Health. NSW Schools Physical Activity and Nutrition Survey (SPANS) 2015: Full Report. Sydney, Australia: NSW Department of Health; 2017.
35. Wyse R, Yoong SL, Dodds P, Campbell L, Delaney T, Nathan N, et al. Online canteens: awareness, use, barriers to use, and the acceptability of potential online strategies to improve public health nutrition in primary schools. Health Promot J Austr 2016 May 04;28(1):67-71. [doi: 10.1071/he15095]
36. Appelhans BM, French SA, Tangney CC, Powell LM, Wang Y. To what extent do food purchases reflect shoppers' diet quality and nutrient intake? Int J Behav Nutr Phys Act 2017 Apr 11;14(1):46 [FREE Full text] [doi: 10.1186/s12966-017-0502-2] [Medline: 28399887]
37. Eyles H, Jiang Y, Mhurchu C. Use of household supermarket sales data to estimate nutrient intakes: a comparison with repeat 24-hour dietary recalls. J Am Diet Assoc 2010 Jan;110(1):106-110 [FREE Full text] [doi: 10.1016/j.jada.2009.10.005] [Medline: 20102835]
38. Wyse R, Gabrielyan G, Wolfenden L, Yoong S, Swigert J, Delaney T, et al. Can changing the position of online menu items increase selection of fruit and vegetable snacks? A cluster randomized trial within an online canteen ordering system in Australian primary schools. Am J Clin Nutr 2019 May 01;109(5):1422-1430. [doi: 10.1093/ajcn/nqy351] [Medline: 30949660]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
NHMRC: National Health and Medical Research Council
NSW: New South Wales
OR: odds ratio
RCT: randomized controlled trial
SEIFA: Socioeconomic Indexes for Areas

Edited by R Kukafka; submitted 30.11.20; peer-reviewed by J Hoenink, N Lister; comments to author 23.03.21; revised version received 16.05.21; accepted 07.06.21; published 07.09.21.

Please cite as:
Wyse R, Delaney T, Stacey F, Zoetemeyer R, Lecathelinais C, Lamont H, Ball K, Campbell K, Rissel C, Attia J, Wiggers J, Yoong SL, Oldmeadow C, Sutherland R, Nathan N, Reilly K, Wolfenden L
Effectiveness of a Multistrategy Behavioral Intervention to Increase the Nutritional Quality of Primary School Students' Web-Based Canteen Lunch Orders (Click & Crunch): Cluster Randomized Controlled Trial
J Med Internet Res 2021;23(9):e26054
URL: https://www.jmir.org/2021/9/e26054
doi:10.2196/26054
PMID:34491207

Corrigenda and Addenda

Correction: Distant Supervision for Mental Health Management in Social Media: Suicide Risk Classification System Development Study

Guanghui Fu[1*], MSc; Changwei Song[1*], BSc; Jianqiang Li[1], PhD; Yue Ma[2], PhD; Pan Chen[3], MSc; Ruiqian Wang[1], BSc; Bing Xiang Yang[3], PhD; Zhisheng Huang[4], PhD

[1]School of Software Engineering, Beijing University of Technology, Beijing, China

[2]Interdisciplinary Laboratory of Digital Sciences, Centre national de la recherche scientifique, Université Paris-Saclay, Orsay, France

[3]School of Health Sciences, Wuhan University, Wuhan, China

[4]Department of Artificial Intelligence, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[*]these authors contributed equally

Corresponding Author:
Bing Xiang Yang, PhD
School of Health Sciences
Wuhan University
Wuchang District
Wuhan, 430072
China
Phone: 86 15902731922
Email: 00009312@whu.edu.cn

Related Article:

Correction of: https://www.jmir.org/2021/8/e26119

(J Med Internet Res 2021;23(9):e33229) doi:10.2196/33229

In "Distant Supervision for Mental Health Management in Social Media: Suicide Risk Classification System Development Study" (J Med Internet Res 2021;23(8):e26119), one error was noted.

In the originally published article, the email address of the corresponding author was incorrect.

The email address was originally published incorrectly as *"00009312@whu.wdu.cn"*.

This has been corrected to *"00009312@whu.edu.cn"*.

The correction will appear in the online version of the paper on the JMIR Publications website on September 3, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 28.08.21; this is a non–peer-reviewed article;accepted 31.08.21; published 03.09.21.

Please cite as:
Fu G, Song C, Li J, Ma Y, Chen P, Wang R, Yang BX, Huang Z
Correction: Distant Supervision for Mental Health Management in Social Media: Suicide Risk Classification System Development Study
J Med Internet Res 2021;23(9):e33229
URL: https://www.jmir.org/2021/9/e33229
doi:10.2196/33229
PMID:34478400

Corrigenda and Addenda

Addendum to the Acknowledgments: Stepwise-Hierarchical Pooled Analysis for Synergistic Interpretation of Meta-analyses Involving Randomized and Observational Studies: Methodology Development

In-Soo Shin[1], PhD; Chai Hong Rim[2], MD, PhD

[1]Graduate School of Education, Dongguk University, Seoul, Republic of Korea

[2]Department of Radiation Oncology, Ansan Hospital, Korea University, Gyeonggido, Republic of Korea

Corresponding Author:
Chai Hong Rim, MD, PhD
Department of Radiation Oncology
Ansan Hospital
Korea University
Jeokgeumro 123
Danwon Gu
Gyeonggido, 15355
Republic of Korea
Phone: 82 314126850
Fax: 82 314126851
Email: crusion3@naver.com

Related Article:

Correction of: https://www.jmir.org/2021/9/e29642

(J Med Internet Res 2021;23(9):e33534) doi:10.2196/33534

In “Stepwise-Hierarchical Pooled Analysis for Synergistic Interpretation of Meta-analyses Involving Randomized and Observational Studies: Methodology Development” (J Med Internet Res 2021;23(9):e29642), two errors were noted.

In the originally published article, the *Acknowledgments* section was incorrect. This section originally read as follows:

> *This study was supported by the National Research Fund of Korea (NRF-2018R1D1A1B07046998). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

This has been corrected to:

> *This study was supported by the National Research Fund of Korea (NRF-2018R1D1A1B07046998 and NRF-2019M2D2A1A01031560). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

As well, in the originally published article, the postal code of the Corresponding Author was incorrectly displayed as "13555." This has been corrected to "15355."

The correction will appear in the online version of the paper on the JMIR Publications website on September 20, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 11.09.21; this is a non–peer-reviewed article;accepted 13.09.21; published 20.09.21.

Please cite as:
Shin IS, Rim CH
Addendum to the Acknowledgments: Stepwise-Hierarchical Pooled Analysis for Synergistic Interpretation of Meta-analyses Involving Randomized and Observational Studies: Methodology Development
J Med Internet Res 2021;23(9):e33534
URL: https://www.jmir.org/2021/9/e33534
doi:10.2196/33534
PMID:34543226

Original Paper

Experience of and Worry About Discrimination, Social Media Use, and Depression Among Asians in the United States During the COVID-19 Pandemic: Cross-sectional Survey Study

Shuya Pan[1,2], PhD; Chia-chen Yang[3], PhD; Jiun-Yi Tsai[4], PhD; Chenyu Dong[1], PhD

[1]School of Journalism and Communication, Renmin University of China, Beijing, China

[2]Center of Journalism and Social Development, Renmin University of China, Beijing, China

[3]School of Educational Foundations, Leadership and Aviation, Oklahoma State University, Stillwater, OK, United States

[4]School of Communication, College of Social & Behavioral Sciences, Northern Arizona University, Flagstaff, AZ, United States

Corresponding Author:
Shuya Pan, PhD
School of Journalism and Communication
Renmin University of China
No. 59 Zhongguancun Street, Haidian District
Beijing, 100802
China
Phone: 86 1082508550
Email: shuya@ruc.edu.cn

Abstract

Background: The COVID-19 outbreak has spurred increasing anti-Asian racism and xenophobia in the United States, which might be detrimental to the psychological well-being of Asian people living in the United States.

Objective: We studied three discrimination-related variables, including (1) experience of discrimination, (2) worry about discrimination, and (3) racism-related social media use during the COVID-19 pandemic among Asians in the United States. We examined how these three variables were related to depression, and how the association between racism-related social media use and depression was moderated by personal experience of and worry about racial discrimination.

Methods: A web-based, cross-sectional survey was conducted. A total of 209 people (mean age 33.69, SD 11.31 years; 96/209, 45.93% female) who identified themselves as Asian and resided in the United States were included in the study.

Results: Experience of discrimination (β=.33, *P*=.001) and racism-related social media use (β=.14, *P*=.045) were positively associated with depressive symptoms. Worry about discrimination (β=.13, *P*=.14) was not associated with depression. Worry about discrimination moderated the relationship between racism-related social media use and depression (β=–.25, *P*=.003) such that a positive relationship was observed among those who had low and medium levels of worry.

Conclusions: The present study provided preliminary evidence that experience of discrimination during the COVID-19 pandemic was a risk factor of depressive symptoms among Asian people in the United States. Meanwhile, racism-related social media use was found to be negatively associated with the well-being of US Asians, and the relationship between social media use and depression was significantly moderated by worry about discrimination. It is critical to develop accessible programs to help US Asians cope with racial discrimination both in real lives and on social media during this unprecedented health crisis, especially among those who have not been mentally prepared for such challenges.

(J Med Internet Res 2021;23(9):e29024) doi:10.2196/29024

KEYWORDS

racial discrimination; experience of discrimination; worry about discrimination; depression; social media use; COVID-19

Introduction

Since the COVID-19 outbreak, anti-Asian racism and xenophobia has been spurred in the United States, and violent attacks against people who appear to be Asian have been documented [1,2]. The growing discrimination against the Asian community, combined with multifaceted pressures from the COVID-19 crisis, can make Asian people more vulnerable to

mental health problems. Recent work has provided insights into the deteriorating impact of COVID-fueled discrimination on Asians' mental health conditions [3,4], but it leaves two important gaps. First, Asians were more likely to use social media for COVID-19–related information than other media sources, thereby amplifying the possibility to encounter discrimination-related news reports or personal narratives [5,6]. However, there has not been sufficient attention toward how using social media to read, post, and talk about discrimination and racism associated with COVID-19 would relate to Asians' well-being. Second, worry emerges as a common emotional response among these individuals after having directly experienced discrimination and/or reading about personal stories shared on social platforms. Worry could negatively influence one's mental health, depending on the levels of exposure to racial discrimination. More empirical research is needed to fully explore these contributing factors to mental health outcomes, as Asian ethnic groups face lingering metal health consequences from the COVID-19 pandemic.

In an attempt to empirically examine the psychological consequences of racial discrimination associated with the COVID-19 outbreak among Asian people in the United States, we looked into three factors related to discrimination during this period and examined their associations with depressive symptoms among US Asians. These factors included: (1) experience of discrimination, (2) worry about discrimination, and (3) racism-related social media use during the COVID-19 outbreak. We also investigated whether the relationship between social media use and depression was contingent upon one's experience of and worry about discrimination.

Racial discrimination is defined as a phenomenon wherein people are unfairly treated or disadvantaged because of their race or phenotypic difference [7]. This could occur in interpersonal interactions, at the workplace, in education settings, or in receiving services and justice [8]. In some worse cases, racial discrimination could take forms of hate crimes or physical attacks [9]. It is not a new phenomenon that the outbreak of an infectious disease could trigger discrimination against racial and ethnic minorities [10]. Several recent studies have documented the salient evidence that Asians in the United States have encountered increasing levels of racial discrimination since the COVID-19 outbreak [11,12].

A huge amount of research has provided empirical evidence that racial discrimination can have negative impacts on the well-being of the stigmatized [13-15]. Racial discrimination can likely damage well-being through the process where one feels being rejected or excluded by the society and thus finds it harder to take control of their lives [16]. The above-described findings were also applicable to the Asian population in the United States. [17-19]. Recent studies have found that racial discrimination against Asian people is negatively associated with psychological well-being among this group during the COVID-19 outbreak [20,21]. One study found that direct experiences of racial discrimination during this period were associated with higher levels of anxiety and depressive symptoms among Chinese American parents and youth [20]. Another study found that vicarious racism, defined as indirect experiences of discrimination experienced by hearing about or seeing racist acts against other members of one's racial group, was also associated with higher levels of depression and anxiety among Asian Americans [21].

Current conceptualizations of racial discrimination vary across literatures, but almost all attempt to measure it in a subjective way, asking about the perception of racial discrimination in cognitive, affective, and behavioral domains [22]. Some scholars explain that it is because most incidents of discrimination are subtle and hard to document [23]. Importantly, scholars have pointed out the necessity to differentiate two related but distinct constructs of perceived racial discrimination: one is the perception of discrimination actually encountered by oneself and the other is the general worry about the fact that oneself could be a potential target of racial discrimination in the future [24]. It is worthwhile to examine the two constructs' unique relationship with depression, as previous literature suggested that the two might have different psychological implications. Although experienced discrimination was consistently associated with poor well-being [20,21], the role of worry seemed less conclusive. Worry is a normal mental reaction to impending events in our lives, and it only poses a threat when it becomes persistent and less controllable [25,26]. Under some situations, it can be instrumental by motivating people to cope, to conduct more analytical thinking, and to take more goal-oriented actions [27,28]. In the current study, we included both direct experience of discrimination during the COVID-19 outbreak among our study participants, as well as their worry about being discriminated. Together, we proposed the following hypothesis and research questions:

- Hypothesis 1 (H1): Experience of discrimination would be positively associated with depression among US Asians during the COVID-19 outbreak.
- Research question (RQ) 1: How would worry about discrimination be associated with depression among US Asians during the COVID-19 outbreak?

In addition to the experience of and worry about discrimination, social media also plays a role in US Asians' psychological well-being during the COVID-19 pandemic [19]. During the pandemic, racism-related content has become more visible on social media. A study found a nearly 10-fold increase of tweets containing "Chinese virus" or "China virus" in the week after former US President Trump mentioned this term on March 16, 2020, compared to the tweets in the week prior [29]. The sentiment analysis of these tweets suggested that the majority contained negative emotions, such as fear, sadness, anger, and disgust [29]. Another study also analyzed the sentiment of race-related tweets before and following the emergence of COVID-19 and found that the proportion of negative tweets referencing Asians increased by 68.4% from November 2019 to March 2020 [30].

Social media use during this period can thus expose Asian users to greater risk for mental health concerns. Incidental and frequent encounter with racist posts on social media might make racism-related information in memory particularly accessible, leading to a perceived social reality where being Asian is unsafe and devalued [31]. A study showed that social media exposure to COVID-related information was associated with higher odds

of anxiety and depression among the public in China during the COVID-19 outbreak, after controlling for various demographic factors [32]. The authors attributed the findings to misinformation on social media and users' risk perception of contracting COVID-19. Although the interpretation is reasonable, for US Asians, engagement with COVID-19–related information on social media, which commonly included anti-Asian sentiments [29,30], may trigger another layer of fear—fear for becoming the target of discrimination or victimization [33]. Nevertheless, to our knowledge, no research has been conducted on how reading, posting, and talking about COVID-19–related racism on social media would relate to US Asians' mental health. To bridge the gap, we proposed the following hypothesis:

- Hypothesis 2 (H2): Racism-related social media use would be positively related to depression among US Asians during the COVID-19 outbreak.

Finally, this study attempted to examine how the association between racism-related social media use and depression would be moderated by personal experience of and worry about discrimination. As posited in cultivation theory [34,35], the effect of media exposure on personal opinions and feelings would be moderated by personal experiences. They further proposed that the moderation effect could go in two different directions, which were defined as the mainstream effect (ie, people would be more influenced if they have limited experiences of the media content) and the resonance effect (ie, people would be more influenced if they have more experiences of the media content).

The vulnerability-stress model also argues that the extent to which a stressful event could trigger a mental disorder would depend on an individual's vulnerability, which could be psychological, biological, or situational factors [36]. We considered depression as the result of the interaction between worry about discrimination and racism-related social media use. As mentioned above, worry is an emotional reaction to future uncertainties, which could play a positive role in problem-solving in some situations [27,28]. In our case, worry might buffer the influence of racism-related social media use on depression by motivating people to solve the problem. Meanwhile, it is also possible for worry to play a negative role in such relationships if it does not direct people to take positive actions. As the moderating roles of discrimination-related experience and worry remained unclear, we proposed the following two additional research questions:

- RQ2: How would the experience of discrimination moderate the relationship between social media use and depression among US Asians during the COVID-19 outbreak?
- RQ3: How would worry about discrimination moderate the relationship between social media use and depression among US Asians during the COVID-19 outbreak?

Methods

Procedures

A web-based, cross-sectional survey approved by the institutional review board was first created on Qualtrics and then distributed to qualified participants via Amazon Mechanical Turk (MTurk) in mid-May 2020. Eligible participants included those belonging to Asian ethnic groups and residing in the United States. The first page of the questionnaire was a consent form providing the participants with brief background on the study and information about anonymity, confidentiality, and compensation. The rest of the pages asked questions about demographic information, media usage, discrimination-related experiences and perceptions during the COVID-19 outbreak, and some psychological measurements. Attention checkers were included to ensure data quality. The survey took an average of 15 minutes to complete.

Participants

We recruited 242 eligible participants who were Asians residing in the United States, and a total of 209 participants were included in this study after using a listwise deletion approach (ie, those who answered "not sure" or "prefer not to answer" in any questions were coded as having a missing value and were excluded from the analyses). To determine if we achieved the needed sample size for our regression analyses, an a priori sample size calculator for multiple regressions was used. The result showed that, with the anticipated medium effect size set at 0.15, the statistical power level set at 0.80, and the significance level set at P=.05, the minimum sample size required for the current study was 131. Thus, we had reached the sample size requirement in this study.

The participants were between the ages of 18 and 73 years (mean 33.69, SD 11.31 years). Among them, 96 (45.93%) were female and 113 (54.07%) were male participants. Most participants were either single (n=105, 50.24%) or married (n=97, 46.41%). The majority of the participants had a bachelor's degree (n=106, 50.72%), with the remaining holding a degree below (n=50, 23.92%) or above the bachelor's level (n=53, 25.36%). About half of the participants (n=97, 46.41%) has a full-time employment. The income levels were relatively evenly distributed among participants, but more than half of the participants (n=115, 55.02%) reported that their income had somehow been affected by the COVID-19 pandemic. Meanwhile, most participants did not identify themselves as Chinese (n=135, 64.59%). Table 1 presents demographic information of the sample.

Table 1. Descriptive analysis of the sample (N=209).

Variables	Respondents, n (%)
Age in years (range: 18-73 years)	
≤30	92 (44.02)
>30	117 (55.98)
Gender	
Male	113 (54.07)
Female	96 (45.93)
Marital status	
Single	105 (50.24)
Married	97 (46.41)
Widowed	2 (0.96)
Divorced	5 (2.39)
Education	
Less than high school degree	1 (0.48)
High school graduate	8 (3.83)
Some college but no degree	29 (13.88)
Associate degree	12 (5.74)
Bachelor's degree	106 (50.72)
Master's degree	39 (18.66)
Doctoral degree	14 (6.70)
Employment status	
Employed, working 40 or more hours per week	97 (46.41)
Employed, working 1-39 hours per week	46 (22.01)
Not employed, looking for work	41 (19.62)
Not employed, *not* looking for work	21 (10.05)
Retired	4 (1.91)
Income (US $)	
<10,000	15 (7.18)
10,000-19,999	12 (5.74)
20,000-29,999	12 (5.74)
30,000-39,999	13 (6.22)
40,000-49,999	18 (8.61)
50,000-59,999	18 (8.61)
60,000-69,999	11 (5.26)
70,000-79,999	19 (9.09)
80,000-89,999	15 (7.18)
90,000-99,999	17 (8.13)
100,000-149,999	37 (17.70)
≥150,000	22 (10.53)
Income affected	
Yes	115 (55.02)
No	94 (44.98)
Chinese	

Variables	Respondents, n (%)
Yes	74 (35.41)
No	135 (64.59)

Measurements

Experience of Discrimination

The 5-item Everyday Discrimination Scale [37] was adapted to measure how often the respondents had experienced discrimination since the COVID-19 outbreak. The scale ranged from 1 ("never") to 5 ("a lot"). The five items were as follows: (1) "You are treated with less courtesy or respect than other people"; (2) "You receive poorer service than other people"; (3) "People will act as if they think you are dangerous"; (4) "People act as if they are afraid of you"; (5) "You are threatened or harassed." The item "People act as if they think you are not smart" from the original scale was changed as "People act as if you are dangerous" to better indicate the common bias against Asians during the COVID-19 pandemic. Cronbach α for this scale was .94 (mean score 9.60, SD 4.78).

Worry About Discrimination

Participants were also asked to rate the extent to which they worry about the listed things in Everyday Discrimination Scale [37]. The items were changed to future tense since worry tends to be an emotion related to things that may happen in the future. Cronbach α for this scale was .96 (mean score 12.96, SD 5.91).

Racism-Related Social Media Use

We created a 4-item scale to measure the extent to which the participants used social media to read, post, and talk about racism-related information associated with COVID-19 since the outbreak of the pandemic. The 4 items on this scale were as follows: (1) "On social media, how much do you pay attention to and read about racism related to COVID-19?" (2) "On social media, how much do you post or repost information and news about racism related to COVID-19?" (3) "On social media, how much do you discuss racism related to COVID-19 with others through commenting?" and (4) "On social media, how much do you discuss racism related to COVID-19 with others through private messaging?" A 7-point Likert scale, ranging from 1 ("not at all") to 7 ("very much"), was used. Cronbach α for this scale was .85 (mean score 12.76, SD 6.40).

Depression

A 10-item depression scale [38] was adopted to measure the level of depression among US Asians. Respondents were asked to rate how often they felt they "[C]ouldn't shake off the blues even with help from your family and your friends," for example. The scale ranged from 1 ("rarely or none of the time") to 4 ("most or all of the time"). After performing confirmative factor analysis, three reverse-coded items were removed (see details below). Cronbach α for this 7-item scale was .88 (mean score 21.51, SD 5.97).

Control Variables

Respondents reported basic demographic information, including age, gender, marital status, education level, employment status, and income. In addition, self-identification as Chinese was also considered as a control factor because of the close connection between the COVID-19 outbreak and China. We also asked respondents whether their family incomes had been affected since the COVID-19 outbreak, as it has been documented as a key stressor during the pandemic [39].

Data Analysis

To address the issues of measurement errors, we performed confirmatory factor analysis in Mplus (version 7; Muthén & Muthén) by considering experience of discrimination, worry about discrimination, social media use, and depression as *latent variables*. The initial model fit was mediocre, and 3 reverse-coded items of the depression scale had low loadings (<0.40). After removing these 3 items, the measurement model fit was acceptable (comparative fit index=0.92, Tucker–Lewis index=0.91, root mean square error of approximation= 0.08).

We then used SPSS software (version 26.0; IBM Corp) to perform a 3-step hierarchical regression analysis. The first model included *control variables*. We dichotomized gender, marital status, employment status, and self-identification as Chinese, with those who were male, married, employed full-time, and self-identified as Chinese being coded as "1," and the other categories coded as "0." In the second model, the following 3 key independent variables were added: (1) experience of discrimination, (2) worry about discrimination, and (3) racism-related social media use. These variables were mean-centered. In the third model, the two interaction terms examining the moderating roles of experience of discrimination and worry about discrimination were added. No collinearity problems were identified after the mean-centering approach. The significance level was set at $P<.05$.

Results

As shown in Table 2, all three regression models were significant: model 1 ($R^2=0.11$, $F_{8,200}= 3.07$; $P=.003$), model 2 ($R^2=0.35$, $F_{11,197}= 9.71$; $P<.001$), and model 3 ($R^2=0.38$, $F_{13,195}= 9.21$; $P<.001$). Model 2 with three discrimination-related variables significantly increased the explaining power for depression (R^2 change=0.24; F change=24.51; $P<.001$).

Experience of discrimination was positively associated with depressive symptoms (β=.33, $P=.001$), supporting hypothesis 1. Racism-related social media use (β=.14, $P=.045$) was also positively related to depression, supporting hypothesis 2. Worry about discrimination (β=.13, $P=.14$) was not associated with depression (RQ1).

To answer RQ2 and RQ3, we found the significant interactive effect of social media use and worry about discrimination on depression (β=–.25, $P=.003$). To visualize the moderation effect, we used Process version 3.5 to conduct simple slope analysis and produce the figure of the interaction. As depicted in Figure

1, among participants with low (1 SD below the mean) and medium levels of worry, social media use was positively and significantly related to depression (simple effect coefficient=0.29 for low worry, 95% CI 0.15-0.44; simple slope coefficient=0.11 for medium worry, 95% CI 0.0023-0.21). No significant relationship between social media use and depression was found at a high level of worry.

In addition, female Asians were found to be more depressed than male Asians (β=–.14, *P*=.02), and Chinese Asians were less depressed than those Asians who were not Chinese (β=–.12, *P*=.045).

Table 2. Ordinary least squares multiple regression analysis of depression scores as a function of discrimination-related variables and control variables (N=209).

Variables	Model 1		Model 2		Model 3	
	β	*P* value	β	*P* value	β	*P* value
Age	–.18	.02	–.15	.029	–.13	.05
Gender (male)	–.12	.08	–.14	.020	–.14	.02
Marital status (married)	–.08	.29	–.10	.16	–.11	.12
Education	–.04	.62	–.06	.33	–.07	.27
Employed (full time)	.04	.62	.03	.61	.02	.72
Income	–.06	.45	–.04	.52	–.06	.37
Income affected (yes)	.12	.07	.07	.28	.06	.31
Chinese	.00	.99	–.11	.06	–.12	.045
Experience of discrimination	N/A[a]	N/A	.34	<.001	.33	.001
Worry about discrimination	N/A	N/A	.07	.40	.13	.14
Social media use	N/A	N/A	.18	.012	.14	.045
SM[b]×Experience[c]	N/A	N/A	N/A	N/A	.16	.10
SM×Worry[d]	N/A	N/A	N/A	N/A	–.25	.003

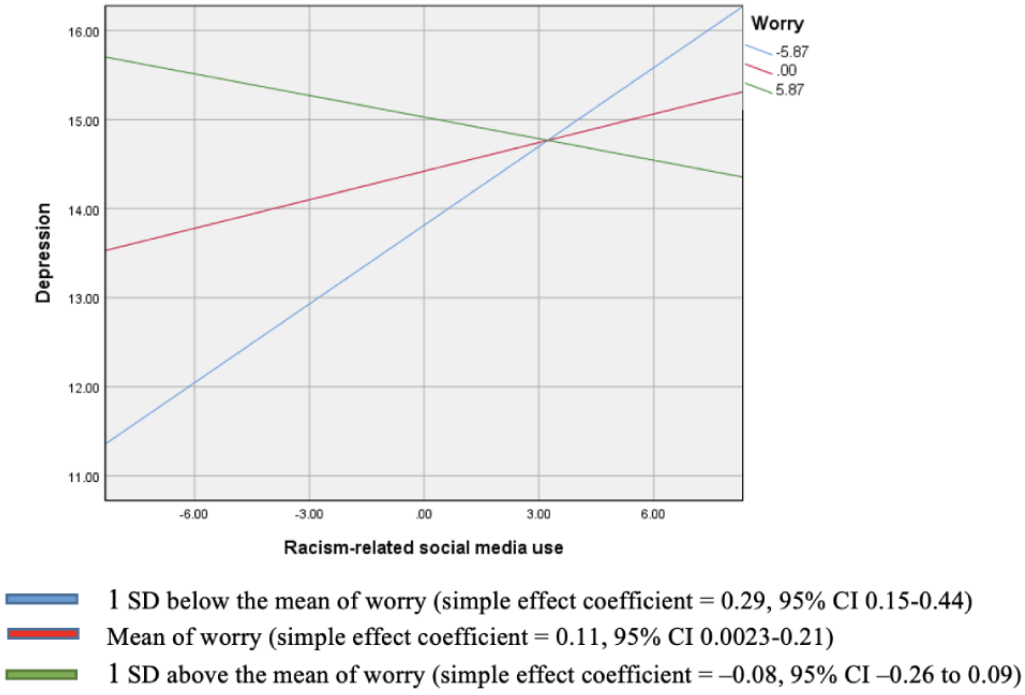
[a]N/A: not applicable.

[b]SM: racism-related social media use.

[c]Experience: experience of discrimination.

[d]Worry: worry about discrimination.

Figure 1. Interactive plot showing predicted values of depression as the function of racism-related social media use and worry about discrimination.



Discussion

Principal Findings

The present study provides preliminary evidence that experience of discrimination during the COVID-19 pandemic is a risk factor for depressive symptoms among Asians in the United States. Meanwhile, racism-related social media use was also found to be negatively related to the well-being of US Asians, and the relationship between social media use and depression was moderated by worry about discrimination, such that social media use was associated with greater depression among those who were less worried about discrimination.

Drawing on survey data from 209 Asians residing in the United States, the study presents the findings consistent with earlier research [20,21] showing that experience of racial discrimination induced by the COVID-19 pandemic was negatively related to the well-being among Asian people. Effective interventions should be designed to provide timely assistance to Asians who have experienced discrimination during the pandemic to prevent them from developing further depressive symptoms. One limitation of the current study is the use of an existing measurement of experienced racial discrimination, which may fail to capture the specific experiences of racial discrimination associated with the COVID-19 pandemic. Further research could focus on understanding the unique elements of racial discrimination in the COVID-19 context so that more accurate conclusions could be made.

Using social media to read, post, and talk about racism was also found to be associated with higher depression. It is noteworthy that the relationship was significant when controlling for direct experience of discrimination (albeit with a smaller effect size than direct experience). Given how fast-growing and visible anti-Asian sentiments have been on social media during the pandemic [29,30], and that Asians were more likely to use social media than other media sources to obtain COVID-19–related information [5,6], our finding deserves serious attention. It is also worth mentioning that our social media use scale includes a variety of social media activities (eg, reading, posting, commenting, private messaging) through which individuals may be exposed to and engage with racism-related information. Relative to single-item questions asking about participants' social media use in one particular way, such as exposure to or consumption of a certain type of information [32,40], our scale has the strength of more comprehensively capturing multiple ways of social media use during the pandemic.

Worry about discrimination was not directly associated with depression in the present study. When worry about discrimination was considered alongside the actual experience of discrimination, it was the actual experience that was detrimental. As shown in previous literature, the benefits and detriments of worry may have cancelled out each other, leading to a nonsignificant direct association [25-28]. It may be worthwhile for future research to explore possible moderators of this relationship, such as coping styles and social support-seeking behaviors, to better understand the implications of this type of worry for mental health.

Although worry was not a significant independent variable, it was a significant moderator for the relationship between social media use and depression. Specifically, among Asians with low and medium levels of worry about discrimination, social media use was related to higher levels of depression. A possible explanation for our finding might be that Asians with low to medium worry were less prepared for anti-Asian discrimination. Thus, they felt more overwhelmed and depressed when exposed to greater amount of racism-related information on social media. Future research should explore in detail how US Asians with low and high worry cope with racial discrimination in different ways that may buffer or exacerbate the influence of social media use on depression. Meanwhile, it is also worthwhile to explore more factors that might moderate the relationship between social media use and depression, such as coping strategies, perceived social support, and so forth.

In terms of demographic factors considered in our regression model, we found that female Asians were more depressed than male Asians, which is consistent with findings of previous studies [41,42]. Surprisingly, Chinese Asians were found to be less depressed than Asians who were not Chinese. Although Asians are normally being treated as one single ethnic group in the United States, research does indicate that there might be differences in the mental health mechanisms among various Asian ethnicities. For example, it was indicated that Chinese Americans had lower prevalence of depression than Korean and Vietnamese Americans [43]. Further study is needed to identify how the mental health conditions of different Asian groups have been impacted by the COVID-19 pandemic.

Limitations and Conclusions

The study has some limitations. The first is that the research model was not evaluated with a random sample. Therefore, readers are advised to take caution when generalizing the results to the overall Asian population in the United States or Asian groups in another country. Another limitation is that the cross-sectional study does not allow for drawing causal conclusions, and thus the findings should be interpreted as correlations rather than as having a cause-and-effect relationship.

Despite all these limitations, the present study extends the current literature on rising anti-Asian discrimination and associated mental health outcomes in several ways. First, we differentiated the actual experience of discrimination and worry about discrimination and revealed their unique relationship with depression among US Asians in the context of the COVID-19 pandemic. Another theoretical contribution is providing important insights into the specific role of racism-related social media use in mental health among a minority group. Importantly, the results suggest that US Asians with higher versus lower levels of worry face different degrees of susceptibility to the negative influences of racism-related social media use on mental health.

From a practical perspective, immediate and effective interventions should be implemented to decrease racist and discriminatory behaviors against US Asians, as direct experience of discrimination was found to be a robust risk factor for depression among this group, after controlling for various demographic variables. It is also critical to develop accessible

programs to help US Asians cope with racial discrimination both in real lives and on social media during this unprecedented health crisis, especially among those who have not been mentally prepared for such challenges.

Acknowledgments

This study was supported by Journalism and Marxism Research Center, Renmin University of China (Project Number: 19MXG06).

Conflicts of Interest

None declared.

References

1. Covid-19 fueling anti-Asian racism and xenophobia worldwide. Human Rights Watch. URL: https://www.hrw.org/news/2020/05/12/covid-19-fueling-anti-asian-racism-and-xenophobia-worldwide [accessed 2020-06-01]
2. Haynes S. As coronavirus spreads, so does xenophobia and anti-Asian racism. Time. URL: https://time.com/5797836/coronavirus-racism-stereotypes-attacks [accessed 2020-05-18]
3. Wu C, Qian Y, Wilkes R. Anti-Asian discrimination and the Asian-white mental health gap during COVID-19. Ethnic and Racial Studies 2020 Dec 14;44(5):819-835. [doi: 10.1080/01419870.2020.1851739]
4. Choi S, Hong JY, Kim YJ, Park H. Predicting psychological distress amid the COVID-19 pandemic by machine learning: discrimination and coping mechanisms of Korean immigrants in the U.S. Int J Environ Res Public Health 2020 Aug 20;17(17):6057 [FREE Full text] [doi: 10.3390/ijerph17176057] [Medline: 32825349]
5. Liu PL. COVID-19 information on social media and preventive behaviors: managing the pandemic through personal responsibility. Soc Sci Med 2021 May;277:113928 [FREE Full text] [doi: 10.1016/j.socscimed.2021.113928] [Medline: 33865093]
6. Yu N, Pan S, Yang C, Tsai J. Exploring the role of media sources on COVID-19-related discrimination experiences and concerns among Asian people in the United States: cross-sectional survey study. J Med Internet Res 2020 Nov 06;22(11):e21684 [FREE Full text] [doi: 10.2196/21684] [Medline: 33108307]
7. Fox S, Stallworth LE. Racial/ethnic bullying: exploring links between bullying and racism in the US workplace. J Vocat Behav 2005 Jun;66(3):438-456. [doi: 10.1016/j.jvb.2004.01.002]
8. Chakraborty A, McKenzie K. Does racial discrimination cause mental illness? Br J Psychiatry 2002 Jun;180:475-477. [doi: 10.1192/bjp.180.6.475] [Medline: 12042221]
9. Barnes A, Ephross P. The impact of hate violence on victimsmotional and behavioral responses to attacks. Social Work 1994;39(3):247-251. [doi: 10.1093/sw/39.3.247]
10. Gover AR, Harper SB, Langton L. Anti-Asian hate crime during the COVID-19 pandemic: exploring the reproduction of inequality. Am J Crim Justice 2020 Jul 07:1-21 [FREE Full text] [doi: 10.1007/s12103-020-09545-1] [Medline: 32837171]
11. Ruiz N, Horowitz J, Tamir C. Many Black and Asian Americans say they have experienced discrimination amid the COVID-19 outbreak. Pew Research Center. 2020 Jul 01. URL: https://www.pewresearch.org/social-trends/2020/07/01/many-black-and-asian-americans-say-they-have-experienced-discrimination-amid-the-covid-19-outbreak/ [accessed 2021-08-24]
12. Stop AAPI Hate Reporting Center. Asian Pacific Policy and Planning Council. URL: http://www.asianpacificpolicyandplanningcouncil.org/stop-aapi-hate/ [accessed 2020-07-17]
13. Benner AD, Wang Y, Shen Y, Boyle AE, Polk R, Cheng Y. Racial/ethnic discrimination and well-being during adolescence: A meta-analytic review. Am Psychol 2018 Oct;73(7):855-883 [FREE Full text] [doi: 10.1037/amp0000204] [Medline: 30024216]
14. Brown TN, Williams DR, Jackson JS, Neighbors HW, Torres M, Sellers SL, et al. "Being black and feeling blue": the mental health consequences of racial discrimination. Race and Society 2000 Jan;2(2):117-131. [doi: 10.1016/s1090-9524(00)00010-3]
15. Feinstein BA, Goldfried MR, Davila J. The relationship between experiences of discrimination and mental health among lesbians and gay men: An examination of internalized homonegativity and rejection sensitivity as potential mechanisms. J Consult Clin Psychol 2012 Oct;80(5):917-927. [doi: 10.1037/a0029425]
16. Wirth JH, Williams KD. 'They don't like our kind': consequences of being ostracized while possessing a group membership. Group Processes & Intergroup Relations 2009 Jan 01;12(1):111-127. [doi: 10.1177/1368430208098780]
17. Hwang W, Goto S. The impact of perceived racial discrimination on the mental health of Asian American and Latino college students. Cultur Divers Ethnic Minor Psychol 2008 Oct;14(4):326-335. [doi: 10.1037/1099-9809.14.4.326] [Medline: 18954168]
18. Yip T, Gee GC, Takeuchi DT. Racial discrimination and psychological distress: the impact of ethnic identity and age among immigrant and United States-born Asian adults. Dev Psychol 2008 May;44(3):787-800 [FREE Full text] [doi: 10.1037/0012-1649.44.3.787] [Medline: 18473644]

19. Yang C, Tsai J, Pan S. Discrimination and well-being among Asians/Asian Americans during COVID-19: the role of social media. Cyberpsychol Behav Soc Netw 2020 Dec;23(12):865-870. [doi: 10.1089/cyber.2020.0394] [Medline: 32762541]
20. Cheah CS, Wang C, Ren H, Zong X, Cho HS, Xue X. COVID-19 racism and mental health in Chinese American families. Pediatrics 2020 Nov;146(5):e2020021816. [doi: 10.1542/peds.2020-021816] [Medline: 32873719]
21. Chae DH, Yip T, Martz CD, Chung K, Richeson JA, Hajat A, et al. Vicarious racism and vigilance during the COVID-19 pandemic: mental health implications among Asian and Black Americans. Public Health Rep 2021 May 25;136(4):508-517. [doi: 10.1177/00333549211018675] [Medline: 34034574]
22. Atkins R. Instruments measuring perceived racism/racial discrimination: review and critique of factor analytic techniques. Int J Health Serv 2014 Oct 01;44(4):711-734. [doi: 10.2190/hs.44.4.c]
23. Dovidio JF, Gaertner SL. Changes in the expression and assessment of racial prejudice. In: Knopke HJ, Norrell RJ, Rogers RW, editors. Opening doors: perspectives on race relations in contemporary America. Tuscaloosa, AL: The University of Alabama Press; 1991:119-148.
24. Paradies YC. Defining, conceptualizing and characterizing racism in health research. Crit Public Health 2006 Jun;16(2):143-157. [doi: 10.1080/09581590600828881]
25. Craske MG, Rapee RM, Jackel L, Barlow DH. Qualitative dimensions of worry in DSM-III-R generalized anxiety disorder subjects and nonanxious controls. Behav Res Ther 1989;27(4):397-402. [doi: 10.1016/0005-7967(89)90010-7]
26. Wells A. Meta-cognition and worry: a cognitive model of generalized anxiety disorder. Behav Cogn Psychother 2009 Jun 16;23(3):301-320. [doi: 10.1017/s1352465800015897]
27. Davey GCL, Tallis F, Capuzzo N. Beliefs about the consequences of worrying. Cogn Ther Res 1996 Oct;20(5):499-520. [doi: 10.1007/bf02227910]
28. Sweeny K, Dooley MD. The surprising upsides of worry. Soc Personal Psychol Compass 2017 Apr 18;11(4):e12311. [doi: 10.1111/spc3.12311]
29. Dubey AD. The resurgence of cyber racism during the COVID-19 pandemic and its aftereffects: analysis of sentiments and emotions in tweets. JMIR Public Health Surveill 2020 Oct 15;6(4):e19833-e19551 [FREE Full text] [doi: 10.2196/19833] [Medline: 32936772]
30. Nguyen TT, Criss S, Dwivedi P, Huang D, Keralis J, Hsu E, et al. Exploring U.S. shifts in anti-Asian sentiment with the emergence of COVID-19. Int J Environ Res Public Health 2020 Sep 25;17(19):7032 [FREE Full text] [doi: 10.3390/ijerph17197032] [Medline: 32993005]
31. Shrum LJ. Media Consumption and Perceptions of Social Reality: Effects and Underlying Processes. In: Bryant J, Oliver MB, editors. Media Effects: Advances in Theory and Research. New York, NY: Routledge; 2002:50-73.
32. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]
33. Custers K, Dorrance Hall E, Bushnell Smith S, McNallie J. The indirect association between television exposure and self-protective behavior as a result of worry about crime: the moderating role of gender. Mass Commun Soc 2017 Apr 17;20(5):637-662. [doi: 10.1080/15205436.2017.1317353]
34. Shrum J, Bischak D. Mainstreaming, resonance, and impersonal impact. Human Comm Res 2001 Apr;27(2):187-215. [doi: 10.1111/j.1468-2958.2001.tb00780.x]
35. Gerbner G, Gross L, Signorielli N, Morgan M. Aging with television: images on television drama and conceptions of social reality. J Commun 1980;30(1):37-47. [doi: 10.1111/j.1460-2466.1980.tb01766.x] [Medline: 7372841]
36. Ingram R, Luxton D. Vulnerability-Stress Models. In: Hankin BL, Abela JRZ, editors. Development of Psychopathology: a vulnerability stress perspective (pp. 32-46).; Thousand Oaks. Thousand Oaks, CA: Sage Publications, Inc; 2005:32-46.
37. Sternthal MJ, Slopen N, Williams DR. Racial disparities in health. Du Bois Rev 2011 Apr 15;8(1):95-113. [doi: 10.1017/s1742058x11000087]
38. Meadows SO, Brown JS, Elder GH. Depressive symptoms, stress, and support: gendered trajectories from adolescence to young adulthood. J Youth Adolescence 2006 Mar 18;35(1):89-99. [doi: 10.1007/s10964-005-9021-6]
39. Chiwaya N, Wu J. The coronavirus has destroyed the job market in every state. NBC News. 2020 Apr 14. URL: https://www.nbcnews.com/business/economy/unemployment-claims-state-see-how-covid-19-has-destroyed-job-n1183686 [accessed 2020-07-05]
40. Ni MY, Yang L, Leung CMC, Li N, Yao XI, Wang Y, et al. Mental health, risk factors, and social media use during the COVID-19 epidemic and cordon sanitaire among the community and health professionals in Wuhan, China: cross-sectional survey. JMIR Ment Health 2020 May 12;7(5):e19009 [FREE Full text] [doi: 10.2196/19009] [Medline: 32365044]
41. Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, de Girolamo G, et al. Cross-national epidemiology of DSM-IV major depressive episode. BMC Med 2011 Jul 26;9(1):90 [FREE Full text] [doi: 10.1186/1741-7015-9-90] [Medline: 21791035]
42. Salk RH, Hyde JS, Abramson LY. Gender differences in depression in representative national samples: meta-analyses of diagnoses and symptoms. Psychol Bull 2017 Aug;143(8):783-822 [FREE Full text] [doi: 10.1037/bul0000102] [Medline: 28447828]

43. Chau V, Bowie JV, Juon H. The association of perceived discrimination and depressive symptoms among Chinese, Korean, and Vietnamese Americans. Cultur Divers Ethnic Minor Psychol 2018 Jul;24(3):389-399 [FREE Full text] [doi: 10.1037/cdp0000183] [Medline: 29389149]

Abbreviations

H1: hypothesis 1
H2: hypothesis 2
RQ: research question

Edited by G Eysenbach; submitted 22.03.21; peer-reviewed by D Zhang, D Liu; comments to author 14.04.21; revised version received 28.06.21; accepted 01.08.21; published 01.09.21.

Please cite as:
Pan S, Yang CC, Tsai JY, Dong C
Experience of and Worry About Discrimination, Social Media Use, and Depression Among Asians in the United States During the COVID-19 Pandemic: Cross-sectional Survey Study
J Med Internet Res 2021;23(9):e29024
URL: https://www.jmir.org/2021/9/e29024
doi:10.2196/29024
PMID:34351304

Original Paper

Effect of the Period From COVID-19 Symptom Onset to Confirmation on Disease Duration: Quantitative Analysis of Publicly Available Patient Data

Myung-Bae Park[1*], PhD; Eun Young Park[2*], MD, PhD; Tae Sic Lee[3], MD, PhD; Jinhee Lee[4], MD, PhD

[1]Department of Gerontology, Division of Health and Welfare, Pai Chai University, Seo-gu, Republic of Korea

[2]Department of Obstetrics and Gynecology, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea

[3]Department of Biomedical Science and Engineering, Gwangju Institute of Science and Technology, Gwangju, Republic of Korea

[4]Department of Psychiatry, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jinhee Lee, MD, PhD
Department of Psychiatry
Yonsei University Wonju College of Medicine
Ilsan-ro 20
Wonju, 26426
Republic of Korea
Phone: 82 1074470582
Email: jinh.lee95@yonsei.ac.kr

Abstract

Background: In general, early intervention in disease based on early diagnosis is considered to be very important for improving health outcomes. However, there is still insufficient evidence regarding how medical care that is based on the early diagnosis of confirmed cases can affect the outcome of COVID-19 treatment.

Objective: We aimed to investigate the effect of the duration from the onset of clinical symptoms to confirmation of COVID-19 on the duration from the onset of symptoms to the resolution of COVID-19 (release from quarantine).

Methods: For preliminary data collection, we performed data crawling to extract data from social networks, blogs, and official websites operated by local governments. We collected data from the 4002 confirmed cases in 33 cities reported up to May 31, 2020, for whom sex and age information could be verified. Subsequently, 2494 patients with unclear symptom onset dates and 1349 patients who had not been released or had no data about their release dates were excluded. Thus, 159 patients were finally included in this study. To investigate whether rapid confirmation reduces the prevalence period, we divided the duration from symptom onset to confirmation into quartiles of ≤1, ≤3, ≤6, and ≥7 days, respectively. We investigated the duration from symptom onset to release and that from confirmation to release according to these quartiles. Furthermore, we performed multiple regression analysis to investigate the effects of rapid confirmation after symptom onset on the treatment period, duration of prevalence, and duration until release from isolation.

Results: We performed multiple regression analysis to investigate the association between rapid confirmation after symptom onset and the total prevalence period (faster release from isolation). The time from symptom onset to confirmation showed a negative association with the time from confirmation to release (t_1=−3.58; P<.001) and a positive association with the time from symptom onset to release (t_1=5.86; P<.001); these associations were statistically significant.

Conclusions: The duration from COVID-19 symptom onset to confirmation date is an important variable for predicting disease prevalence, and these results support the hypothesis that a short duration of symptom onset to confirmation can reduce the time from symptom onset to release.

(J Med Internet Res 2021;23(9):e29576) doi:10.2196/29576

KEYWORDS

COVID-19; SARS-CoV-2; symptoms onset; duration of prevalence; confirmation; South Korea; data crawling; social media; Internet; dataset; symptom; duration; outcome; diagnosis; prevalence

Introduction

The COVID-19 Outbreak

COVID-19 was first reported on December 31, 2019, in cases of pneumonia with an unknown etiology in Wuhan, China [1]; the disease subsequently spread to neighboring countries, including South Korea and Japan [2]. On March 11, 2020, the World Health Organization (WHO) reported a total of 118,000 confirmed cases and approximately 4000 deaths due to COVID-19 in 114 countries. On the afternoon of the same day, the WHO formally declared COVID-19 to be a global pandemic [3]. The first confirmed case in South Korea, a Chinese woman in her 30s, was reported on January 20, 2020, followed by another confirmed case, a 55-year-old man who had visited Wuhan, China. Subsequently, COVID-19 appeared to be stably controlled for almost a month; however, the number of cases increased explosively after February 20, whereby South Korea became the second country after China to experience the beginning of an epidemic. Subsequently, on February 29, the daily incidence of confirmed cases peaked at 900 and has since been following a downward trend. As of November 30, 2020, a total of 64 million cumulative confirmed cases of COVID-19 had been reported worldwide [4].

Treatment and Management of Confirmed COVID-19 Cases in South Korea

The WHO defines a confirmed case of COVID-19 as a person with a laboratory-confirmed COVID-19 infection, regardless of clinical signs and symptoms [5]. The symptoms mainly include fever, cough, shortness of breath, and breathing difficulties [6], although more than 15% of cases are asymptomatic [7]. In South Korea, individuals who have entered the country or have been in close contact with a confirmed COVID-19 case must undergo a test for COVID-19. The rapid antigen test is an efficient tool for rapid confirmation of SARS-CoV-2 infection, although it is not used as an official tool for confirmatory diagnosis in view of the capacity of laboratories to conduct the requisite number of tests and considering the number of confirmed and suspected cases. In the early stages of the COVID-19 epidemic, confirmatory testing involved two stages: pan-coronavirus and base sequencing analysis. However, the real-time polymerase chain reaction (RT-PCR) has subsequently been used for basic confirmation [8]. Individuals who are suspected of having COVID-19 based on the clinical presentation of symptoms or close contact with a confirmed patient with COVID-19 are diagnosed based on test results from two clinical samples, one each from the upper and lower respiratory tract [8,9]. RT-PCR requires at least 24 hours for the results to become available, and everyone who is tested is quarantined until the final diagnosis is determined.

The treatment strategies for COVID-19 can generally be divided into supportive care, respiratory support, symptomatic treatment, nutritional support, and psychological intervention [10]. Moreover, antibiotics can be administered to prevent secondary infections [11]. Furthermore, remdesivir is prescribed as an antiviral treatment for SARS-CoV-2, although its effectiveness is debated [12,13]. In South Korea, patients are classified according to disease severity, and patients with high severity are treated in hospitals dedicated to infectious diseases and at national inpatient treatment centers. Patients who are classified as having mild symptoms or those who do not require inpatient treatment due to improvement in clinical symptoms may be admitted to residential treatment centers. In these facilities, the medical staff monitors patients at least twice per day.

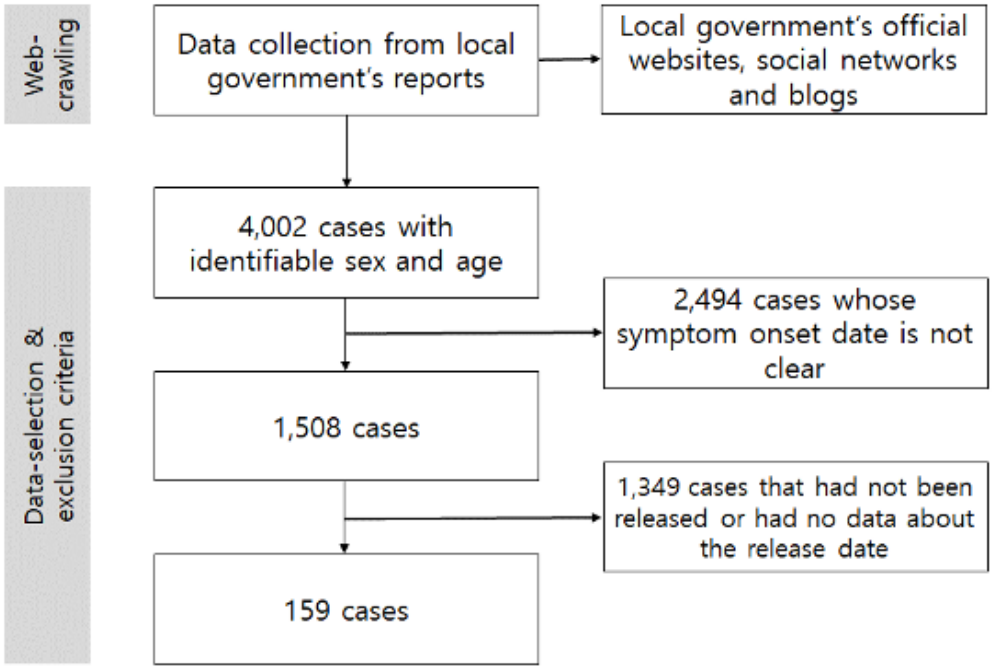
COVID-19 is highly contagious, therefore, confirmed patients must avoid contact with other people and should immediately be placed in quarantine [14-16]. In general, early intervention based on early diagnosis of the disease is considered to be very important for improving health outcomes [17-19]. However, most of the studies on COVID-19 conducted to date have focused on preventing the spread of the disease by quickly identifying confirmed patients and minimizing their contact with other people [15,20]. In contrast, there is very little evidence of how medical care that is based on the early diagnosis of confirmed cases of COVID-19 can affect the treatment outcomes. Accordingly, we aimed to investigate the effects of the duration from the onset of clinical symptoms to confirmation on the duration from the onset of clinical symptoms to the resolution of COVID-19 (release from quarantine).

Methods

Data Collection

South Korea comprises 228 si (cities), gun (counties), and gu (districts), with 17 metropolitan city and province levels, including Seoul, its capital. All local governments disclose basic information on official websites about confirmed cases. Almost all Korean local governments have provided anonymized public information regarding confirmed COVID-19 cases, such as symptom date, release date, and age, up to the first half of 2020. We performed web crawling using Python's Selenium module and additionally used the Beautiful Soup and Pandas libraries to collect data. In the early stage of the pandemic, the local government deidentified the information of the confirmed persons with COVID-19 and disclosed their movements, region, and age. At this time, many people connected to the local government homepages to check these data, and the server was overloaded. Accordingly, many local governments provided information to their official social networks and blogs to distribute data traffic and improve access to information. Therefore, we only collected information on local government official websites, social networks, and blogs, and we did not use information from other sites that are not guaranteed to be reliable. For preliminary data collection, we performed data crawling to extract data from social networks, blogs, and official websites operated by local governments. Second, the data were manually reviewed and revised. Data obtained in this manner underwent a final review and revision process that included the use of various check codes to complete the final data set. Data were extracted and reviewed from June 1-7, 2020 (Figure 1).

Figure 1. Flowchart of the data extraction process.



Study Design and Participants

We collected data from the 4002 confirmed cases from 33 cities reported up to May 31, 2020, for whom sex and age information could be verified. Subsequently, 2494 patients with an unclear symptom onset date and 1349 patients who had not been released or had no data about their release date were excluded. Thus, 159 patients were finally included in this study.

A confirmed case of COVID-19 was defined based on a confirmed diagnosis from positive RT-PCR results on 2 or more clinical specimens, regardless of clinical signs and symptoms [21]. For a patient to be released from isolation, at least 7 days must pass from the day of confirmation; the patient must be afebrile without the administration of any antipyretic agents; and the clinical symptoms must show an improving trend. In addition, the patient must test negative on 2 consecutive PCR tests, each conducted at least 24 hours apart [22].

Statistical Analysis

We performed descriptive analysis of the mean, minimum, and maximum values for the durations from symptom onset to release, symptom onset to confirmation, and confirmation to release according to the sex and age of the patients. To investigate whether rapid confirmation reduces the prevalence period, we divided the period from symptom onset to confirmation into quartiles, 1Q, 2Q, 3Q, and 4Q, of ≤1, ≤3, ≤6, and ≥7 days, respectively. We investigated the durations from symptom onset to release and confirmation to release according to these quartiles. Furthermore, we performed multiple regression analysis to investigate the effects of rapid confirmation after symptom onset on the treatment period, duration of prevalence, and duration until release from isolation.

To measure the predictive power of duration from symptom onset to release, we used logistic regression (LR) after setting the duration from symptom onset to release and other variables (age, sex, and symptom onset to confirmation) as the dependent and independent variables, respectively. We transformed the number of days from symptoms to release into tertiles (T1, T2, and T3). Two scenarios were evaluated: (1) the classification of early symptom onset to release after setting T1 and T2-T3 as 1 and 0 coding, respectively, and (2) the classification of late symptom onset to release after setting T3 and T1-T2 as 1 and 0 coding, respectively. We randomly divided the data into training and test sets by ratios of 70% and 30%, respectively. We determined two lists of variables with the following combinations: (1) age + sex and (2) age + sex + duration from symptom onset to confirmation. We constructed an LR model using a training set to predict early or late symptom onset to release. Thereafter, we measured the predictive performance of early or late symptom onset to release in the test sets. With 100 iterations of the random division of training and test sets, we measured the average performance of the classification for early or late symptom onset to release. For the 2-tailed *t* tests, we considered a *P* value <.05 to be statistically significant.

Ethical Approval

This study was conducted in accordance with the Declaration of Helsinki, and all of the materials used in the article were publicly available data. Moreover, all of those data are nonidentifying data, and they are available for use by anyone.

Results

The study population (N=159) included 67 men (42.1%) and 92 women (57.9%), whereas the age groups appeared in the following order: 40-59 years (n=71, 44.7%), 20-39 years (n=59, 37.1%), ≥60 years (n=21, 13.2%), and 0-19 years (n=8, 5.0%). The mean duration from symptom onset to confirmation was 6.1 days, and the mean duration from confirmation to release was 25.3 days. The mean duration from symptom onset to confirmation was shorter among females (5.6 days) than among males (6.7 days), whereas the mean duration from confirmation to release was similar for males (25.4 days) and females (25.2

days). The mean duration from symptom onset to release was 32.1 and 30.8 days among male and female patients, respectively. With regard to age, the mean duration from symptom onset to confirmation was longest in the 0-19 years group (7.9 days), followed by the 40-59 years (7.3 days), 20-39 years (5.0 days), and ≥60 years (4.3 days) groups. The mean duration from confirmation to release was longest in the ≥60 years group (27.8 days), whereas the mean duration from symptom onset to release was longest in the 40-59 years group (33 days; Table 1).

Both the duration from confirmation to release and the duration from symptom onset to release were investigated by quartiles of duration from symptom onset to confirmation, and 4Q, 3Q, 2Q, and 1Q were ≥7 days, 4-6 days, 2-3 days, and 1 day, respectively. The duration of symptom onset to release was shortest, at 28.5 days, when the duration of symptom onset to confirmation was 1Q (≤1 day); this was followed sequentially by 2Q, 3Q, and 4Q, with 28.9, 31.6, and 36.5 days, respectively. The results indicated that the prevalence increased with increasing duration of symptom onset to confirmation (Figure 2).

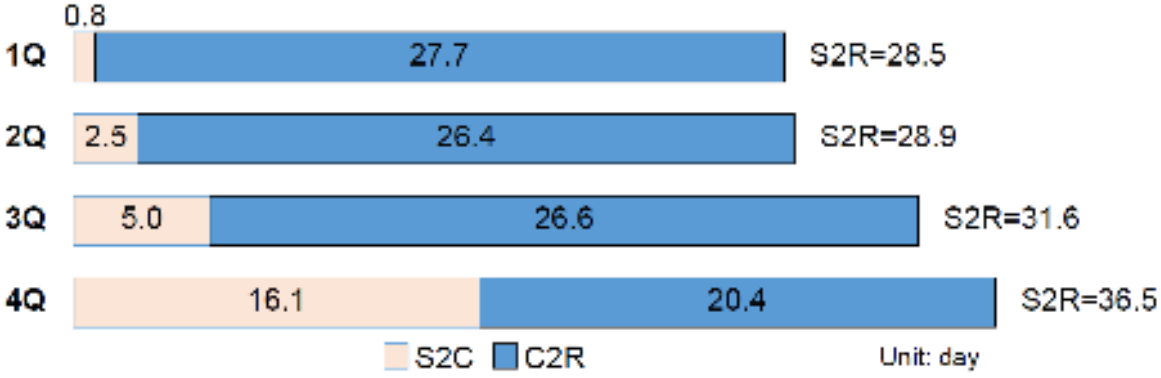
Table 1. Characteristics of confirmed COVID-19 cases (N=159) and their durations from symptom onset to confirmation, confirmation to release, and symptom onset to release.

Characteristic	Value, n (%)	Dependent variable		Independent variables			
		Duration of symptom onset to confirmation		Duration of confirmation to release		Duration of symptom onset to release	
		Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Total sample	N/A[a]	6.1 (8.9)	−1 to 48	25.3 (12.0)	4 to 72	31.4 (12.8)	7 to 73
Sex							
Male	67 (42.1)	6.7 (10.4)	−1 to 48	25.4 (12.7)	7 to 72	32.1 (13.4)	11 to 73
Female	92 (57.9)	5.6 (7.6)	0 to 47	25.2 (11.6)	4 to 69	30.8 (12.4)	7 to 70
Age (years)[b]							
0-19	8 (5.0)	7.9 (11.0)	−1 to 33	22.3 (12.9)	12 to 44	30.1 (15.7)	14 to 54
20-39	59 (37.1)	5.0 (6.6)	0 to 48	24.4 (12.1)	4 to 69	29.4 (13.3)	7 to 70
40-59	71 (44.7)	7.3 (11.0)	0 to 47	25.7 (11.7)	7 to 72	33 (11.9)	14 to 73
≥60	21 (13.2)	4.3 (3.6)	0 to 12	27.8 (12.9)	6 to 54	32.1 (13.6)	9 to 60

[a]N/A: not applicable.

[b]Mean 42 years (SD 14.7), range 1 to 81.

Figure 2. The duration from the confirmation date to release (C2R) and from the symptom onset to release (S2R) according to the quartile of the duration from symptom onset to confirmation (S2C) quartile. 1Q: ~1 day (mean age: 42.5 years), 2Q: 2–3 days (mean age: 43.8 years), 3Q: 4-6 days (mean age: 39.5 years), and 4Q: ≥7 days (mean age: 42.6 years). 1Q: first quartile; 2Q: second quartile; 3Q: third quartile; 4Q; fourth quartile; C2R: duration from confirmation date to release; S2C: duration from symptom onset to confirmation; S2R: duration from symptom onset to release.



We performed multiple regression analysis to investigate the association between rapid confirmation after symptom onset and the total prevalence period (faster release from isolation). Sex and age did not show a significant association with the duration from confirmation to release or the duration from symptom onset to release. However, the duration from symptom onset to confirmation showed a negative association with the duration from confirmation to release (t_1=−3.58; P<.001) and a positive association with the duration from symptom onset to release (t_1=5.86; P<.001); these associations were statistically significant (Table 2).

We measured the informative power of the duration from symptom onset to confirmation to predict early or late symptom onset to release. We found that the duration from symptom onset to confirmation resulted in improved performance for both models (P<.001 for both the early and late symptom onset

to release models) compared to the model that used age and sex as input features (Figure 3).

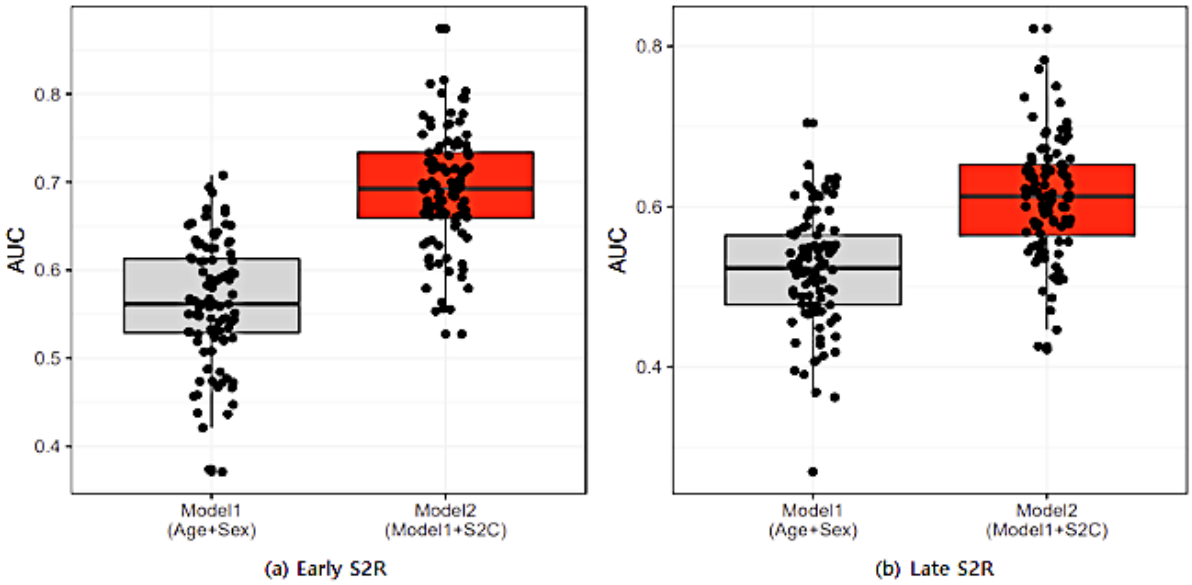
Table 2. Association between the durations from symptom onset to confirmation, confirmation to release, and symptom onset to release based on the results of multiple regression analysis.

Characteristics and categories	Duration from confirmation to release[a]		Duration from symptom onset to release[b]	
	T *(df)*	*P* value	T *(df)*	*P* value
Sex				
Male (reference) versus female	−0.55 (1)	.58	−0.55 (1)	.58
Age (years)				
≥60 (reference) versus 40-59	0.30 (1)	.76	0.30 (1)	.76
≥60 versus 20-39	0.83 (1)	.41	0.83 (1)	.41
≥60 versus 0-19	0.94 (1)	.34	0.94 (1)	.34
Duration of symptom onset to confirmation (days)	−3.58 (1)	<.001	5.86 (1)	<.001
F	2.99	.01	7.66	<.001

[a]R^2=0.089; adjusted R^2=0.059.

[b]R^2=0.200; adjusted R^2=0.174.

Figure 3. Predictive power determined through logistic regression after setting S2C and other variables (age, sex, and S2C). The analysis comprises 100 iterations by the random division of training and test data sets. AUC: area under the curve; S2C; duration from symptom onset to confirmation; S2R: duration from symptom onset to release.



Discussion

Principal Findings

Rapid diagnosis of confirmed cases enables prompt infection control and environmental decontamination of patients with suspected infection and places, such as environmental decontamination of related places and the management of close contacts [9]. Most of the studies published to date have tended to approach the early detection of COVID-19 from the perspective of prevention of the infection and spread of the disease [23]. Reducing the time to isolation through rapid diagnosis has been proven to help prevent the spread of SARS-CoV-2 infection [15,21,23]. In this context, studies published to date have claimed that early detection of COVID-19 is important because it can lower the mortality rate [24,25]. However, this claim is an inference based on general medical knowledge about pneumonia and complications and is not rooted in actual COVID-19 data.

Through linear regression analysis, we verified that the duration from symptom onset to confirmation is a factor that influences the duration from symptom onset to release. In addition, we used LR to test the predictors of duration from symptom onset to release, and we found that the inclusion of the duration from symptom onset to confirmation as a variable in the model significantly increased the predictive power. Therefore, the duration from symptom onset to confirmation is an important variable for predicting prevalence, and these results support the hypothesis that a short duration from symptom onset to confirmation can reduce the duration from symptom onset to release. Consequently, this study demonstrated that rapid diagnosis leads to faster release from isolation. If COVID-19 could be diagnosed early, the symptoms and progression of the

disease could be controlled with only simple treatment, and the reduced number of patients with severe disease could result in stable availability of hospital beds [20]. In the current COVID-19 pandemic, confirmed COVID-19 patients in many countries are not receiving proper treatment because of a shortage of medical resources; thus, efficient use of limited medical resources is extremely important. Under such circumstances, treating patients and releasing them from isolation as quickly as possible is one of the most important strategies for infectious disease control. Clinically, shortening the treatment period could be one of the best methods to improve the personal safety and quality of life of patients as well as the safety of medical staff and to reduce the workload of medical staff.

Any person with symptoms or suspicion of SARS-CoV-2 infection should be tested immediately, and intervention must take place as soon as possible upon confirmation. Delays in the confirmation of COVID-19 increase the social burden of the spread of this infectious disease but also contribute to the clinical burden of an increase in the prevalence period. The results of the regression analysis in this study showed that sex and age were unassociated with the prevalence and isolation periods of patients with COVID-19. These findings are not consistent with a previous study which reported that the prognosis of COVID-19 is influenced by age [10]. An earlier study focused on the qualitative aspect of patients' health or medical condition after treatment or release from isolation [10], whereas this study investigated the prevalence period. Therefore, a direct comparison is difficult because of the differences in the level of variables. The COVID-19-related mortality rate is higher among the older population [5,26]; therefore, rapid diagnosis is important for older persons. The older population has a high likelihood of having underlying diseases and onset of complications, whereas the risk of disease progression to severe conditions is also higher. Thus, higher mortality is to be expected and, therefore, early intervention through rapid confirmation is expected to reduce the mortality rate [27]. Consequently, considering the results of this and previous studies, rapid confirmation could shorten the prevalence period, lead to better prognosis, and reduce the mortality rate. Along with the advancement in the diagnosis of COVID-19, such as nucleic acid tests, among which the PCR method is considered as the "gold standard" for detection of the virus, rapid confirmation is very important to not only prevent the spread of disease but also enable better patient care.

This study has some limitations. First, we could not distinguish between severely and mildly ill patients, and it is possible that the prevalence period may differ between these two groups. Second, the study only considered a quantitative variable of duration of symptom onset to release, instead of clinical outcomes such as complications and posttreatment prognosis. Third, a sufficiently large sample size could not be obtained. Fourth, because we only used crawled data available on the web, other factors besides the duration from symptom onset to confirmation that could affect the duration from symptom onset to release could have resulted in uncorrected bias in the results of the study. Nonetheless, in this study, we obtained information on the confirmed diagnosis date for every patient, although many cases did not have a symptom onset date. This was due to the fact that many patients could not accurately remember their symptom onset date. To overcome these challenges, efforts to secure a record on the date of symptom onset in the clinical field are necessary, and additional studies should be conducted that consider disease severity and treatment.

Conclusion

The duration from symptom onset to the date of confirmation of a disease is an important variable for predicting prevalence, and our results support the hypothesis that a short duration from symptom onset to confirmation of COVID-19 can reduce the duration from symptom onset to release. Consequently, this study demonstrated that rapid diagnosis leads to faster release from isolation.

Data Availability

To obtain the processed data, please contact the authors to request the data.

Acknowledgments

This work was supported by the PaiChai University research grant in 2020.

Authors' Contributions

MBP initiated the idea and led the formal analysis. JHL and EUP initiated the idea and reviewed and edited the final draft of the article. TSL reviewed related articles and performed the statistical analysis. All authors reviewed and approved the final version of this article.

Conflicts of Interest

None declared.

References

1. Novel Coronavirus (2019-nCoV) Situation Report - 1. World Health Organization. 2020 Jan 21. URL: https://apps.who.int/iris/bitstream/handle/10665/330760/nCoVsitrep21Jan2020-eng.pdf?sequence=3&isAllowed=y [accessed 2021-07-27]

2. World Health Organization. Novel Coronavirus (2019-nCoV) Situation Report - 8. World Health Organization. 2020 Jan 28. URL: https://apps.who.int/iris/bitstream/handle/10665/330773/nCoVsitrep28Jan2020-eng.pdf?sequence=1&isAllowed=y [accessed 2021-07-27]
3. WHO Director-General's opening remarks at the media briefing on COVID-19 World Health Organization; 2020 updated 11 March 2020;. World Health Organization. 2020 Mar 11. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2021-07-27]
4. COVID-19 Dashboard. CoronaBoard. URL: https://coronaboard.kr/en/ [accessed 2021-07-27]
5. World Health Organization. Clinical management of severe acute respiratory infection (SARI) when COVID-19 disease is suspected. Interim guidance. Pediatr Med Rodz 2020 Jul 14;16(1):9-26. [doi: 10.15557/pimr.2020.0003]
6. Similarities and Differences between Flu and COVID-19. US Centers for Disease Control and Prevention. URL: https://tinyurl.com/4cb8659h [accessed 2021-07-27]
7. Kronbichler A, Kresse D, Yoon S, Lee KH, Effenberger M, Shin JI. Asymptomatic patients as a source of COVID-19 infections: a systematic review and meta-analysis. Int J Infect Dis 2020 Sep;98:180-186 [FREE Full text] [doi: 10.1016/j.ijid.2020.06.052] [Medline: 32562846]
8. Guidelines on temporary living/testing facilities for inbound travelers. Korea Disease Control and Prevention Agency. 2020 Jun 18. URL: http://ncov.mohw.go.kr/en/guidelineView.do?brdId=18&brdGubun=181&dataGubun=&ncvContSeq=2615&contSeq=2615&board_id=&gubun=# [accessed 2021-07-27]
9. Corman V, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu D, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro Surveill 2020 Jan;25(3) [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.3.2000045] [Medline: 31992387]
10. Liu J, Liu S. The management of coronavirus disease 2019 (COVID-19). J Med Virol 2020 Sep;92(9):1484-1490 [FREE Full text] [doi: 10.1002/jmv.25965] [Medline: 32369222]
11. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb;395(10223):497-506. [doi: 10.1016/s0140-6736(20)30183-5]
12. Ko W, Rolain J, Lee N, Chen P, Huang C, Lee P, et al. Arguments in favour of remdesivir for treating SARS-CoV-2 infections. Int J Antimicrob Agents 2020 Apr;55(4):105933 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.105933] [Medline: 32147516]
13. Ferrara F, Porta RL, D'Aiuto V, Vitiello A. Remdesivir and COVID-19. Ir J Med Sci 2020 Oct 17;190(3):1237-1238. [doi: 10.1007/s11845-020-02401-5]
14. Caly L, Druce J, Roberts J, Bond K, Tran T, Kostecki R, et al. Isolation and rapid sharing of the 2019 novel coronavirus (SARS-CoV-2) from the first patient diagnosed with COVID-19 in Australia. Med J Aust 2020 Jun;212(10):459-462 [FREE Full text] [doi: 10.5694/mja2.50569] [Medline: 32237278]
15. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. Lancet Global Health 2020 Apr;8(4):e488-e496. [doi: 10.1016/s2214-109x(20)30074-7]
16. Kim J, Chung Y, Jo HJ, Lee N, Kim MS, Woo SH, et al. Identification of coronavirus Isolated from a Patient in Korea with COVID-19. Osong Public Health Res Perspect 2020 Feb;11(1):3-7 [FREE Full text] [doi: 10.24171/j.phrp.2020.11.1.02] [Medline: 32149036]
17. Locatelli F, Vecchio LD, Pozzoni P. The importance of early detection of chronic kidney disease. Nephrol Dial Transplant 2002 Nov 01;17 Suppl 11(suppl 11):2-7. [doi: 10.1093/ndt/17.suppl_11.2] [Medline: 12386248]
18. Ott J, Ullrich A, Miller A. The importance of early symptom recognition in the context of early detection and cancer survival. Eur J Cancer 2009 Nov;45(16):2743-2748. [doi: 10.1016/j.ejca.2009.08.009] [Medline: 19765977]
19. Veenema TG, Tõke J. Early detection and surveillance for biopreparedness and emerging infectious diseases. Online J Issues Nurs 2006 Jan 31;11(1):3. [Medline: 16629504]
20. Peck K. Early diagnosis and rapid isolation: response to COVID-19 outbreak in Korea. Clin Microbiol Infect 2020 Jul;26(7):805-807 [FREE Full text] [doi: 10.1016/j.cmi.2020.04.025] [Medline: 32344168]
21. World Health Organization. Clinical management of severe acute respiratory infection (SARI) when COVID-19 disease is suspected. Interim guidance. Pediatr Med Rodz 2020 Jul 14;16(1):9-26. [doi: 10.15557/pimr.2020.0003]
22. All About Korea's Response to COVID-19. Korea Disease Control and Prevention Agency. URL: http://www.kdca.go.kr/board/board.es?mid=a30402000000&bid=0030 [accessed 2021-07-28]
23. Pang J, Wang MX, Ang IYH, Tan SHX, Lewis RF, Chen JI, et al. Potential rapid diagnostics, vaccine and therapeutics for 2019 novel coronavirus (2019-nCoV): a systematic review. J Clin Med 2020 Feb 26;9(3):623 [FREE Full text] [doi: 10.3390/jcm9030623] [Medline: 32110875]
24. Richardson S, Hirsch J, Narasimhan M, Crawford J, McGinn T, Davidson K, the Northwell COVID-19 Research Consortium, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. JAMA 2020 May 26;323(20):2052-2059 [FREE Full text] [doi: 10.1001/jama.2020.6775] [Medline: 32320003]
25. Goyal DK, Mansab F, Iqbal A, Bhatti S. Early intervention likely improves mortality in COVID-19 infection. Clin Med (Lond) 2020 May 01;20(3):248-250 [FREE Full text] [doi: 10.7861/clinmed.2020-0214] [Medline: 32357975]

26. Cascella M, Rajnik M, Cuomo A, Dulebohn S, Di NR. Features, evaluation and treatment of coronavirus (COVID-19). In: StatPearls. Treasure Island, FL: StatPearls Publishing; 2021.
27. Goyal DK, Mansab F, Iqbal A, Bhatti S. Early intervention likely improves mortality in COVID-19 infection. Clin Med (Lond) 2020 May 01;20(3):248-250 [FREE Full text] [doi: 10.7861/clinmed.2020-0214] [Medline: 32357975]

Abbreviations

LR: logistic regression
RT-PCR: real-time polymerase chain reaction
WHO: World Health Organization

Edited by G Eysenbach; submitted 14.04.21; peer-reviewed by H Lee; comments to author 10.05.21; revised version received 31.05.21; accepted 09.06.21; published 01.09.21.

Please cite as:
Park MB, Park EY, Lee TS, Lee J
Effect of the Period From COVID-19 Symptom Onset to Confirmation on Disease Duration: Quantitative Analysis of Publicly Available Patient Data
J Med Internet Res 2021;23(9):e29576
URL: https://www.jmir.org/2021/9/e29576
doi:10.2196/29576
PMID:34280114

Original Paper

Health Belief Model Perspective on the Control of COVID-19 Vaccine Hesitancy and the Promotion of Vaccination in China: Web-Based Cross-sectional Study

Hao Chen[1*], MSc; Xiaomei Li[1*], MSc; Junling Gao[1*], PhD; Xiaoxi Liu[1], BSc; Yimeng Mao[1], MSc; Ruru Wang[1], MPH; Pinpin Zheng[1], PhD; Qianyi Xiao[1], PhD; Yingnan Jia[1], PhD; Hua Fu[1], PhD; Junming Dai[1], PhD

Department of Preventive Medicine and Health Education, School of Public Health, Fudan University, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Junming Dai, PhD
Department of Preventive Medicine and Health Education
School of Public Health
Fudan University
No.138# Yixueyuan Road, Xuhui District
Shanghai, 200032
China
Phone: 86 021 54237358
Email: jmdai@fudan.edu.cn

Abstract

Background: The control of vaccine hesitancy and the promotion of vaccination are key protective measures against COVID-19.

Objective: This study assesses the prevalence of vaccine hesitancy and the vaccination rate and examines the association between factors of the health belief model (HBM) and vaccination.

Methods: A convenience sample of 2531 valid participants from 31 provinces and autonomous regions of mainland China were enrolled in this online survey study from January 1 to 24, 2021. Multivariable logistic regression was used to identify the associations of the vaccination rate and HBM factors with the prevalence of vaccine hesitancy after other covariates were controlled.

Results: The prevalence of vaccine hesitancy was 44.3% (95% CI 42.3%-46.2%), and the vaccination rate was 10.4% (9.2%-11.6%). The factors that directly promoted vaccination behavior were a lack of vaccine hesitancy (odds ratio [OR] 7.75, 95% CI 5.03-11.93), agreement with recommendations from friends or family for vaccination (OR 3.11, 95% CI 1.75-5.52), and absence of perceived barriers to COVID-19 vaccination (OR 0.51, 95% CI 0.35-0.75). The factors that were directly associated with a higher vaccine hesitancy rate were a high level of perceived barriers (OR 1.63, 95% CI 1.36-1.95) and perceived benefits (OR 0.51, 95% CI 0.32-0.79). A mediating effect of self-efficacy, influenced by perceived barriers (standardized structure coefficient [SSC]=−0.71, P<.001), perceived benefits (SSC=0.58, P<.001), agreement with recommendations from authorities (SSC=0.27, P<.001), and agreement with recommendations from friends or family (SSC=0.31, P<.001), was negatively associated with vaccination (SSC=−0.45, P<.001) via vaccine hesitancy (SSC=−0.32, P<.001).

Conclusions: It may be possible to increase the vaccination rate by reducing vaccine hesitancy and perceived barriers to vaccination and by encouraging volunteers to advocate for vaccination to their friends and family members. It is also important to reduce vaccine hesitancy by enhancing self-efficacy for vaccination, due to its crucial mediating function.

(J Med Internet Res 2021;23(9):e29329) doi:10.2196/29329

KEYWORDS

COVID-19 pandemic; vaccination behavior; vaccine hesitancy; health belief model

Introduction

COVID-19 has spread worldwide, causing more than 88 million infections and more than 1.9 million deaths as of January 2021 [1]. Due to the lack of effective treatments, the development and use of a new COVID-19 vaccine has become an important strategy to control the epidemic. Since COVID-19 broke out, according to the World Health Organization (WHO), 60 new coronavirus-inactivated vaccines and more than 10 nucleic acid vaccines, vector vaccines, and protein subunit vaccines have been developed [2]. Vaccination is recognized as the most successful and cost-effective public health intervention in the world today, and it has made a very large contribution to improving global health by reducing the incidence and deaths of many infectious diseases [3,4]. China and the whole world are experiencing the third wave of epidemics, so it is especially important to establish herd immunity by vaccinating against COVID-19 [5].

On December 30, 2020, the first homegrown COVID-19 vaccine in China was approved for marketing by the China National Medical Products Administration, and open volunteer vaccination to the public was announced through official media. On January 9, 2021, the National Health Commission promised free vaccinations for the Chinese population [6]. As of February 2021, the COVID-19 vaccine in China is suitable for people aged 18 to 59 years; the COVID-19 vaccine is not suitable for pregnant women, lactating women, and people with the following conditions: acute stages of fever, infections and other diseases, immune deficiency or immune disorders, serious liver and kidney diseases, hypertension, diabetic complications, and malignant tumors with uncontrolled drugs [7]. As of February 2021, common adverse reactions to vaccines in China mainly include headache, fever, local redness or lumps at the inoculation site, and cough, as well as loss of appetite, vomiting, and diarrhea in some people [8]. In first month of COVID-19 vaccinations, up to January 26, 2021, 22.8 million doses of the COVID-19 vaccine were administered in China, and less than 5% of the vaccine-eligible population among them, the main group, was at high risk of infection in all regions [9]. Considering the occupational exposure risk of COVID-19 infection, some populations with priority for vaccination were those with occupations at border ports, in key places such as international and domestic transportation, and in key industries such as medical and health care as well as basic social operation services. These populations are mass vaccinated on the basis of individual willingness [6]. According to the director of the National Health Commission, National Bureau of Disease Control and Prevention, all residents could be vaccinated in an orderly manner where there is an ample supply of vaccine and where vaccination units are health service centers, township health centers, or general hospitals located in their respective jurisdictions. Local governments have been required to make public in a timely manner the vaccination sites and units that can administer vaccines in their respective jurisdictions, including their locations and service hours [8]. From the beginning of 2020 to February 2021, the State Council Information Office has held regular press conferences to invite experts from relevant departments to brief the population on the joint prevention and control of COVID-19 [10]. Not only have the daily numbers of new cases, close contacts, patients who recovered, and patients who died been announced, but the latest number of vaccinated people as well as side effects and psychological changes following vaccination have also been announced. In addition to communications at the national level, provinces, municipalities, and autonomous regions released the latest information about the epidemic through various channels, such as press conferences or short videos, specific to their own situations, to ensure mastery and understanding of information regarding the epidemic and vaccines [11,12]. As of May 2021, 6 months after countries had begun carrying out vaccinations, the global number of COVID-19 vaccinations has exceeded 1.5 billion doses. Among them, nearly 60% are concentrated in China (420 million doses), the United States (270 million doses), and India (180 million doses). Except for a few countries with a vaccination rate exceeding 50% (eg, Israel), most countries in the world have a vaccination rate below 20% [13]. According to a recent study, predicted vaccine coverage of 55% to 82% of the population is needed to achieve COVID-19 herd immunity [5]. In addition to the supply of vaccines, individuals' psychological mechanisms of vaccine behavior are particularly critical to vaccination [14]. Therefore, it is of great significance to explore the possible influencing factors of individuals' vaccination willingness when vaccination rates are low in order to improve COVID-19 vaccination willingness and coverage in China and other parts of the world.

Although vaccines are currently an effective means of improving global health, in many parts of the world there are still quite a few people who question the necessity of vaccination, postpone vaccination, or even refuse vaccination; this is especially true when vaccines first came to market and were met with considerable hesitation and even outright opposition [15]. In 2012, the WHO established the Strategic Advisory Group of Experts (SAGE) working group to address and define vaccine hesitancy and its scope [16]. Vaccination hesitancy was defined as the refusal or delay of vaccination when vaccination services were available [17], and vaccination hesitancy was listed among the 10 threats to global health in 2019 [18]. Vaccine hesitancy is reflected in many factors, including confidence in the efficacy and safety of the vaccine and in the health service system providing the vaccine, such as the reliability and competence of the health service system and the professionals involved in the vaccination service [19]. In the first month after vaccines became available to all vaccine-eligible members of the Chinese population, a nationwide cross-sectional study reported the prevalence of COVID-19 vaccination hesitancy to be 35.5%. After an instance of illegal marketing of vaccines, 32.4% of parents became hesitant of vaccines [20]; rapid sociocultural changes have also contributed to vaccine hesitancy [21,22]. A study on COVID-19 vaccine hesitancy of Italian college students showed that among the 735 students who answered questions about their vaccination intentions, more than 1 in 10 students showed hesitancy [23]. An investigation during Israel's mandatory quarantine revealed that nurses and medical workers showed high levels of vaccine hesitancy [24]. According to a literature review, 68.4% of the global population is willing to receive the vaccination [25].

Recent studies of factors associated with COVID-19 vaccination have identified a number of demographic, cognitive, and psychosocial factors, including age, gender, educational level, insurance status, attitudes toward the vaccine, confidence in government information, perceived susceptibility to COVID-19, and perceived benefits and side effects of the vaccine [26,27]. In the current age of Web 2.0, the spread of false news about vaccine safety and validity on social media, such as that COVID-19 vaccination can affect individuals' reproductive function, influence vaccination willingness and confidence [28]. Several typical behavioral theories, such as the health belief model (HBM), the theory of planned behavior (TPB) [29], and the diffusion of innovation theory (DIT) [14], have been used to explain COVID-19 vaccination intent combined with demographic, cognitive, and psychosocial factors. The HBM is a widely used theory that proposes a variety of psychological factors that affect people's health protective behaviors, such as attitudes, beliefs, and intentions [30-32]. The HBM assumes that health-related actions depend on the simultaneous occurrence of three factors [33]: (1) the presence of sufficient motivation (or health concern) to make the health problem salient or relevant, (2) the belief that a person is vulnerable to serious health problems or the sequelae of that illness or condition is often referred to as a perceived threat, and (3) believing that following a specific health recommendation will help reduce the perceived threat at a subjectively acceptable cost. The TPB assumes that an individual's behavioral posture, activity attraction, and behavioral control jointly affect and direct the individual's behavior [34]. The DIT aims to disseminate innovation awareness, technology, or innovative ideas related to the masses, so that patients can develop innovative thinking or health awareness. In recent years, the DIT has been gradually introduced into medical and health industries, mainly for the guidance of health education strategies [14]. The HBM has been one of the most widely used theories in understanding health and illness behaviors, and due to its design, it has been previously used in vaccination studies to identify behavior relationships [35,36]. When compared with other models that explain behavior and resulting actions, the HBM was specifically developed to focus on preventative health research [35-38], which has been modified since its early use in the 1950s to be more inclusive and encourage interventions that improve health behaviors [39]. Thus, the HBM was chosen as the preferred model to investigate intention and behavior regarding COVID-19 vaccination. There are six main components of the HBM: perceived susceptibility, perceived severity, perceived benefits, perceived barriers, self-efficacy for health protective behaviors, and cues to action [40]. Previous studies, including those on H1N1 [41], hepatitis [42], human papillomavirus (HPV) [43], and measles [44], have identified HBM factors as important predictors of vaccination intentions. Therefore, it is necessary to explore the possible influence of these factors on people's willingness to vaccinate against COVID-19 in order to improve individual immunity and slow the epidemic. Although the aforementioned studies suggested that there were associations between HBM constructs and vaccine acceptance or hesitancy, relatively few studies have focused on COVID-19 vaccination behavior, especially in China and other countries where vaccinations are available to the domestic population [45].

In summary, we explored whether HBM constructs were associated with vaccine hesitancy and vaccination at the time when COVID-19 vaccination became available to the public in mainland China. A previous study identified that vaccine intention and willingness were important predictors of vaccination behavior, with more than 50% of the explained variance in influenza [46] and HPV [34] vaccinations. However, a gap seems to exist between intention and vaccination behavior [47], such as the willingness of students to receive the HPV vaccine predicting less than 10% of actual vaccinations [34]. Our first hypothesis (Hypothesis 1) was that vaccine hesitancy was negatively associated with COVID-19 vaccination behavior. In particular, we examined our major hypothesis (Hypothesis 2), which was that the HBM constructs of perceived barriers, self-efficacy, and cues to action would predict vaccine hesitancy and vaccination behavior. As in a previous study, self-efficacy is defined as the confidence in one's ability to facilitate decisions to carry out a health behavior such as vaccination, which is useful only to the extent that one feels one can adequately implement the steps needed to perform the behavior [48]. Evidence based on the HBM poses several mechanisms regarding how self-efficacy is associated with vaccine intention and behaviors. Self-efficacy was able to mediate the relationship between perceived barriers to HPV vaccination and HPV vaccine intentions among young women [49]. A similar mediation effect was found in the association between perceived severity and susceptibility and the intent to receive the Zika vaccine [50]. It was also suggested that self-efficacy could influence the path from cues to action (eg, physician recommendation, family members recommendation, media coverage, and public health communication) to HPV vaccine uptake [51] and acceptance of the H1N1 vaccine [52]. The aforementioned studies suggested our third major hypothesis (Hypothesis 3), which was that self-efficacy of the COVID-19 vaccine would mediate the influence of other HBM constructs on vaccine hesitancy and vaccination.

Methods

Study Design and Participants

From January 1 to 24, 2021, we used convenience and snowball sampling to recruit a sample of 2580 participants from 31 out of a total of 34 provinces and autonomous regions in China, with each area consisting of at least 30 participants; we then conducted a web-based cross-sectional study. A digital questionnaire link was sent to a WeChat "Friends circle," a function that can be used to share personal photos or public website links in one's "Moments" to make them visible to friends on platforms such as Twitter and Facebook. This questionnaire link, on the Wenjuanxing platform, could then be forwarded or shared by participants with friends in their WeChat contact list whom they considered appropriate for this survey; their friends were also encouraged to send the link to their friend networks. The snowball sampling process continued until a sufficient sample size was reached. The first page of the questionnaire contained an electronic consent form. Each

respondent received a small monetary reward of ¥5 (a currency exchange rate of ¥1=US $0.15 is applicable) after authentically completing the questionnaire, which took approximately 5 to 10 minutes. To prevent repeated entries from the same individual, who may attempt multiple entries for the enrollment reward, additional measures were adopted: (1) the same IP address was only allowed to be used once to fill in the questionnaire, which was a built-in function of the Wenjuanxing platform, and (2) participants were only allowed to fill in questionnaires after logging in to their WeChat accounts—they needed to register with this platform with a personal identity card—and each WeChat account could only be used once to fill in the questionnaire. The minimum sample size was calculated to be 1100 by using the following formula:

where the latest reported prevalence of COVID-19 vaccination hesitancy (p) was 35.5%, based on research that was conducted in China, nationwide, from January 10 to January 22, 2021 [53]. The type I error () was .05; thus, $z_{1-/2}$=1.96, the precision (d) was 0.04, and the design effect (deff) was 2 [54]. The inclusion criteria for participants' enrollment were as follows: (1) aged 18 to 59 years, (2) able to understand the questionnaire by themselves, and (3) could use online services, such as mobile phones, computers, and tablet computers. The questionnaires of participants who met the following exclusion criteria were discarded: (1) aged less than 18 years (n=16) or more than 59 years (n=32) and not eligible for vaccination until April 2021 in China and (2) returned invalid questionnaires (n=32). Questionnaires were deemed invalid if the following occurred: (1) participant gave one or two wrong answers to two quality control questions, including "Where is capital of China?" and "What's three plus five?"; (2) occurrence of a logic check result error, which occurred when the participant selected both "no disease" and "any type of disease" in response to the question "Do you have any type of the following diseases or diagnosed medical histories"; and (3) participant took less than the minimum time of 3 minutes to complete the questionnaire. Cognitive interviewing with 5 subjects was done to refine the questionnaires through the web-based platform WeChat. Participants were required to respond to each item by answering three questions: (1) "What does '……' mean to you?", (2) "Can you repeat this question in your own words?", and (3) "When you think about '……' what comes to your mind?" We also asked participants to answer three questions for the overall survey, including the following: (1) "Are there additional questions you believe should be asked?", (2) "Are there questions you believe should be deleted?", and (3) "Are there questions you believe should be modified?" The entire questionnaire was tested and modified to appropriately conduct the survey. Finally, 2531 participants were included in this study. All participants consented to written ethics approval before the survey was conducted. This study was approved by the Institutional Review Board of Fudan University, School of Public Health (IRB00002408&FWA00002399), and approval expired on March 3, 2021.

Measurements

Vaccine Hesitancy and Vaccination

Vaccine hesitancy was assessed with a one-item self-report measure that quantified the demand for, and acceptance of, vaccination: "How willing would you be to get the COVID-19 vaccine?" The respondents were asked to answer the question using the following 7-point scale recommended by the SAGE working group on vaccine hesitancy: "accept all [vaccines]," "accept but unsure," "accept some," "delay," "refuse some," "refuse but unsure," and "refuse all" [17]. Vaccine hesitancy was defined as any response on the scale except for "accept all" or "accept but unsure." Vaccination was assessed by asking the participants to answer "yes" or "no" to a single question: "Have you gotten the COVID-19 vaccine?"

Health Belief Model

Items derived from the HBM were adopted from a previous study or modified to measure the participants' beliefs about COVID-19 vaccination. Five essential dimensions of health beliefs were measured as follows: (1) perceived susceptibility to COVID-19 in the future (three items; eg, "I was vulnerable to infection with SARS-CoV-2"), (2) perceived severity of COVID-19 infection (four items; eg, "It would be very harmful for me if I got COVID-19"), (3) perceived benefits of COVID-19 vaccination (three items; eg, "COVID-19 vaccination can protect me from infection with SARS-CoV-2"), (4) perceived barriers to COVID-19 vaccination (six items; eg, "The COVID-19 vaccine might have side effects, such as fever or soreness in the arm"), and (5) self-efficacy for COVID-19 vaccination (five items; eg, "I believe I can deal with side effects of the COVID-19 vaccine with doctors' help"). Cues to action refer to external recommendations that might affect individuals' health-related behaviors. In this study, the Cronbach α coefficients indicating internal consistency (ie, reliability) were .78 for the total HBM factors, .84 for perceived susceptibility to COVID-19, .80 for perceived severity of COVID-19 infection, .83 for perceived benefits of COVID-19 vaccination, .80 for perceived barriers to COVID-19 vaccination, and .82 for self-efficacy for COVID-19 vaccination. The sampling adequacy for the HBM factor scale was excellent (Kaiser-Meyer-Olkin=0.82). Inter-item correlations were sufficiently large for principal component analysis (PCA) (Bartlett test of sphericity: $^{2}_{210}$=23,122.6, P<.001). The PCA revealed five factors, which in combination explained 68.58% of the variance, and each factor accounted for 24.23%, 20.55%, 10.32%, 8.16%, and 5.32% of the explained variance, respectively. An examination of the factor loadings after rotation suggested, as expected, that factor 1 (perceived barriers to COVID-19 vaccination) had six items with loading factors between 0.74 and 0.79, factor 2 (self-efficacy for COVID-19 vaccination) included five items with loading factors between 0.71 and 0.80, factor 3 (perceived severity of COVID-19 infection) included four items with loading factors between 0.67 and 0.85, factor 4 (perceived benefits of COVID-19 vaccination) included three items with loading factors between 0.68 and 0.85, and factor 5 (perceived susceptibility to COVID-19) included three items with loading factors between 0.78 and 0.89.

External Cues to Action

External cues to action were assessed based on four cues used in previous surveys [36,55]: recommendations from authorities, recommendations from friends or family, vaccination of authorities, and vaccination of friends or family. Participants were asked to state their level of agreement with each of the statements, with a score of 1 for positive responses (strongly agree or agree) and a score of 0 for neutral or negative responses (neither agree nor disagree, disagree, or strongly disagree). The Cronbach α coefficient for cues to action was .82. The sampling adequacy for the cues to action scale was excellent (Kaiser-Meyer-Olkin=0.75). Inter-item correlations were sufficiently large for PCA (Bartlett test of sphericity: $^{2}_{6}$=2829.1, *P*<.001). The PCA revealed a single factor, which in combination explained 59.72% of the variance, and an examination of the factor loadings after rotation suggested, as expected, that the single factor included four items whose loading factors were between 0.65 and 0.84.

Demographic and Health-Related Characteristics

Demographic characteristics in this study included gender, age, educational level (high school degree and below, bachelor's degree, or master's degree and above), marital status (married or not married [including unmarried, divorced, and widowed]), occupation (medical worker or nonmedical worker), region (urban or rural), monthly salary (<¥6000, ¥6000-¥10,000, or >¥10,000), and family members with backgrounds in medical work or with medical education (yes or no). Health-related characteristics included self-rated health and self-reported chronic diseases having been diagnosed by doctors. Self-rated health was evaluated by a single question: "How is your perceived health in general?"; responses included "excellent," "very good," "good," "general," or "poor" [56]. We listed 16 common chronic diseases, such as hypertension and diabetes, and categorized the number of reported chronic diseases into 0, 1 or 2, and 3 or over.

Statistical Analysis

Frequencies were first calculated for all variables, and the prevalence and 95% CIs of vaccine hesitancy and vaccination were determined according to the participants' demographics, health-related characteristics, and HBM factors. Multivariable logistic regression analyses were used to explore the demographic and health-related characteristics (Table 1) as well as the HBM factors (Table 2) associated with vaccine hesitancy and vaccination. We then ran the multivariable logistic regression again to determine the HBM factors associated with vaccine hesitancy and vaccination after controlling for covariates (ie, demographic and health-related characteristics), with a significance level of *P*<.05. Odds ratios (ORs) with 95% CIs were calculated for each independent variable and were visualized in forest plots (Figures 1 and 2). All of the analyses were performed using SAS software, version 9.4 (SAS Institute Inc), and all tests were two-tailed with a significance level of *P*<.05. We used the forest plot package in R software, version 3.5.3 (The R Foundation), to generate the forest plots. We used Mplus, version 8.4 (Muthén & Muthén), to establish structural equation modeling (SEM) and to assess the standardized structure coefficients (SSCs) among the HBM factors of vaccine hesitancy and vaccination. The mean- and variance-adjusted weighted least squares method was employed as the method of estimation because the analyses included categorical endogenous variables (ie, vaccine hesitancy and vaccination), and the link was the probability unit in the current model [57]. We freed covariances between error terms based on their modification indices during the estimation process to improve model fit. The most common indices and acceptable reference values included the magnitude of χ^2 divided by its degrees of freedom (χ^2/df <5), the comparative fit index (CFI >0.90), the Tucker-Lewis index (TLI >0.90), and the root mean square error of approximation (RMSEA <0.08), which were used to determine whether the data fit the model [58].

Figure 1. Associations between the health belief model and vaccine hesitancy.

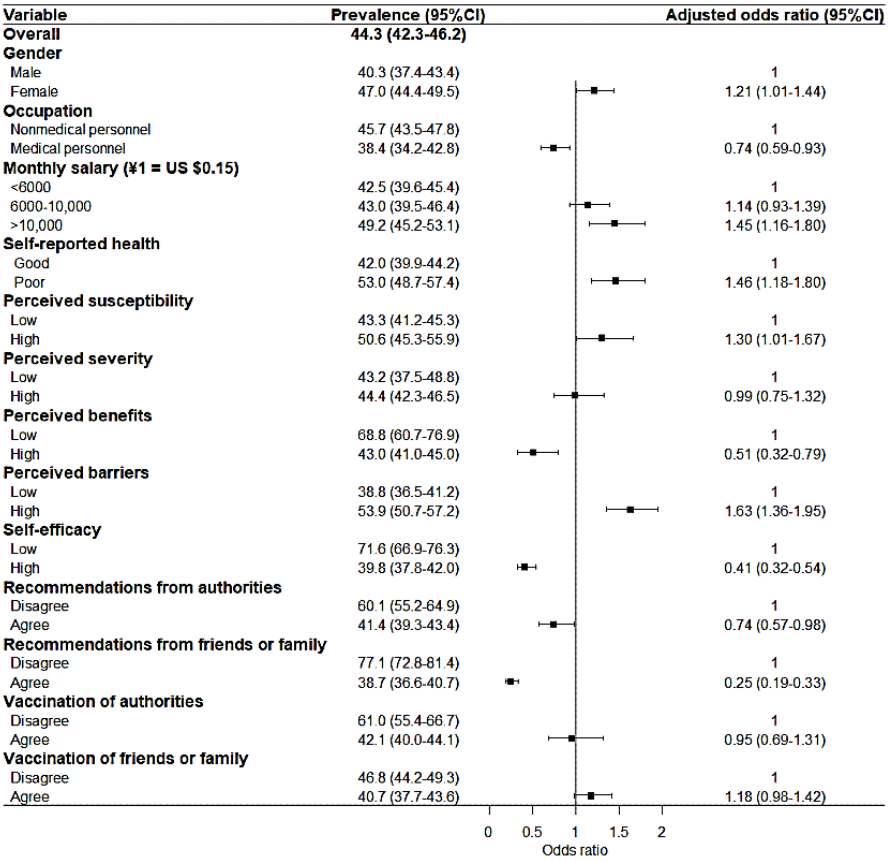
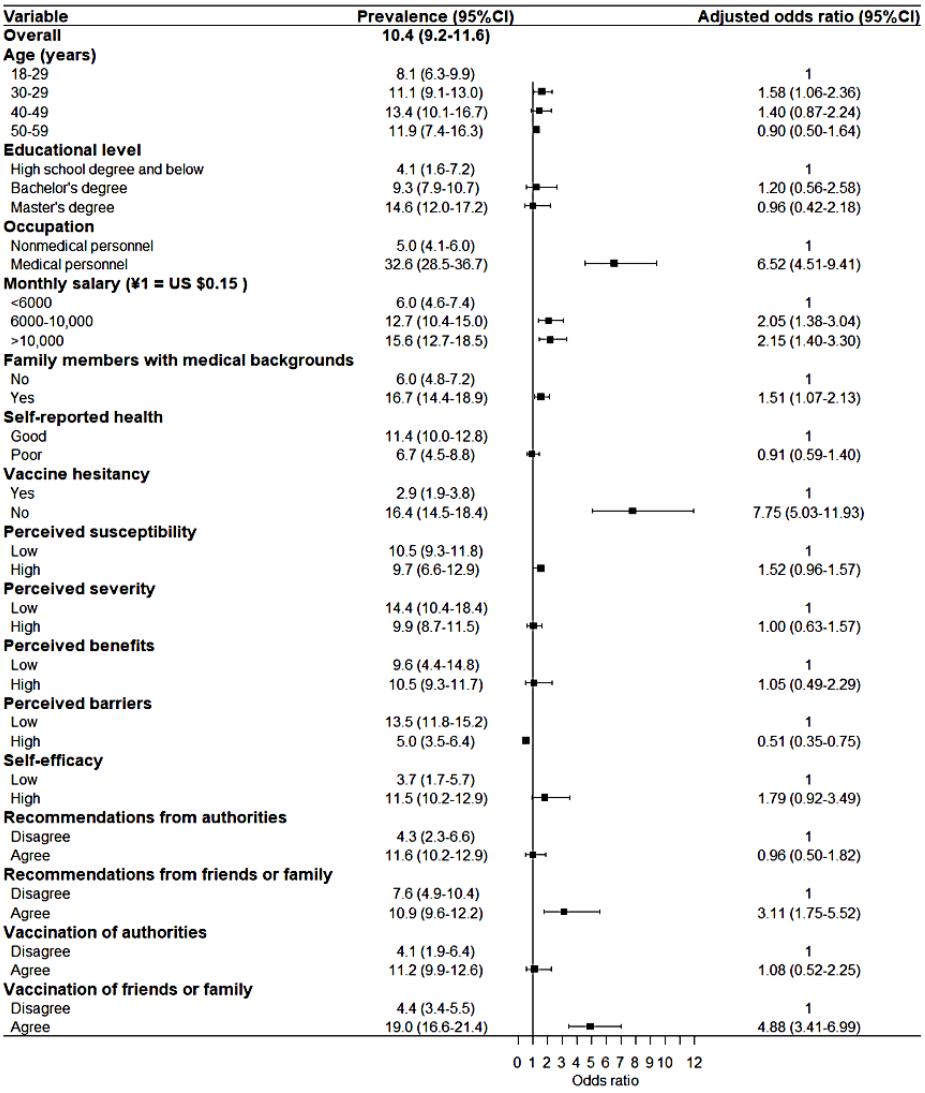


Figure 2. Associations between the health belief model and vaccination rate.



Results

Participant Characteristics

Our analysis included 2531 participants aged between 18 and 59 years (mean 33.92 years, SD 8.94); 58.7% (1486/2531) of the participants were female. Most of the participants were married (1660/2531, 65.6%), had a bachelor's degree (1609/2531, 63.6%), were nonmedical personnel (2034/2531, 80.4%), lived in urban areas (2262/2531, 89.4%), reported good health (2020/2531, 79.8%), and did not have chronic diseases (1617/2531, 63.9%). Slightly less than half of the participants reported monthly salaries lower than ¥6000 (1128/2531, 44.6%) and had family members with medical personnel backgrounds (1056/2531, 41.7%) (Table 1).

Table 1. Distribution of vaccine hesitancy and vaccination rate by participant demographics and health-related characteristics.

Characteristics	Participants (N=2531), n (%)	Vaccine hesitancy			Vaccination rate		
		Vaccine hesitancy, n (%)	OR[a] (95% CI)	*P* value	Vaccination, n (%)	OR (95% CI)	*P* value
Age (years)							
18-29	926 (36.6)	412 (44.5)	1		75 (8.1)	1	
30-39	993 (39.2)	467 (47.0)	1.11 (0.93-1.33)	.27	110 (11.1)	1.41 (1.04-1.92)	.03
40-49	410 (16.2)	163 (39.8)	0.82 (0.65-1.04)	.11	55 (13.4)	1.76 (1.22-2.54)	.003
50-59	202 (8.0)	78 (39.6)	0.79 (0.58-1.07)	.13	24 (11.9)	1.53 (0.94-2.49)	.09
Gender							
Male	1045 (41.3)	422 (40.4)	1		116 (11.1)	1	
Female	1486 (58.7)	698 (47.0)	1.31 (1.11-1.53)	.001	148 (10.0)	0.89 (0.69-1.15)	.36
Marital status							
Married	1660 (65.6)	725 (43.7)	1		187 (11.3)	1	
Not married	871 (34.4)	395 (45.4)	1.07 (0.91-1.26)	.42	77 (8.8)	0.76 (0.58-1.01)	.06
Educational level							
High school degree and below	204 (8.0)	83 (40.7)	1		9 (4.4)	1	
Bachelor's degree	1609 (63.6)	725 (45.1)	1.20 (0.89-1.61)	.24	150 (9.3)	2.23 (1.12-4.43)	.02
Master's degree and above	718 (28.4)	312 (43.5)	1.12 (0.82-1.54)	.48	105 (14.6)	3.71 (1.84-7.47)	<.001
Occupation							
Nonmedical personnel	2034 (80.4)	929 (45.7)	1		102 (5.0)	1	
Medical personnel	497 (19.6)	191 (38.4)	0.74 (0.61-0.91)	.004	162 (32.6)	9.16 (6.97-12.04)	<.001
Region							
Urban	2262 (89.4)	1016 (44.9)	1		242 (10.7)	1	
Rural	269 (10.6)	104 (38.7)	0.77 (0.60-1.00)	.05	22 (8.2)	0.74 (0.47-1.17)	.20
Monthly salary (¥)[b]							
<6000	1128 (44.6)	478 (42.5)	1		68 (6.0)	1	
6000-10,000	787 (31.1)	338 (43.0)	1.02 (0.85-1.23)	.83	100 (12.7)	2.27 (1.64-3.13)	<.001
>10,000	616 (24.3)	303 (49.2)	1.31 (1.08-1.60)	.007	96 (15.6)	2.88 (2.07-3.99)	<.001
Family members with medical backgrounds							
No	1475 (58.3)	667 (45.2)	1		88 (6.0)	1	
Yes	1056 (41.7)	453 (42.9)	0.91 (0.78-1.06)	.25	176 (16.7)	3.13 (2.28-4.17)	<.001
Self-reported health							
Good	2020 (79.8)	849 (42.0)	1		230 (11.4)	1	
Poor	511 (20.2)	271 (53.0)	1.56 (1.28-1.89)	<.001	34 (6.7)	0.55 (0.38-0.81)	.002
Number of chronic diseases							
0	1617 (63.9)	688 (42.3)	1		178 (11.0)	1	
1	639 (25.4)	300 (47.0)	0.84 (0.70-1.01)	.045	64 (10.0)	1.11 (0.82-1.50)	.49
2 and above	288 (11.2)	136 (49.5)	1.13 (0.86-1.49)	.49	22 (8.0)	0.78 (0.47-1.30)	.34
Vaccine hesitancy							
Yes	1120 (44.3)	N/A[c]	N/A	N/A	32 (2.9)	1	
No	1411 (55.7)	N/A	N/A	N/A	232 (16.4)	6.69 (4.58-9.77)	<.001

[a]OR: odds ratio.

[b]A currency exchange rate of ¥1=US $0.15 is applicable.

[c]N/A: not applicable.

Distribution of Vaccine Hesitancy and Vaccination by Participant Characteristics and Health Belief Model Factors

Overall, 44.3% (1120/2531; 95% CI 42.3%-46.2%) of the participants were classified as vaccine hesitant: 1.4% responded "refuse all," 5.3% responded "refuse but unsure," 3.7% responded "refuse some," 18.8% responded "delay," and 15.1% responded "accept some." Overall, 55.7% (1411/2531) of the participants were classified as vaccine accepting: 25.1% responded "accept but unsure" and 30.6% responded "accept all." Only 10.4% (264/2531; 95% CI 9.2%-11.6%) of the participants had been vaccinated for COVID-19, while the majority (2267/2531, 89.6%) had not been.

According to the multivariable logistic regression analyses including participant characteristics (Table 1), the participants were more likely to be vaccine hesitant if they were female (OR 1.31, 95% CI 1.11-1.53), were nonmedical personnel (OR 1.35, 95% CI 1.10-1.64), had poor self-rated health (OR 1.56, 95% CI 1.28-1.89), or had a monthly salary over ¥10,000 (OR 1.31, 95% CI 1.08-1.60). The participants were more likely to have been vaccinated if they were 30 to 39 years old (OR 1.41, 95% CI 1.04-1.92), had a bachelor's degree (OR 2.23, 95% CI 1.12-4.43), had a master's degree and above (OR 3.71, 95% CI 1.12-4.43), were medical personnel (OR 3.71, 95% CI 1.12-4.43), had good self-rated health (OR 1.82, 95% CI 1.23-2.63), were not vaccine hesitant (OR 6.69, 95% CI 4.58-9.77), had a monthly salary between ¥6000 and ¥10,000 (OR 2.27, 95% CI 1.64-3.13), had a monthly salary over ¥10,000 (OR 2.88, 95% CI 2.07-4.17), or had family members with medical personnel backgrounds (OR 3.13, 95% CI 2.28-4.17).

According to the multivariable regression analyses including the HBM factors (Table 2), the participants were more likely to be vaccine hesitant if they had high perceived susceptibility to COVID-19 (OR 1.34, 95% CI 1.07-1.69) or had high perceived barriers to vaccination (OR 1.84, 95% CI 1.56-2.17). The participants were less likely to be vaccine hesitant if they had high perceived benefits of vaccination (OR 0.34, 95% CI 0.23-0.50), had high self-efficacy for vaccination (OR 0.26, 95% CI 0.20-0.34), agreed with recommendations from authorities (OR 0.47, 95% CI 0.38-0.58), agreed with recommendations from friends or family (OR 0.19, 95% CI 0.14-0.24), agreed with the vaccination of authorities (OR 0.46, 95% CI 0.36-0.60), or agreed with the vaccination of friends or family (OR 0.77, 95% CI 0.66-0.91). The participants were more likely to have been vaccinated if they had high self-efficacy for vaccination (OR 3.39, 95% CI 1.92-6.00), agreed with recommendations from authorities (OR 2.89, 95% CI 1.75-4.78), agreed with the vaccination of authorities (OR 2.94, 95% CI 1.62-5.31), or agreed with the vaccination of friends or family (OR 5.05, 95% CI 3.77-6.76).

Table 2. Distribution of vaccine hesitancy and vaccination by health belief model (HBM) factors and cues to action.

HBM factors and cues to action	Participants (N=2531), n (%)	Vaccine hesitancy			Vaccination		
		Vaccine hesitancy, n (%)	OR[a] (95% CI)	*P* value	Vaccination, n (%)	OR (95% CI)	*P* value
Perceived susceptibility							
Low	2191 (86.6)	948 (43.3)	1		231 (10.4)	1	
High	340 (13.4)	172 (50.6)	1.34 (1.07-1.69)	.01	33 (9.7)	0.91 (0.62-1.34)	.64
Perceived severity							
Low	292 (11.5)	126 (43.2)	1		42 (14.4)	1	
High	2239 (88.5)	994 (44.4)	1.05 (0.82-1.36)	.69	222 (9.9)	0.66 (0.46-0.93)	.02
Perceived benefits							
Low	125 (4.9)	86 (68.8)	1		12 (9.6)	1	
High	2406 (95.1)	1034 (43.0)	0.34 (0.23-0.50)	<.001	252 (10.5)	1.01 (0.95-1.07)	.76
Perceived barriers							
Low	1622 (64.1)	630 (38.8)	1		219 (13.5)	1	
High	909 (35.9)	490 (53.9)	2.08 (1.77-2.45)	<.001	49 (5.0)	0.33 (0.24-0.47)	<.001
Self-efficacy							
Low	352 (13.9)	252 (71.6)	1		13 (3.7)	1	
High	2179 (86.1)	868 (39.8)	0.26 (0.20-0.34)	<.001	251 (11.5)	3.39 (1.92-6.00)	<.001
Recommendations from authorities							
Disagree	393 (15.5)	236 (60.1)	1		17 (4.3)	1	
Agree	2138 (84.5)	884 (41.4)	0.47 (0.38-0.58)	<.001	247 (11.6)	2.89 (1.75-4.78)	<.001
Recommendations from friends or family							
Disagree	367 (14.5)	283 (77.1)	1		28 (7.6)	1	
Agree	2164 (85.5)	837 (38.7)	0.19 (0.14-0.24)	<.001	236 (10.9)	1.48 (0.99-2.23)	.06
Vaccination of authorities							
Disagree	290 (11.5)	177 (61.0)	1		12 (4.1)	1	
Agree	2241 (88.5)	943 (42.1)	0.46 (0.36-0.60)	<.001	252 (11.2)	2.94 (1.62-5.31)	<.001
Vaccination of friends or family							
Disagree	1488 (58.8)	696 (46.8)	1		66 (4.4)	1	
Agree	1043 (41.2)	424 (40.7)	0.77 (0.66-0.91)	.002	198 (19.0)	5.05 (3.77-6.76)	<.001

[a]OR: odds ratio.

Influencing Factors of Vaccine Hesitancy and Vaccination

We included the participant characteristics and HBM factors in the vaccine hesitancy logistic regression, and the influencing factors are shown in Figure 1. The risk factors for vaccine hesitancy were female gender (*P*=47.0%, 95% CI 44.4%-49.5%; OR 1.12, 95% CI 1.01-1.44), monthly salary over ¥10,000 (*P*=49.2%, 95% CI 45.2%-53.1%; OR 1.45, 95% CI 1.16-1.80), poor self-rated health (*P*=53.0%, 95% CI 48.7%-57.4%; OR 1.46, 95% CI 1.18-1.80), high perceived susceptibility to COVID-19 (*P*=50.6%, 95% CI 45.3%-55.9%; OR 1.30, 95% CI 1.01-1.67), and high perceived barriers to vaccination (*P*=53.9%, 95% CI 50.7%-57.2%; OR 1.63, 95% CI 1.36-1.95). Additionally, the protective factors against vaccine hesitancy were occupation as medical personnel (*P*=38.4%, 95% CI 34.2%-42.7%; OR 0.74, 95% CI 0.59-0.93), high perceived benefits of vaccination (*P*=43.0%, 95% CI 41.0%-45.0%; OR 0.51, 95% CI 0.32-0.79), high self-efficacy for vaccination (*P*=38.4%, 95% CI 48.7%-57.4%; OR 1.46, 95% CI 1.18-1.80), agreement with recommendations from authorities (*P*=41.4%, 95% CI 39.3%-43.4%; OR 0.74, 95% CI 0.57-0.98), and agreement with recommendations from friends or family (*P*=41.4%, 95% CI 39.3%-43.4%; OR 0.74, 95% CI 0.57-0.98).

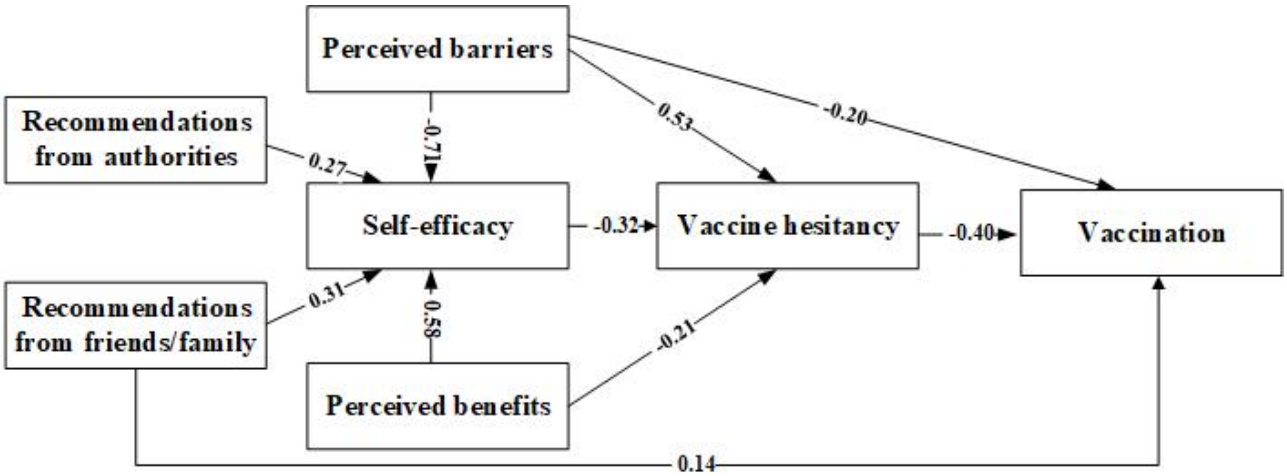
We included the participant characteristics, the HBM factors, and vaccine hesitancy in the vaccination logistic regression, and the influencing factors are shown in Figure 2. The promoting factors for vaccination were occupation as medical personnel (*P*=32.6%, 95% CI 28.5%-36.7%; OR 6.52, 95% CI 4.51-9.41),

monthly salary between ¥6000 and ¥10,000 (*P*=12.7%, 95% CI 10.4%-15.0%; OR 2.05, 95% CI 1.38-3.04), monthly salary over ¥10,000 (*P*=15.6%, 95% CI 12.7%-18.5%; OR 2.15, 95% CI 1.40-3.30), family members with medical personnel backgrounds (*P*=16.7%, 95% CI 14.4%-18.9%; OR 1.51, 95% CI 1.07-2.13), a lack of vaccine hesitancy (*P*=16.4%, 95% CI 14.5%-18.4%; OR 7.75, 95% CI 1.01-1.67), agreement with recommendations from friends or family (*P*=10.9%, 95% CI 9.6%-12.2%; OR 3.11, 95% CI 1.75-5.52), and agreement with the vaccination of friends or family (*P*=19.0%, 95% CI 19.6%-21.4%; OR 4.88, 95% CI 3.41-6.99). Additionally, a lower vaccination rate was associated with higher perceived barriers to COVID-19 vaccination (p=5.0%, 95% CI 3.5%-6.4%; OR 0.51, 95% CI 0.35-0.75).

Structural Equation Modeling of Vaccination

We used SEM to examine the underlying psychological mechanism of vaccination behavior (Figure 3). Based on the goodness-of-fit statistics, SEM showed a better fit to the data than the regression models (χ^2/df=4.62; RMSEA=0.05; CFI = 0.95; TLI = 0.91), and all of the paths were statistically significant (*P*<.05). The findings suggested that a mediating effect of self-efficacy, influenced by perceived barriers (SSC=−0.71, *P*<.001), perceived benefits (SSC=0.58, *P*<.001), agreement with recommendations from authorities (SSC=0.27, *P*<.001), and agreement with recommendations from friends or family (SSC=0.31, *P*<.001), was negatively associated with vaccination (SSC=−0.45, *P*<.001) via vaccine hesitancy (SSC=−0.32, *P*<.001). Additionally, perceived barriers (SSC=0.53, *P*<.001) and perceived benefits (SSC=−0.21, *P*<.001) were directly associated with vaccine hesitancy. Perceived barriers (SSC=−0.20, *P*<.001) and recommendations from friends or family (SSC=0.14, *P*<.001) were directly correlated with vaccination behavior.

Figure 3. The paths among vaccine hesitancy, vaccination, and health belief model factors. The numbers on the lines are the standardized structure coefficients.



Availability of Data and Materials

The data that support the findings of this study are available from the School of Public Health, Fudan University. The data were used under license for this study and are not publicly available. The data are, however, available from the authors upon reasonable request and with permission from the School of Public Health, Fudan University.

Discussion

Principal Findings

The findings of our study suggest that five HBM constructs—in the absence of perceived barriers, a high level of perceived benefits, and self-efficacy—as well as individuals' agreement with recommendations from authorities and friends or family were negatively associated with COVID-19 vaccine hesitancy and positively associated with vaccination behavior. Furthermore, psychological mechanisms were found to mediate the relationship between perceived barriers, perceived benefits, recommendations from authorities and friends or family, and vaccination uptake behavior via vaccine hesitancy. Self-efficacy, perceived barriers, and perceived benefits were correlated with vaccine hesitancy, while perceived barriers and recommendations from friends or family were directly correlated with vaccination.

In this study based in China, the prevalence of vaccine hesitancy was 44.3% (95% CI 42.3%-46.2%), and the vaccination rate was 10.4% (95% CI 9.2%-11.6%), representing high vaccine hesitancy and low vaccination behaviors. Vaccine hesitancy has been universally reported in recent research, with over half of participants (53%) across 19 countries showing vaccine hesitancy, which is similar to our results [59] and in accordance with the decline in vaccine acceptance (from >70% in March 2020 to <50% in October 2020) reported by a recent review [60]. Undoubtedly, eliminating vaccine hesitancy would be beneficial to voluntary vaccination behaviors, as seen in this study, which showed that the vaccination rate was nearly 8 times higher among the participants who were accepting of vaccines compared to those who were vaccine hesitant. In the SEM results, vaccine hesitancy was also strongly negatively associated with vaccination behaviors (Hypothesis 1 confirmed). Therefore, the control of vaccine hesitancy and the promotion of voluntary vaccination still seem to be challenges in the context of the COVID-19 pandemic.

In this study, female participants showed more COVID-19 vaccine hesitancy, which is consistent with previous findings in the literature [61,62]; a possible reason for this finding is that women are more likely to be concerned about side effects [63] and take nonpharmaceutical protective measures (eg, masking and maintaining social distance) [64], while men are more inclined to adopt medical intervention [65]. Medical personnel showed less vaccine hesitancy and a much higher vaccination rate in this study, which may be inconsistent with the general argument that health workers have strong negative attitudes toward vaccines, with strong skepticism about their safety and effectiveness, especially regarding the influenza vaccine [66,67]. Another finding seems unexpected; that is, that the participants with higher monthly salaries were associated with both vaccine hesitancy and a higher vaccination rate; in other words, even though these individuals were vaccine hesitant, they were still vaccinated. Vaccine hesitancy was not only a direct determinant of vaccination but also a perceived barrier. Participants with higher salaries were more likely to have higher socioeconomic status [68], so they could more easily access social resources; that is, they had lower barriers to obtaining vaccines, which could then increase the vaccination rate among this group.

Although some of the HBM factors were not directly associated with the vaccination rate, perceived benefits of vaccination, perceived barriers to vaccination, self-efficacy for vaccination, and recommendations from authorities were correlated with vaccine hesitancy (Hypothesis 2 partially confirmed), which was consistent with previous research among the Malaysian public [36] and the Chinese general population [69]. In all HBM constructs associated with vaccine hesitancy and vaccination, self-efficacy for COVID-19 vaccination was an important predictor of vaccination behaviors, via vaccine hesitancy. This result is similar to the findings of previous studies on influenza vaccination, according to which self-efficacy is a key factor of willingness, which in turn predicts behavior [46,70]. Self-efficacy also plays a mediating role between vaccine hesitancy and other HBM components, including perceived barriers, perceived benefits, and recommendations from authorities and friends or family, and it indirectly influences vaccination uptake. This finding was supported by the HBM hypothesis (Hypothesis 3 partially confirmed) that HBM constructs and cues to action may not share a juxtaposition or parallel relationship, but self-efficacy functioned as a serial mediator [71]. Hilyard et al noted that public self-efficacy for COVID-19 vaccination could be promoted by enhancing the perceived benefits of vaccination, confidence in overcoming possible side effects (ie, perceived barriers), and recommendations from authorities, such as the Obamas' modeling of H1N1 vaccine acceptance for their daughters [52]. In this study, self-efficacy was measured as a specific domain with confidence in the safety of the COVID-19 vaccine, a low prevalence of side effects of the COVID-19 vaccine, and success in dealing with side effects. Vaccine safety or side effects, which are regarded as contributing to the development of disease, are of paramount importance to individual efficacy when deciding whether to vaccinate [72,73] and are even relevant aspects that help explain the antivaccine movement in Europe [74]. A study argues that a perceived risk-benefit balance may influence confidence in vaccine uptake; in other words, a combined decision-making process relies on a trade-off between benefits and risks [66]. In addition to cues to action, this result was consistent with a previous study showing that compliance with recommendations from health workers may also be correlated with confidence in vaccine efficacy [73], because they can share personal knowledge about being immunized and motivate vaccine uptake efficacy [75].

In addition to the direct and mediating effect of self-efficacy, some HBM constructs were directly associated with vaccine hesitancy and vaccination behavior. Perceived barriers were both positively correlated with vaccine hesitancy and detrimental to vaccination, as measured by the safety, side effects, and inaccessibility of the COVID-19 vaccine, in which safety may influence self-efficacy as aforementioned, while inaccessibility would hinder the perceived convenience of COVID-19 vaccination behavior directly. With a more specific formulation, a controlled before-and-after trial study showed that arranging time and transportation were key predictors of both intention and behavior regarding influenza vaccination [76]. A previous survey also found that the side effects and safety of influenza vaccination were the most common reasons for vaccine hesitancy [77]. Perceived benefits were associated with vaccine hesitancy, which was measured by preventing the self and one's family from being infected after COVID-19 vaccination. From an altruistic motivation perspective, people could be vaccinated to protect not only themselves but also their loved ones; in other words, there could be more willingness to receive the vaccine if individuals believe that it helps reduce the transmission of COVID-19 [78]. Recommendations from family were found to be directly associated with vaccination behavior in this study. An online survey in Canada showed that respondents reported that encouragement from both colleagues and employers was beneficial to their vaccination decision-making process [55]. Another finding implied that a recommendation from a spouse or a friend is an important cue to action in determining willingness to accept the Zika virus vaccine [79]. However, perceived susceptibility and severity were not enough to reduce vaccine hesitancy and promote vaccination behavior. A review indicated that perceived barriers were the most powerful single predictor of preventive health behavior across all studies and behaviors, and perceived severity was the least powerful predictor [71].

From the perspective of the HBM on understanding vaccination behavior, it is valuable that self-efficacy is an important and direct predictor of COVID-19 vaccine hesitancy because it can also mediate the influences from cues to action, perceived barriers, and perceived benefits. Furthermore, vaccine hesitancy was strongly correlated with vaccination behavior but was not the only determinant, since perceived barriers and recommendations from friends or family were also associated with vaccination behavior directly and in combination.

In practice, it is valuable for other nations to know what the Chinese vaccine hesitancy and vaccination statuses were at the beginning of the critical period when COVID-19 vaccination became available to the public, free of charge. This finding indicates that health authorities or doctors may be less effective in motivating people to action, while it may be useful to advocate for more volunteers to engage in motivating their

friends or family members. Although the antivaccine movement that occurred in other nations was not popular in mainland China, vaccine hesitancy and refusal were not rare occurrences without mandatory vaccination in this study. Moreover, it is essential to reinforce the publishing of information regarding the safety and validity of COVID-19 vaccines and incentives of vaccination completion, which could then promote public confidence in overcoming vaccination barriers and in achieving benefits after vaccination.

In summary, there was a high prevalence of vaccine hesitancy and low vaccination behavior in China during the first month (January 2021) when vaccinations became available to the vaccine-eligible population. The HBM framework is a useful framework to guide the development of future campaigns to reduce vaccine hesitancy and promote COVID-19 vaccination.

Limitations

There are some potential limitations to this study. First, due to the convenience sampling and snowball recruitment methods that were part of the online survey process, selection bias, such as the participation of fewer respondents with low education attainment and fewer older adults (aged over 50 years), may have affected the generalizability of the results. Second, vaccine hesitancy was measured by a single item derived from a definition from the SAGE working group, which may promote more accurate measurement tools in future research. Furthermore, the vaccination rate in this study may not reflect future trends because only some participants had received the vaccine in a timely manner, vaccinations were available to the public for only 1 month, and there were no incentives except to receive a free vaccination before participating in the study.

Acknowledgments

This work was supported by the National Key R&D Program of China (grants 2018YFC2002000 and 2018YFC2002001) and the National Natural Science Foundation of China (grant 71573048).

Authors' Contributions

JD, JG, HF, PZ, and YJ designed the study and obtained the data. HC and X Li undertook the analysis, supervised by JD, JG, and HF, and wrote the manuscript. X Liu translated the questionnaire. HC, X Li, X Liu, YM, and RW performed the survey. All authors read the final manuscript and agreed with the content.

Conflicts of Interest

None declared.

References

1. COVID-19 Weekly Epidemiological Update. Geneva, Switzerland: World Health Organization; 2021 Jan 12. URL: https://www.who.int/publications/m/item/weekly-epidemiological-update---12-january-2021 [accessed 2021-08-13]
2. WHO Coronavirus (COVID-19) Dashboard. Geneva, Switzerland: World Health Organization; 2021. URL: https://covid19.who.int/ [accessed 2021-08-13]
3. Dubé E, Laberge C, Guay M, Bramadat P, Roy R, Bettinger JA. Vaccine hesitancy: An overview. Hum Vaccin Immunother 2013 Aug;9(8):1763-1773 [FREE Full text] [doi: 10.4161/hv.24657] [Medline: 23584253]
4. Rappuoli R, Mandl CW, Black S, De Gregorio E. Vaccines for the twenty-first century society. Nat Rev Immunol 2011 Nov 04;11(12):865-872 [FREE Full text] [doi: 10.1038/nri3085] [Medline: 22051890]
5. Schaffer DeRoo S, Pudalov NJ, Fu LY. Planning for a COVID-19 vaccination program. JAMA 2020 Jun 23;323(24):2458-2459. [doi: 10.1001/jama.2020.8711] [Medline: 32421155]
6. The new coronavirus vaccine is free for all! Health insurance fund and public finance share the burden. Government of the People's Republic of China. 2021. URL: http://www.gov.cn/fuwu/2021-01/09/content_5578430.htm [accessed 2021-08-13]
7. Experts: China's vaccine adverse reaction monitoring system will continue to track the relevant situation after COVID-19 vaccination. Government of the People's Republic of China. 2021. URL: http://www.gov.cn/xinwen/2021-01/03/content_5576402.htm [accessed 2021-08-13]
8. Who can be vaccinated? Where can I get the vaccine? What to do if there is an adverse reaction? Authorities respond to seven questions about the novel coronavirus vaccine. Government of the People's Republic of China. 2020. URL: http://www.gov.cn/xinwen/2020-12/19/content_5571273.htm [accessed 2021-08-13]
9. National Health Commission. As of January 26, 22.767 million doses of COVID-19 vaccine had been administered. People's Daily Online. 2021. URL: http://health.people.com.cn/n1/2021/0127/c14739-32014069.html [accessed 2021-08-13]
10. The State Information Office of the People's Republic of China held a press conference on the joint prevention and control of novel coronavirus pneumonia. The State Information Office of the People's Republic of China. 2020. URL: http://www.scio.gov.cn/xwfbh/xwbfbh/wqfbh/42311/42478/index.htm [accessed 2021-08-13]
11. Local press release. The State Council Information Office of the People's Republic of China. 2021. URL: http://www.scio.gov.cn/xwFbh/gssxwfbh/index.htm [accessed 2021-08-13]

12. Shandong held a press conference on the epidemic prevention and control work and the next steps. The State Council Information Office of the People's Republic of China. 2020. URL: http://www.scio.gov.cn/xwFbh/gssxwfbh/xwfbh/shandong/Document/1672973/1672973.htm [accessed 2021-08-13]
13. Novel coronavirus (COVID-19): Global vaccination status inquiry. Win.d. URL: http://wx.wind.com.cn/WindSariWeb/sari/messageVaccin.html?lan=en [accessed 2021-08-13]
14. Mo PK, Luo S, Wang S, Zhao J, Zhang G, Li L, et al. Intention to receive the COVID-19 vaccination in China: Application of the diffusion of innovations theory and the moderating role of openness to experience. Vaccines (Basel) 2021 Feb 05;9(2):129 [FREE Full text] [doi: 10.3390/vaccines9020129] [Medline: 33562894]
15. Siddiqui M, Salmon DA, Omer SB. Epidemiology of vaccine hesitancy in the United States. Hum Vaccin Immunother 2013 Dec;9(12):2643-2648 [FREE Full text] [doi: 10.4161/hv.27243] [Medline: 24247148]
16. Immunization, vaccines and biologicals. World Health Organization. URL: https://www.who.int/teams/regulation-prequalification/eul/immunization-vaccines-and-biologicals [accessed 2021-08-13]
17. MacDonald NE, SAGE Working Group on Vaccine Hesitancy. Vaccine hesitancy: Definition, scope and determinants. Vaccine 2015 Aug 14;33(34):4161-4164 [FREE Full text] [doi: 10.1016/j.vaccine.2015.04.036] [Medline: 25896383]
18. Ten threats to global health in 2019. World Health Organization. URL: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 [accessed 2021-08-13]
19. SAGE Working Group on Vaccine Hesitancy. What influences vaccine acceptance: A model of determinants of vaccine hesitancy. World Health Organization. 2013 Mar 18. URL: https://www.who.int/immunization/sage/meetings/2013/april/1_Model_analyze_driversofvaccineConfidence_22_March.pdf [accessed 2021-08-13]
20. Yu WZ, Ji SS, Liu J. Continuous monitoring of parental confidence on vaccination following the Shandong illegal vaccine selling event. China Vaccines Immun 2016 Dec 6;24(02):230-236.
21. Larson HJ, Cooper LZ, Eskola J, Katz SL, Ratzan S. Addressing the vaccine confidence gap. Lancet 2011 Aug;378(9790):526-535. [doi: 10.1016/s0140-6736(11)60678-8]
22. Larson H, Paterson PB, Erondu N. The globalization of risk and risk perception. Drug Saf 2012 Dec 13;35(11):1053-1059. [doi: 10.1007/bf03261991]
23. Barello S, Nania T, Dellafiore F, Graffigna G, Caruso R. 'Vaccine hesitancy' among university students in Italy during the COVID-19 pandemic. Eur J Epidemiol 2020 Aug;35(8):781-783 [FREE Full text] [doi: 10.1007/s10654-020-00670-z] [Medline: 32761440]
24. Dror AA, Eisenbach N, Taiber S, Morozov NG, Mizrachi M, Zigron A, et al. Vaccine hesitancy: The next challenge in the fight against COVID-19. Eur J Epidemiol 2020 Aug;35(8):775-779. [doi: 10.1007/s10654-020-00671-y] [Medline: 32785815]
25. Wang W, Wu Q, Yang J, Dong K, Chen X, Bai X, et al. Global, regional, and national estimates of target population sizes for COVID-19 vaccination: Descriptive study. BMJ 2020 Dec 15;371:m4704 [FREE Full text] [doi: 10.1136/bmj.m4704] [Medline: 33323388]
26. Neumann-Böhme S, Varghese NE, Sabat I, Barros PP, Brouwer W, van Exel J, et al. Once we have it, will we use it? A European survey on willingness to be vaccinated against COVID-19. Eur J Health Econ 2020 Sep;21(7):977-982 [FREE Full text] [doi: 10.1007/s10198-020-01208-6] [Medline: 32591957]
27. Guidry JP, Laestadius LI, Vraga EK, Miller CA, Perrin PB, Burton CW, et al. Willingness to get the COVID-19 vaccine with and without emergency use authorization. Am J Infect Control 2021 Feb;49(2):137-142 [FREE Full text] [doi: 10.1016/j.ajic.2020.11.018] [Medline: 33227323]
28. Ward JK, Peretti-Watel P, Bocquier A, Seror V, Verger P. Vaccine hesitancy and coercion: All eyes on France. Nat Immunol 2019 Oct;20(10):1257-1259. [doi: 10.1038/s41590-019-0488-9] [Medline: 31477920]
29. Sherman SM, Smith LE, Sim J, Amlôt R, Cutts M, Dasch H, et al. COVID-19 vaccination intention in the UK: Results from the COVID-19 vaccination acceptability study (CoVAccS), a nationally representative cross-sectional survey. Hum Vaccin Immunother 2021 Jun 03;17(6):1612-1621 [FREE Full text] [doi: 10.1080/21645515.2020.1846397] [Medline: 33242386]
30. Abraham C, Sheeran P. The health belief model. In: Ayers S, Baum A, McManus C, Newman S, Wallston K, Weinman J, et al, editors. Cambridge Handbook of Psychology, Health and Medicine. 2nd edition. Cambridge, UK: Cambridge University Press; 2007:97-102.
31. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
32. Mckenna SP. Predicting health behaviour: Research and practice with social cognition models. Saf Sci 1996 Dec;24(3):229-230. [doi: 10.1016/s0925-7535(97)81483-x]
33. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. Health Educ Q 1988;15(2):175-183. [doi: 10.1177/109019818801500203] [Medline: 3378902]
34. Juraskova I, O'Brien M, Mullan B, Bari R, Laidsaar-Powell R, McCaffery K. HPV vaccination and the effect of information framing on intentions and behaviour: An application of the theory of planned behaviour and moral norm. Int J Behav Med 2012 Dec;19(4):518-525. [doi: 10.1007/s12529-011-9182-5] [Medline: 21879340]

35. Coe AB, Gatewood SB, Moczygemba LR, Goode J, Beckner JO. The use of the health belief model to assess predictors of intent to receive the novel (2009) H1N1 influenza vaccine. Innov Pharm 2012;3(2):1-11 [FREE Full text] [doi: 10.24926/iip.v3i2.257] [Medline: 22844651]
36. Wong LP, Alias H, Wong P, Lee HY, AbuBakar S. The use of the health belief model to assess predictors of intent to receive the COVID-19 vaccine and willingness to pay. Hum Vaccin Immunother 2020 Sep 01;16(9):2204-2214 [FREE Full text] [doi: 10.1080/21645515.2020.1790279] [Medline: 32730103]
37. Mercadante AR, Law AV. Will they, or won't they? Examining patients' vaccine intention for flu and COVID-19 using the Health Belief Model. Res Social Adm Pharm 2021 Sep;17(9):1596-1605 [FREE Full text] [doi: 10.1016/j.sapharm.2020.12.012] [Medline: 33431259]
38. Quah SR. The health belief model and preventive health behaviour in Singapore. Soc Sci Med 1985 Jan;21(3):351-363. [doi: 10.1016/0277-9536(85)90112-1]
39. Orji R, Vassileva J, Mandryk R. Towards an effective health interventions design: An extension of the health belief model. Online J Public Health Inform 2012;4(3):1-31 [FREE Full text] [doi: 10.5210/ojphi.v4i3.4321] [Medline: 23569653]
40. Rahman M, Berenson AB, Herrera SR. Perceived susceptibility to pregnancy and its association with safer sex, contraceptive adherence and subsequent pregnancy among adolescent and young adult women. Contraception 2013 Apr;87(4):437-442 [FREE Full text] [doi: 10.1016/j.contraception.2012.09.009] [Medline: 23083528]
41. Myers LB, Goodwin R. Determinants of adults' intention to vaccinate against pandemic swine flu. BMC Public Health 2011 Jan 06;11(1):15 [FREE Full text] [doi: 10.1186/1471-2458-11-15] [Medline: 21211000]
42. Khodaveisi M, Salehi Khah M, Bashirian S, Karami M, Khodaveisi M. The effect of health belief model-based training on preventive behaviors of hepatitis B in addicts. Int J High Risk Behav Addict 2018 Jan 09;7(2):e58579 [FREE Full text] [doi: 10.5812/ijhrba.58579]
43. Donadiki E, Jiménez-García R, Hernández-Barrera V, Sourtzi P, Carrasco-Garrido P, López de Andrés A, et al. Health Belief Model applied to non-compliance with HPV vaccine among female university students. Public Health 2014 Mar;128(3):268-273. [doi: 10.1016/j.puhe.2013.12.004] [Medline: 24529635]
44. Wagner AL, Boulton ML, Sun X, Mukherjee B, Huang Z, Harmsen IA, et al. Perceptions of measles, pneumonia, and meningitis vaccines among caregivers in Shanghai, China, and the health belief model: A cross-sectional study. BMC Pediatr 2017 Jun 12;17(1):143 [FREE Full text] [doi: 10.1186/s12887-017-0900-2] [Medline: 28606106]
45. Rabin C, Dutra S. Predicting engagement in behaviors to reduce the spread of COVID-19: The roles of the health belief model and political party affiliation. Psychol Health Med 2021 Apr 27:1-10. [doi: 10.1080/13548506.2021.1921229] [Medline: 33906540]
46. Fall E, Izaute M, Chakroun-Baggioni N. How can the health belief model and self-determination theory predict both influenza vaccination and vaccination intention? A longitudinal study among university students. Psychol Health 2018 Jun;33(6):746-764. [doi: 10.1080/08870446.2017.1401623] [Medline: 29132225]
47. Webb TL, Sheeran P. How do implementation intentions promote goal attainment? A test of component processes. J Exp Soc Psychol 2007 Mar;43(2):295-302. [doi: 10.1016/j.jesp.2006.02.001]
48. Brewer NT, Rimer BK. Perspectives on health behavior theories that focus on individuals. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior and Health Education: Theory, Research, and Practice. 4th edition. San Francisco, CA: Jossey-Bass; 2008:149-165.
49. Christy SM, Winger JG, Mosher CE. Does self-efficacy mediate the relationships between social-cognitive factors and intentions to receive HPV vaccination among young women? Clin Nurs Res 2019 Jul;28(6):708-725 [FREE Full text] [doi: 10.1177/1054773817741590] [Medline: 29134823]
50. Guidry JP, Carlyle KE, Perrin PB, LaRose JG, Ryan M, Messner M. A path model of psychosocial constructs predicting future Zika vaccine uptake intent. Vaccine 2019 Aug 23;37(36):5233-5241. [doi: 10.1016/j.vaccine.2019.07.064] [Medline: 31375439]
51. Gerend MA, Shepherd JE. Predicting human papillomavirus vaccine uptake in young adult women: Comparing the health belief model and theory of planned behavior. Ann Behav Med 2012 Oct;44(2):171-180 [FREE Full text] [doi: 10.1007/s12160-012-9366-5] [Medline: 22547155]
52. Hilyard KM, Quinn SC, Kim KH, Musa D, Freimuth VS. Determinants of parental acceptance of the H1N1 vaccine. Health Educ Behav 2014 Jun;41(3):307-314 [FREE Full text] [doi: 10.1177/1090198113515244] [Medline: 24369176]
53. Wang C, Han B, Zhao T, Liu H, Liu B, Chen L, et al. Vaccination willingness, vaccine hesitancy, and estimated coverage at the first round of COVID-19 vaccination in China: A national cross-sectional study. Vaccine 2021 May 18;39(21):2833-2842 [FREE Full text] [doi: 10.1016/j.vaccine.2021.04.020] [Medline: 33896661]
54. Wang Q, Xiu S, Zhao S, Wang J, Han Y, Dong S, et al. Vaccine hesitancy: COVID-19 and influenza vaccine willingness among parents in Wuxi, China-A cross-sectional study. Vaccines (Basel) 2021 Apr 01;9(4):342 [FREE Full text] [doi: 10.3390/vaccines9040342] [Medline: 33916277]
55. Nowrouzi-Kia B, McGeer A. External cues to action and influenza vaccination among post-graduate trainee physicians in Toronto, Canada. Vaccine 2014 Jun 24;32(30):3830-3834. [doi: 10.1016/j.vaccine.2014.04.067] [Medline: 24837775]

56. Kondo N, van Dam RM, Sembajwe G, Subramanian SV, Kawachi I, Yamagata Z. Income inequality and health: The role of population size, inequality threshold, period effects and lag effects. J Epidemiol Community Health 2012 Jun;66(6):e11. [doi: 10.1136/jech-2011-200321] [Medline: 22012964]
57. Muthén LK, Muthén BO. Mplus User's Guide. 8th edition. Los Angeles, CA: Muthén & Muthén; 2017 Apr. URL: https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf [accessed 2021-08-17]
58. Byrne BM, Crombie G. Modeling and testing change: An introduction to the latent growth curve model. Underst Stat 2003 Aug;2(3):177-203. [doi: 10.1207/s15328031us0203_02]
59. Lazarus JV, Ratzan SC, Palayew A, Gostin LO, Larson HJ, Rabin K, et al. A global survey of potential acceptance of a COVID-19 vaccine. Nat Med 2021 Feb;27(2):225-228 [FREE Full text] [doi: 10.1038/s41591-020-1124-9] [Medline: 33082575]
60. Lin C, Tu P, Beitsch LM. Confidence and receptivity for COVID-19 vaccines: A rapid systematic review. Vaccines (Basel) 2020 Dec 30;9(1):16 [FREE Full text] [doi: 10.3390/vaccines9010016] [Medline: 33396832]
61. Bish A, Yardley L, Nicoll A, Michie S. Factors associated with uptake of vaccination against pandemic influenza: A systematic review. Vaccine 2011 Sep 02;29(38):6472-6484. [doi: 10.1016/j.vaccine.2011.06.107] [Medline: 21756960]
62. Baumgaertner B, Ridenhour BJ, Justwan F, Carlisle JE, Miller CR. Risk of disease and willingness to vaccinate in the United States: A population-based survey. PLoS Med 2020 Oct;17(10):e1003354 [FREE Full text] [doi: 10.1371/journal.pmed.1003354] [Medline: 33057373]
63. Freed GL, Clark SJ, Butchart AT, Singer DC, Davis MM. Parental vaccine safety concerns in 2009. Pediatrics 2010 Apr;125(4):654-659. [doi: 10.1542/peds.2009-1962] [Medline: 20194286]
64. Bish A, Michie S. Demographic and attitudinal determinants of protective behaviours during a pandemic: A review. Br J Health Psychol 2010 Nov;15(Pt 4):797-824 [FREE Full text] [doi: 10.1348/135910710X485826] [Medline: 20109274]
65. Moran KR, Del Valle SY. A meta-analysis of the association between gender and protective behaviors in response to respiratory epidemics and pandemics. PLoS One 2016;11(10):e0164541 [FREE Full text] [doi: 10.1371/journal.pone.0164541] [Medline: 27768704]
66. Rey D, Fressard L, Cortaredona S, Bocquier A, Gautier A, Peretti-Watel P. Vaccine hesitancy in the French population in 2016, and its association with vaccine uptake and perceived vaccine risk-benefit balance. Euro Surveill 2018 Apr 23;23(17):1-10 [FREE Full text] [doi: 10.2807/1560-7917.es.2018.23.17.17-00816]
67. Raude J, Fressard L, Gautier A, Pulcini C, Peretti-Watel P, Verger P. Opening the 'vaccine hesitancy' black box: How trust in institutions affects French GPs' vaccination practices. Expert Rev Vaccines 2016 Jul;15(7):937-948. [doi: 10.1080/14760584.2016.1184092] [Medline: 27140417]
68. Wang K, Wong EL, Ho K, Cheung AW, Yau PS, Dong D, et al. Change of willingness to accept COVID-19 vaccine and reasons of vaccine hesitancy of working people at different waves of local epidemic in Hong Kong, China: Repeated cross-sectional surveys. Vaccines (Basel) 2021 Jan 18;9(1):62 [FREE Full text] [doi: 10.3390/vaccines9010062] [Medline: 33477725]
69. Lin Y, Hu Z, Zhao Q, Alias H, Danaee M, Wong LP. Understanding COVID-19 vaccine demand and hesitancy: A nationwide online survey in China. PLoS Negl Trop Dis 2020 Dec;14(12):e0008961 [FREE Full text] [doi: 10.1371/journal.pntd.0008961] [Medline: 33332359]
70. Ernsting A, Gellert P, Schneider M, Lippke S. A mediator model to predict workplace influenza vaccination behaviour--An application of the health action process approach. Psychol Health 2013;28(5):579-592. [doi: 10.1080/08870446.2012.753072] [Medline: 23259583]
71. Jones CL, Jensen JD, Scherr CL, Brown NR, Christy K, Weaver J. The Health Belief Model as an explanatory framework in communication research: Exploring parallel, serial, and moderated mediation. Health Commun 2015;30(6):566-576 [FREE Full text] [doi: 10.1080/10410236.2013.873363] [Medline: 25010519]
72. Chantler TEA, Lees A, Moxon ER, Mant D, Pollard AJ, Fiztpatrick R. The role familiarity with science and medicine plays in parents' decision making about enrolling a child in vaccine research. Qual Health Res 2007 Mar;17(3):311-322. [doi: 10.1177/1049732306298561] [Medline: 17301340]
73. Reynolds D, O'Connell KA. Testing a model for parental acceptance of human papillomavirus vaccine in 9- to 18-year-old girls: A theory-guided study. J Pediatr Nurs 2012 Dec;27(6):614-625. [doi: 10.1016/j.pedn.2011.09.005] [Medline: 22020360]
74. Fournet N, Mollema L, Ruijs WL, Harmsen IA, Keck F, Durand JY, et al. Under-vaccinated groups in Europe and their beliefs, attitudes and reasons for non-vaccination: Two systematic reviews. BMC Public Health 2018 Jan 30;18(1):196 [FREE Full text] [doi: 10.1186/s12889-018-5103-8] [Medline: 29378545]
75. Kempe A, O'Leary ST, Kennedy A, Crane LA, Allison MA, Beaty BL, et al. Physician response to parental requests to spread out the recommended vaccine schedule. Pediatrics 2015 Apr;135(4):666-677 [FREE Full text] [doi: 10.1542/peds.2014-3474] [Medline: 25733753]
76. Payaprom Y, Bennett P, Alabaster E, Tantipong H. Using the Health Action Process Approach and implementation intentions to increase flu vaccine uptake in high risk Thai individuals: A controlled before-after trial. Health Psychol 2011 Jul;30(4):492-500. [doi: 10.1037/a0023580] [Medline: 21534678]

77. Lau AYS, Sintchenko V, Crimmins J, Magrabi F, Gallego B, Coiera E. Impact of a web-based personally controlled health management system on influenza vaccination and health services utilization rates: A randomized controlled trial. J Am Med Inform Assoc 2012;19(5):719-727 [FREE Full text] [doi: 10.1136/amiajnl-2011-000433] [Medline: 22582203]
78. Cerda AA, García LY. Hesitation and refusal factors in individuals' decision-making processes regarding a coronavirus disease 2019 vaccination. Front Public Health 2021;9:626852 [FREE Full text] [doi: 10.3389/fpubh.2021.626852] [Medline: 33968880]
79. Wong LP, Alias H, Hassan J, AbuBakar S. Attitudes towards Zika screening and vaccination acceptability among pregnant women in Malaysia. Vaccine 2017 Oct 13;35(43):5912-5917. [doi: 10.1016/j.vaccine.2017.08.074] [Medline: 28886944]

Abbreviations

CFI: comparative fit index
d: precision
deff: design effect
DIT: diffusion of innovation theory
HBM: health belief model
HPV: human papillomavirus
OR: odds ratio
p: prevalence of COVID-19 vaccination hesitancy
PCA: principal component analysis
RMSEA: root mean square error of approximation
SAGE: Strategic Advisory Group of Experts
SEM: structural equation modeling
SSC: standardized structure coefficient
TLI: Tucker-Lewis index
TPB: theory of planned behavior
WHO: World Health Organization

Edited by C Basch; submitted 02.04.21; peer-reviewed by Q Yang, PKH Mo; comments to author 06.05.21; revised version received 26.06.21; accepted 12.07.21; published 06.09.21.

Please cite as:
Chen H, Li X, Gao J, Liu X, Mao Y, Wang R, Zheng P, Xiao Q, Jia Y, Fu H, Dai J
Health Belief Model Perspective on the Control of COVID-19 Vaccine Hesitancy and the Promotion of Vaccination in China: Web-Based Cross-sectional Study
J Med Internet Res 2021;23(9):e29329
URL: https://www.jmir.org/2021/9/e29329
doi:10.2196/29329
PMID:34280115

Original Paper

The Pandemic Year 2020: World Map of Coronavirus Research

Doris Klingelhöfer[1], MSc; Markus Braun[1]; Dörthe Brüggmann[1], MD, PhD; David A Groneberg[1], MD, PhD

Institute of Occupational, Social and Environmental Medicine, Goethe University, Frankfurt, Germany

Corresponding Author:
Doris Klingelhöfer, MSc
Institute of Occupational, Social and Environmental Medicine
Goethe University
Theodor-Stern-Kai 7
Frankfurt, 60590
Germany
Phone: 49 69 63016650
Email: klingelhoefer@med.uni-frankfurt.de

Abstract

Background: SARS-CoV-2 is one of the most threatening pandemics in human history. As of the date of this analysis, it had claimed about 2 million lives worldwide, and the number is rising sharply. Governments, societies, and scientists are equally challenged under this burden.

Objective: This study aimed to map global coronavirus research in 2020 according to various influencing factors to highlight incentives or necessities for further research.

Methods: The application of established and advanced bibliometric methods combined with the visualization technique of density-equalizing mapping provided a global picture of incentives and efforts on coronavirus research in 2020. Countries' funding patterns and their epidemiological and socioeconomic characteristics as well as their publication performance data were included.

Results: Research output exploded in 2020 with momentum, including citation and networking parameters. China and the United States were the countries with the highest publication performance. Globally, however, publication output correlated significantly with COVID-19 cases. Research funding has also increased immensely.

Conclusions: Nonetheless, the abrupt decline in publication efforts following previous coronavirus epidemics should demonstrate to global researchers that they should not lose interest even after containment, as the next epidemiological challenge is certain to come. Validated reporting worldwide and the inclusion of low-income countries are additionally important for a successful future research strategy.

(J Med Internet Res 2021;23(9):e30692) doi:10.2196/30692

KEYWORDS

COVID-19; SARS-CoV-2; incidence, research funding; socioeconomic factors; bibliometrics; bibliometric analysis; global health; public health; health database; online research; research database

Introduction

In December 2019, a new coronavirus (CoV) variant infected some patients in China. It was transmitted at a seafood and wet animal wholesale market in Wuhan city, Hubei Province. This novel zoonotic coronavirus was named SARS-CoV-2 because it also causes severe acute respiratory syndrome (SARS) [1]. Previously detected coronaviruses such as SARS-CoV and Middle East Respiratory Syndrome (MERS)-CoV led to temporary pandemics and consequently to serious public health challenges, which, however, came to an end with the containment of the diseases and their spread. The first cases of the infectious disease caused by SARS-CoV-2, designated COVID-19, indicated the beginning of an outbreak that would become a still ongoing global pandemic on a scale not seen since the Spanish flu in 1918, which killed up to 50 million people [2].

As of this analysis (January 12, 2021), more than 93 million COVID-19 cases have been confirmed, with the number continuing to rise rapidly. Over 2 million people have died in association with SARS-CoV-2 infection as of that date [3]. The enormous impact is catastrophic and affects all areas of public, political, economic, and private life. For sure, it will for a long time. The associated demands and restrictions on citizens and the social systems of all nations are undeniable.

Indeed, it was expected that the number of publications on CoV would increase sharply in 2020, but to what extent and with what participation were not clear [4]. There already have been some studies on the general output of COVID-19–related publications [5-7]. In addition, however, it is necessary to identify and evaluate the general and national research efforts according to additional influences such as epidemiological and funding characteristics to enable successful and determined planning, funding, and implementation of science-based research in the future that reaches all necessary areas through balanced multidisciplinary research.

To achieve this objective, this study mapped the world according to various influencing factors, leading to an advanced and meaningful assessment of the research of the first COVID-19 pandemic year 2020, which will certainly not be the last.

Methods

Methodological Platform and Data Source

This study followed the methodological approach of the bibliometric platform New Quality and Quantity in Science (NewQIS) [8]. For the first time, NewQIS combined bibliometric analyses with density-equalizing mapping (DEMPs) [9] to depict the global publication landscape on scientific topics. DEMPs enable the rapid acquisition of large-scale data. In this process, countries are distorted according to the density-equalizing principle applied by an algorithm developed by Gastner and Newman [9]. The result is a distorted world map according to the respective evaluation parameter, with countries with high values enlarged and countries with low values reduced.

The aim was to provide solid information on research patterns in terms of trends, incentives, challenges, obstacles, and necessities for all parties involved. Socioeconomic parameters and research-specific conditions at the country level were included in the analyses to assess regional performance according to the need for valuable and balanced research that is accessible and appropriate for all parts of the world.

Search Strategy

To capture all CoV-related articles published in 2020, the following elaborated search term was applied in the Web of Science Core Collection (WoS) search field: Title: “corona virus*” OR “coronavirus*” OR “SARS” OR “MERS” OR “covid-19” OR “severe acute respiratory syndrome” OR “middle east respiratory syndrome” AND Topic: virus* OR epidem* OR CoV OR Co-V OR patient* OR outbreak* OR “corona virus” OR “coronavirus” OR “covid-19” OR “severe acute respiratory syndrome” OR “middle east respiratory syndrome”. This string ensured the representativeness of the database generated. Then, the entries were filtered by original articles to base the evaluation on actual research on coronavirus. No language filter was applied. The year 2020 was chosen as the time frame. The date of data collection was January 12, 2021.

The metadata of the datasets collected in this way were stored and sorted according to the individual evaluation parameters and linked by assigning identification numbers to each entry. Some parameters had to be additionally corrected manually, such as matching institution names and funding sources.

Utilized Parameters and Analyses

In addition to established bibliometric parameters such as publication performance, citation parameters, and networking, CoV-specific parameters were also analyzed in this study. These relate to epidemiological characteristics (numbers of cases associated with COVID-19 [3]), socioeconomic characteristics (gross domestic product [GDP], population size), and funding characteristics of the publishing countries. The socioeconomic and epidemiological parameters were used as absolute numbers to allow comparison with absolute publication numbers. The use of per capita values would also require the use of publication figures per capita, which would be redundant when calculating the ratio. International collaborations were defined by the participation of at least two countries of origin, as indicated in the author's affiliations. China and Taiwan were analyzed separately.

In addition, an analysis of the development of publication and citation numbers in 10 time intervals, into which the year 2020 was divided, was carried out. Furthermore, an analysis of research areas was performed employing clustering with the application VOSviewer developed by van Eck and Waltman [10]. Author keywords were clustered and displayed by nodes and connecting lines that represent the different research areas.

Methodological Limitations

Although the methodology used provides a valid source of data, some limitations must be considered when evaluating the results. First, all analyses can only be as good as the database used. For all NewQIS studies, WoS serves as the standard data source. Despite the often-documented English bias and the limitation of a somewhat restricted dataset, according to the high indexing requirements [11], the database provides representative and qualitative results for the further analyses of this study [12]. Some entries had to be manually unified, such as funding sources, which is a nonrepeatable, standardized procedure. In addition, citation-based analyses are prone to error and cannot be considered a proxy for research quality. Nevertheless, the combination of applied ones provides a deep insight into publication performance.

Results

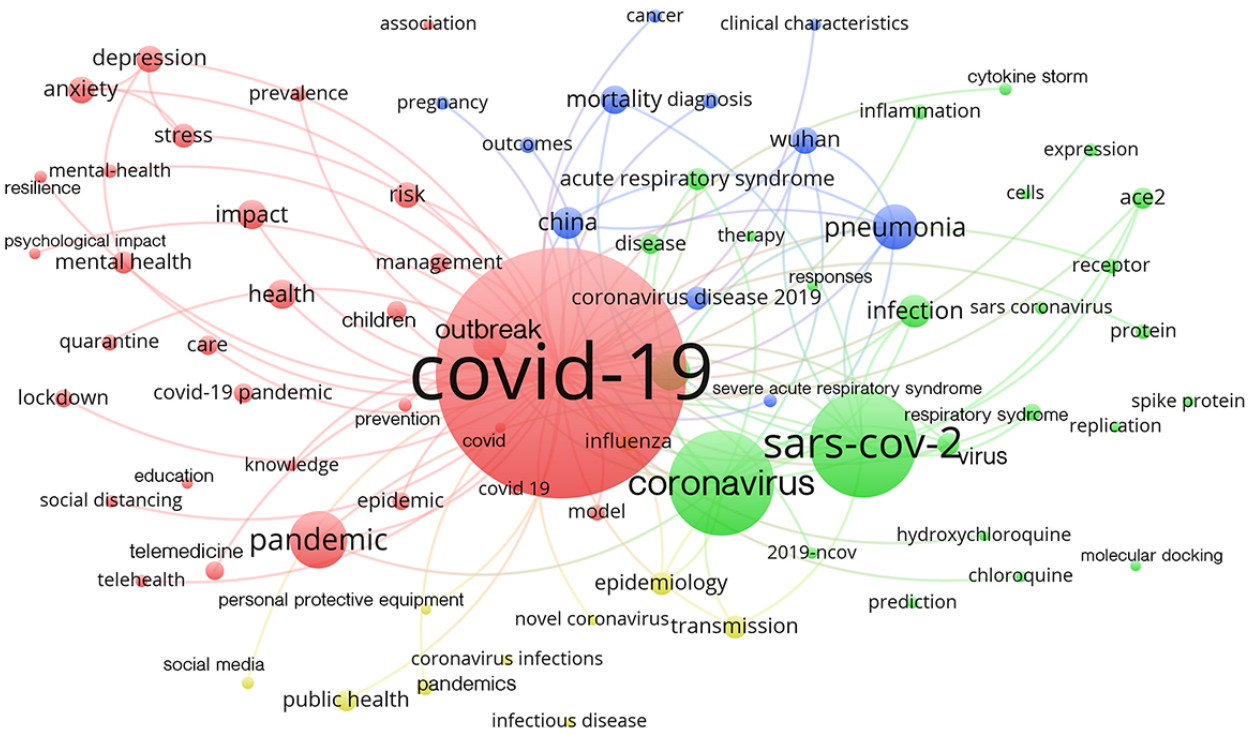
Main Research Foci of CoV Research in 2020

COVID-19 as the most frequently occurring keyword indicates that almost all coronavirus research in 2020 related to SARS-CoV-2 infection, as confirmed by a manually performed review of the articles included in the database.

In addition, analysis of the most frequently used keywords (threshold: 200 occurrences) revealed 4 thematic clusters dealing with the psychological and physical impacts of COVID-19, immunological and biochemical issues, and epidemiological and public health issues (Figure 1). The articles could mainly be assigned to the WoS categories “General and Internal Medicine” (n=8488), “Public, Environmental & Occupational Medicine” (n=2845), “Infectious Diseases” (n=1830), “Science

& Technology – Other Topics" (n=1671), "Environmental Science & Ecology" (n=1397), and "Psychiatry" (n=1116).

Figure 1. Clusters of keywords (threshold: 200 occurrences). Red: psychological health effects; blue: physical health effects and mortality; green: biochemistry and cell biology; yellow: epidemiology and public health.



Evolution of CoV Research in 2020

Applying search terms yielded 67,437 publications on CoV in 2020, of which just under half were original articles (n=32,402) that comprised the analysis database. Of the other document types, 12,435 were letters, 12,245 were editorial material, 6956 were classified as early access, and 6556 were reviews, to name the most common publication types. Looking at the overall evolution of CoV research, the year 2020 was exorbitantly outstanding (Figure 2). The epidemics of SARS in 2003 and MERS in 2012 caused the number of articles to increase to 679 in 2004 and to 340 in 2016, which decreased as soon as the epidemics were contained [4]. These figures show that the number of articles on CoV in 2020 was almost 50 times higher than in the previous peak year of 2004.

Although the 2020 articles are still very young to generate a significant number of citations and thus recognition in the scientific community, which underlies a citation half-life of 7-8 years for biomedical articles, some articles already received an exceptionally high number of citations. The 10 most frequently cited articles are summarized in Table 1. They were published in high-impact journals and are predominantly from China, where SARS-CoV-2 first appeared. Two international collaborations are also represented in the ranking: 1 is an international collaboration of German, Austrian, and Russian scientists, and the other is a Chinese-Australian partnership. All received financial support through government programs from China or Germany.

Figure 2. Development of the numbers of articles about coronavirus research. The comparative figures from 1970-2019 are taken from a previous study by Klingelhöfer et al [4] to show the immense increase in publication numbers in 2020. MERS: Middle East Respiratory Syndrome; SARS: Severe Acute Respiratory Syndrome.

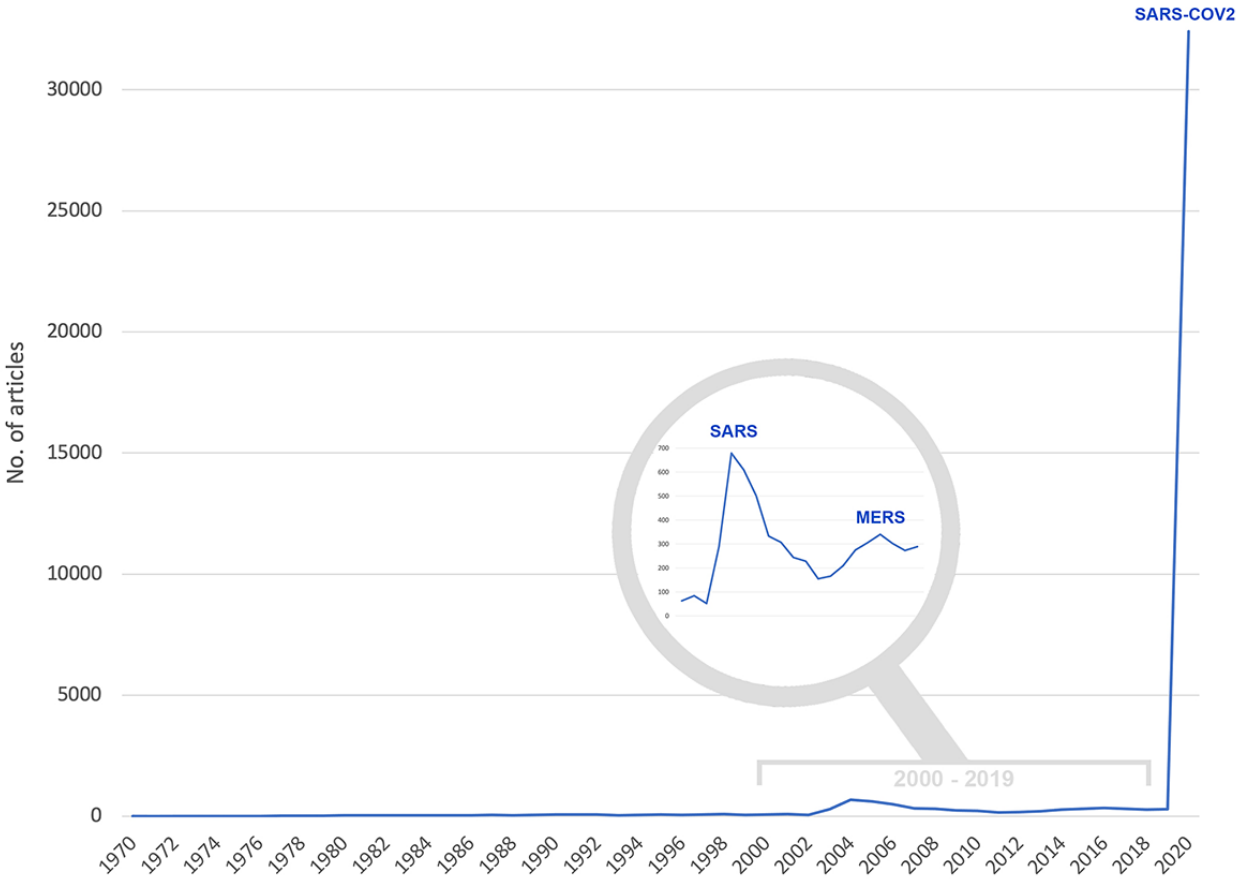


Table 1. Most frequently cited articles on coronavirus (CoV) in 2020 (as of January 12, 2021).

Authors (location)	Number of citations	Title	Source
Huang et al [13] (China)	8978	Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China	The Lancet
Guan et al [14] (China)	5824	Clinical characteristics of coronavirus disease 2019 in China	NEJM[a]
Wang et al [15] (China)	5353	Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China	JAMA[b]
Zhou et al [16] (China)	4987	Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study	The Lancet
Zhu et al [17] (China)	4865	A novel coronavirus from patients with pneumonia in China, 2019	NEJM
Chen et al [18] (China)	4525	Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study	The Lancet
Zhou et al [19] (China)	3762	A pneumonia outbreak associated with a new coronavirus of probable bat origin	Nature
Li et al [20] (China)	3101	Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia	NEJM
Hoffmann et al [21] (Germany, Austria, Russia)	2717	SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor	Cell
Lu et al [22] (China, Australia)	2439	Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding	The Lancet

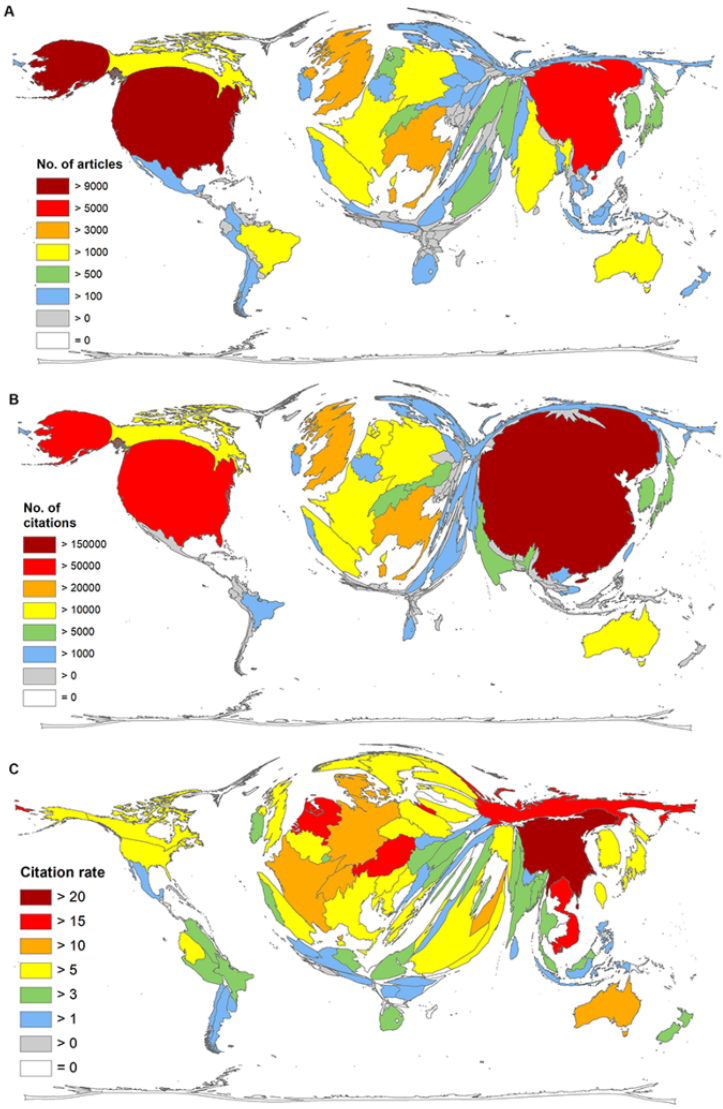
[a]NEJM: New England Journal of Medicine.

[b]JAMA: Journal of the American Medical Association.

Highest Publishing Countries on CoV in 2020

In total, 170 countries or autonomous regions participated in research on CoV in 2020 that was indexed in WoS. By far, the United States had the most publications (n=9018), followed by China (n=5053), Italy (n=3195), the United Kingdom (n=3135), and India (n=1847; Figure 3A).

Figure 3. Density-equalizing maps showing the (A) number of articles, (B) number of citations, and (C) citation rate (citations per article), threshold >30 articles.

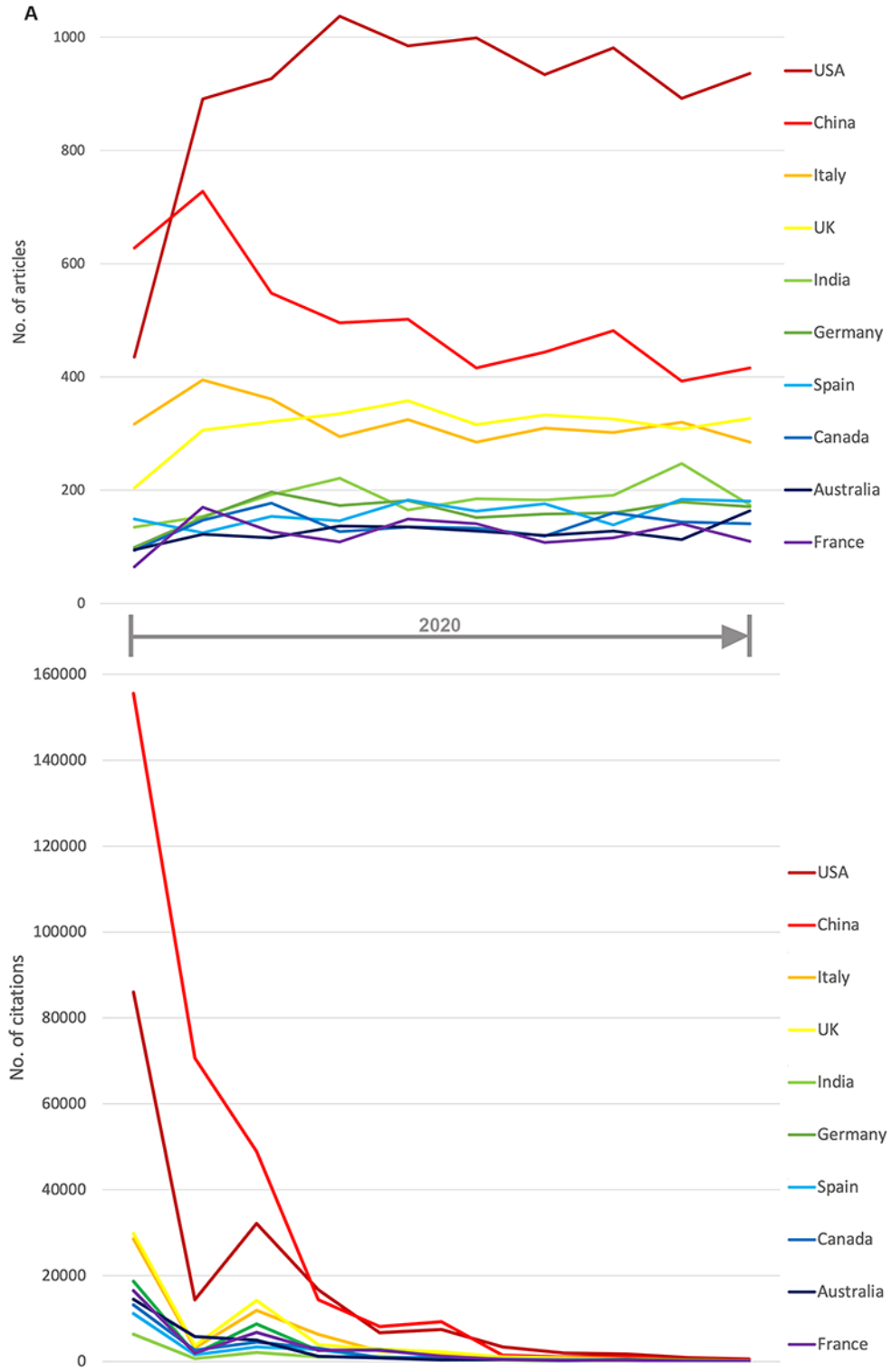


In contrast, China received the most citations (n=155,522). The United States got considerably fewer citations on their CoV-related articles (n=86,003). It is followed by the United Kingdom (n=29,840), Italy (n=28,530), and Germany (n=18,695). India ranked only 12th with 6351 citations (Figure 3B).

Resulting from these numbers, China also led the ranking when looking at the citation rates of countries with at least 30 articles on CoV in 2020 (threshold) with a citation rate of 30.78. It was followed by the Netherlands (citation rate=19.56), Russia (citation rate=17.04), Austria (citation rate=16.18), and Vietnam (citation rate=16.07). In this term, the United States and the United Kingdom ranked nearly similarly, at only 22nd (citation rate=9.54) and 23rd (citation rate=9.52), and India was even only ranked at 79th (citation rate=3.44; Figure 3C).

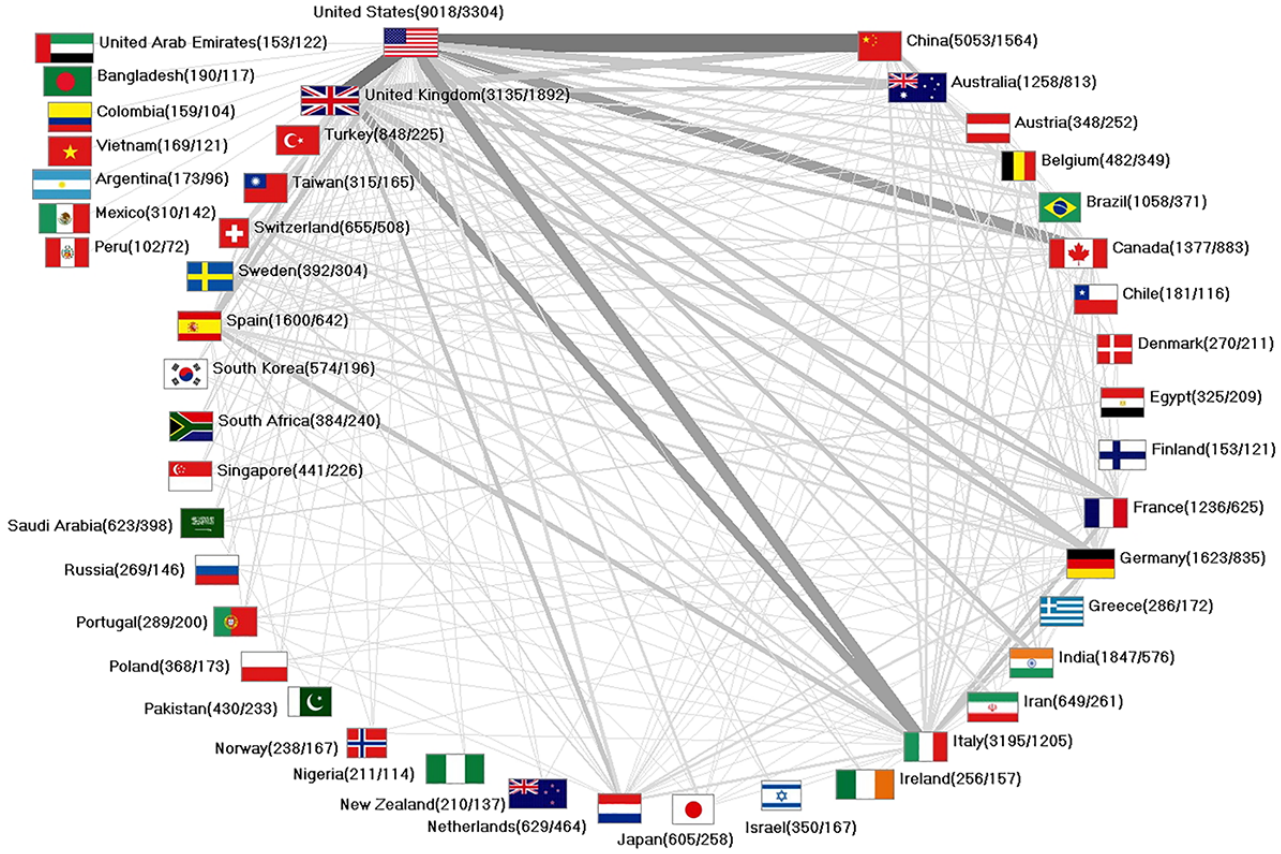
The dominance of US-American and Chinese researchers in terms of CoV research over the entire year 2020 could be shown by the numbers of articles and citations. Nevertheless, their share varied over the year. While Chinese articles dominated in the early phase, US-American articles gained momentum as the year progressed. The United Kingdom's share also increased during the year, eventually overtaking the share of Italy (Figure 4A). The same holds true for the trend in the number of citations, although this depends on the minimum time the articles had to generate citations by other publications (Figure 4B).

Figure 4. Development in 2020 of the 10 countries with the highest number of articles, including the (A) number of articles on the topic of coronavirus and (B) number of citations of the articles on the topic of coronavirus.



A broad international network on CoV research has developed, with the United States and China being the main cooperation countries with 800 collaborations (Figure 5). In general, the United States acts as the main core country when considering international partnerships. In addition to the Chinese collaborations, there were 640 collaborations with the United Kingdom, 496 collaborations with Italy, and 496 collaborations with Canada, to name the most collaborating nations. There were 412 collaborations between the United Kingdom and Italy and 279 collaborations between Italy and Spain.

Figure 5. International network on coronavirus (threshold 35 collaborations).



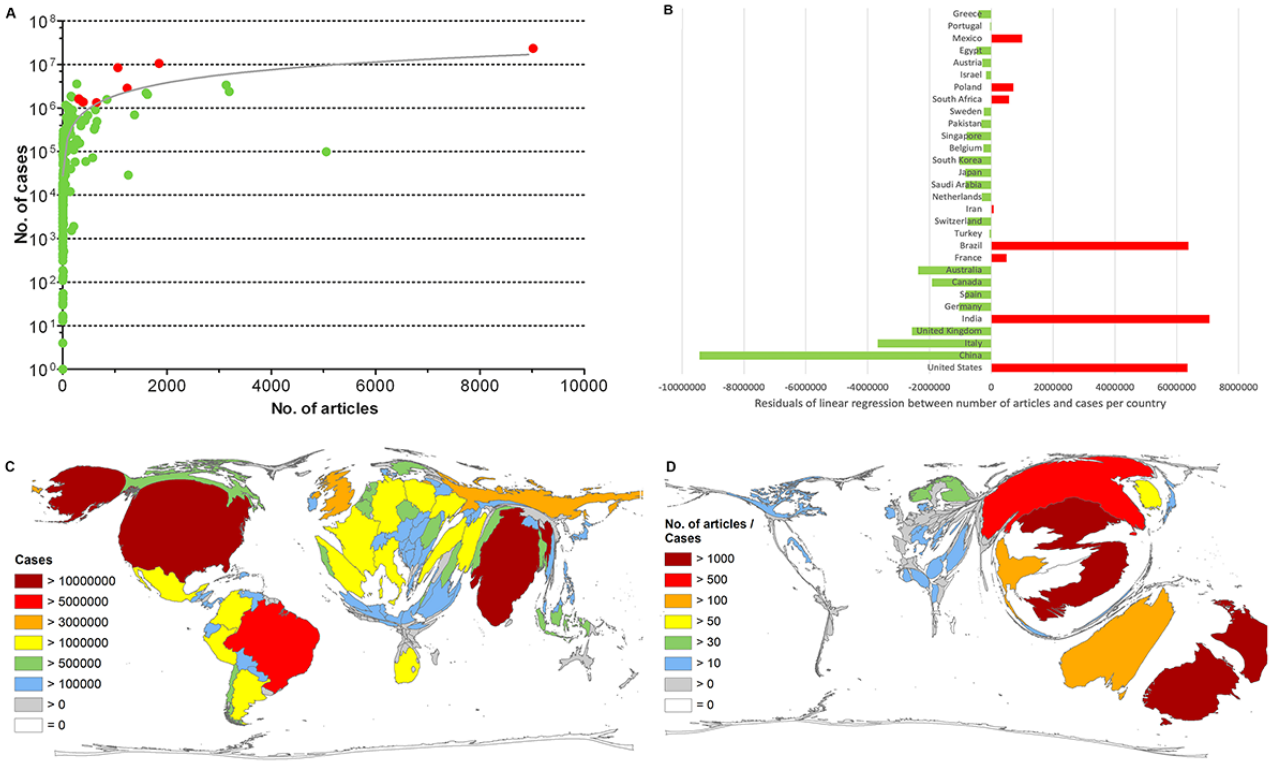
Factors Influencing CoV Research in 2020

Epidemiological Factors

To consider the need for research on CoV according to the national burden of disease, the relationship of the number of articles on CoV to nationally registered cases of COVID-19 (as of January 18, 2021) was analyzed. The coefficient of determination was r^2=0.64 (Figure 6A) showing a significant correlation (Spearman; *P*<.001). China showed the highest distance from the regression line (residual) toward a favorable publication performance (negative values), while India, the United States, and Brazil showed the highest deviations toward a more negative relationship (positive values). The most publishing European countries were also in the positive range of deviation from the regression line but on a lower level than China (Figure 6B).

To provide a picture of the occurrence of COVID-19 cases, a DEMP was generated showing the corresponding distortions on the world map (Figure 6C), with the highest numbers occurring in the United States, India, Brazil, Russia, the United Kingdom, France, Italy, Spain, and Germany, to indicate the countries with more than 2 million cases by January 18, 2021 (date of data collection) [3]. Relating these numbers to COVID-19 cases by calculating the ratio of countries (R_{CASES}) with at least 30 articles on CoV (threshold), southeastern countries were ahead. New Zealand, with only 1900 registered cases, could be ranked first (R_{CASES}=1105.26), followed by Vietnam (R_{CASES}=1099.54), China (R_{CASES}=512.34), Australia (R_{CASES}=438.50), and Thailand (R_{CASES}=119.46). On the other hand, among the most publishing countries besides China, Italy ranked next at 21st, followed by the United Kingdom (26th), Germany (30th), the United States (51st), and India (69th) due to their enormous incidence rates (Figure 6D).

Figure 6. Relationship between the number of articles and COVID-19 cases per country, based on (A) linear regression (logarithmic display of y-axis), with red indicating the countries with an unfavorable ratio in terms of the number of articles among the top 30 countries; (B) residuals of the linear regression of the 30 most publishing countries, with red indicating the countries with an unfavorable ratio in terms of the number of articles; (C) density-equalizing map projection of the number of COVID-19 cases as of January 18, 2021; (D) density-equalizing map projection of the ratio between the number of articles and the number of COVID-19 cases (threshold >30 articles).



Socioeconomic Factors

The inclusion of socioeconomic features of the publishing countries revealed a different ranking. The first ratio related the publication output per country with its economic power (ranked GDP [R_{GDP}]). When looking at the countries with more than 100 articles on CoV in 2020, Jordan, as an upper-middle-income economy, led the ranking (R_{GDP}=1543.10). It was followed only by high-income economies until position 21 (upper-middle-income economy=South Africa). Rank 2 could be attributed to Italy (R_{GDP}=1438.54), followed by Switzerland (R_{GDP}=1325.11), New Zealand (R_{GDP}=1201.37), and Israel (R_{GDP}=1178.45). In addition, the number of articles was significantly correlated with gross expenditures on research and development in purchasing power parity in international $ (ppp$; Spearman, P<.001).

In terms of the population size (R_{POP}) of the same countries, only high-income countries were leading. Switzerland ranked first (R_{POP}=80.08), followed by Norway (R_{POP}=45.20), Israel (R_{POP}=42.82), Belgium (R_{POP}=42.25), and Austria (R_{POP}=39.95; Table 2).

Table 2. Socioeconomic parameters, by country, in descending order by R_{GDP} [23].

Country	n	GDP[a] in US $1000 billion	R_{GDP}[b]	Rank R_{GDP}	Population in millions	R_{POP}[c]	Rank R_{POP}
Jordan	133	0.09	1543.10	UMI[d] 1	8.19	16.25	UMI 1
Italy	3195	2.22	1438.54	HI[e] 1	62.01	51.53	HI 5
Switzerland	655	0.49	1325.11	HI 2	8.18	80.08	HI 1
New Zealand	210	0.17	1201.37	HI 3	4.47	46.93	HI 8
Israel	350	0.30	1178.45	HI 4	8.17	42.82	HI 11
United Kingdom	3135	2.79	1124.46	HI 5	64.43	48.66	HI 6
Australia	1258	1.19	1058.03	HI 6	22.99	54.71	HI 3
Denmark	270	0.26	1019.64	HI 7	5.59	48.27	HI 7
Greece	286	0.29	984.51	HI 8	10.77	26.55	HI 21
Portugal	289	0.30	972.74	HI 9	10.83	26.68	HI 20
Belgium	482	0.51	947.70	HI10	11.41	42.25	HI 12
Spain	1600	1.69	946.75	HI 11	48.56	32.95	HI 17
Singapore	441	0.49	905.73	HI 12	5.78	76.27	HI 2
Austria	348	0.42	836.74	HI 13	8.71	39.95	HI 13
Canada	1377	1.67	822.58	HI 14	35.36	38.94	HI 15
Ireland	256	0.32	789.39	HI 15	4.95	51.69	HI 4
Sweden	392	0.50	786.99	HI 16	9.88	39.67	HI 14
Netherlands	629	0.87	726.41	HI 17	17.02	36.96	HI 16
Norway	238	0.36	652.59	HI 18	5.27	45.20	HI 10
Finland	153	0.24	639.63	HI 19	5.50	27.83	HI 19
South Africa	384	0.74	521.53	UMI 2	54.30	7.07	UMI 5
Turkey	848	1.67	507.78	UMI 3	80.27	10.56	UMI 2
Hungary	135	0.27	504.48	HI 20	9.87	13.67	HI 26
United States	9018	18.56	485.88	HI 21	324.00	27.83	HI18
France	1236	2.74	451.59	HI 22	66.84	18.49	HI 25
Iran	649	1.46	444.83	UMI 4	82.80	7.84	UMI 3
Pakistan	430	0.99	435.13	LMI[f] 1	202.00	2.13	LMI 3
Chile	181	0.44	415.04	HI 23	17.65	10.25	HI 29
Germany	1623	3.98	407.89	HI 24	80.72	20.11	HI 24
Morocco	109	0.28	385.43	LMI 2	33.66	3.24	LMI2
Romania	160	0.44	362.81	HI 25	21.60	7.41	HI 32
Saudi Arabia	623	1.73	359.91	HI 26	28.16	22.12	HI 23
Poland	368	1.05	349.81	HI 27	38.52	9.55	HI 31
Brazil	1058	3.14	337.48	UMI 5	205.82	5.14	UMI 6
Qatar	105	0.33	313.90	HI 28	2.26	46.50	HI 9
Czech Republic	109	0.35	310.63	HI 29	10.64	10.24	HI 30
Bangladesh	190	0.63	302.36	LMI 3	156.19	1.22	LMI 6
South Korea	574	1.93	297.56	HI 30	50.92	11.27	HI 28
Egypt	325	1.11	294.12	LMI 4	94.67	3.43	LMI 1
Vietnam	169	0.59	284.08	LMI 5	95.26	1.77	LMI 4

Country	n	GDP[a] in US $1000 billion	R_{GDP}[b]	Rank R_{GDP}	Population in millions	R_{POP}[c]	Rank R_{POP}
Taiwan	315	1.13	280.00	HI 31	23.46	13.42	HI 27
Malaysia	234	0.86	270.90	UMI 6	30.95	7.56	UMI 4
China	5053	21.27	237.56	UMI 7	1373.54	3.68	UMI 8
Colombia	159	0.69	230.30	UMI 8	47.22	3.37	UMI 9
United Arab Emirates	153	0.67	229.32	HI 32	5.93	25.81	HI 22
India	1847	8.72	211.79	LMI 6	1266.88	1.46	LMI 5
Argentina	173	0.88	196.73	UMI 9	43.89	3.94	UMI 7
Nigeria	211	1.09	193.76	LMI 7	186.05	1.13	LMI 7
Mexico	310	2.31	134.37	UMI 10	123.17	2.52	UMI 10
Thailand	144	1.16	124.03	UMI 11	68.20	2.11	UMI 11
Japan	605	4.93	122.67	HI 33	126.70	4.77	HI 33
Russia	269	3.75	71.83	UMI 12	142.36	1.89	UMI 12
Indonesia	187	3.03	61.76	UMI 13	258.32	0.72	UMI 13

[a]GDP: gross domestic product.

[b]R_{GDP}: ratio of number of articles and GDP in US $1000 billion.

[c]R_{POP}: ratio of number of articles and population in millions.

[d]UMI: upper-middle-income [23].

[e]HI: high income [23].

[f]LMI: lower-middle-income [23].

Funding Factors

In total, 17,590 articles received 27,150 grants from many governments and other funding agencies, including universities, hospitals, research institutions, nonprofit organizations, and private companies, among others (Table 3).

Table 3. Coronavirus research funding in 2020.

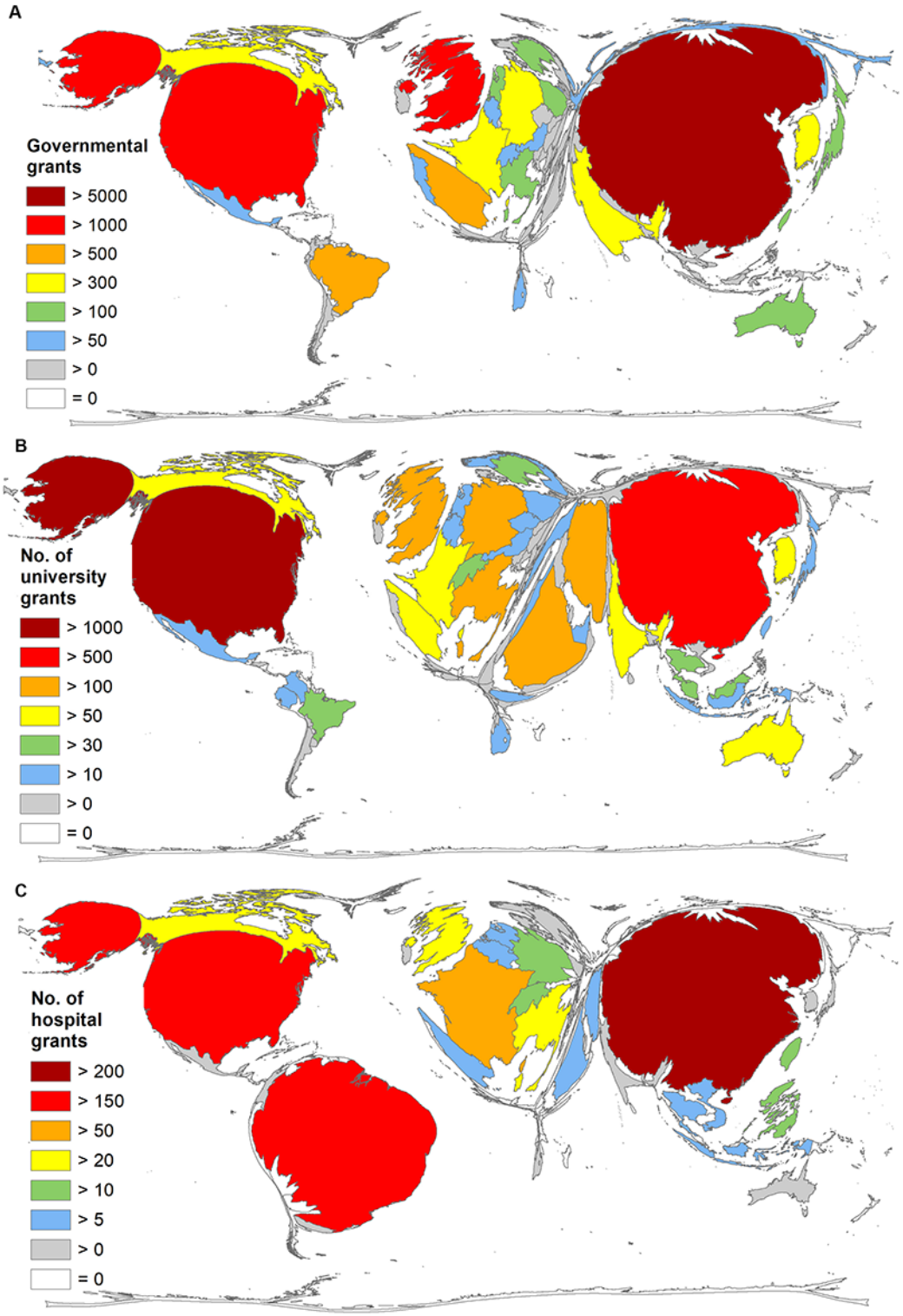
Funder	Number of grants
Governments	17,334
Universities/collages	4264
Trust/foundations	2004
International	967
Companies	947
Hospitals/health care	833
Societies/associations	379
Research institutes	186
Nonprofit organizations	95
Charities	52
Banks (private, nonprivate)	42
Networks/platforms	36
Parishes/churches	6
Publishers/journals	3
Unions	2

The government that funded the most CoV research was that of China (number of grants=6342), followed by the United States (number of grants=3983), the United Kingdom (number of grants=1179), Spain (number of grants=589), and Brazil (number of grants=504; Figure 7A). Thus, China — as the only country — had awarded more grants than published articles,

resulting in a rate of 1.26 grants per article. Applying the methodological 30-article threshold, the next highest rate was achieved by South Korea (0.58 grants per article), followed by Brazil (0.48 grants per article), the United States (0.44 grants per article), and the Czech Republic (0.44 grants per article).

Looking at university funding, the United States supported the most CoV research (number of grants=1023; Figure 7B), followed by China (number of grants=899), the United Kingdom (number of grants=215), Saudi Arabia (number of grants=173), and Italy (number of grants=164). Analysis of hospital grants showed China at the top (number of grants=222; Figure 7C), followed by Brazil (number of grants=161), the United States (number of grants=160), France (number of grants=53), and Italy (number of grants=27).

Figure 7. Density-equalizing maps of funding for coronavirus-related research, including (A) governmental funds, (B) university funds, and (C) hospital funds.



In the analysis of individual funders, the US National Institutes of Health led the way, followed by the Chinese National Natural Science Foundation, the European Union, the British National Institutes for Health Research, and the Wellcome Trust based in London, United Kingdom. The top 10 funding organizations are listed in Table 4.

Table 4. Top 10 funders supporting coronavirus research.

Funder	Number of grants
US National Institutes of Health (NIH)	2505
National Natural Science Foundation of China (NSFC)	1467
European Union (EU)	861
UK National Institute for Health Research (NIHR)	390
French National Research Agency (ANR)	343
Wellcome Trust United Kingdom	192
Brazilian National Council for Scientific and Technological Development (CNPq)	169
Chinese Academy of Sciences (CAS)	132
Bill & Melinda Gates Foundation	120
Chinese Academy of Medical Science (CAMS)	110

Discussion

Principal Findings

Undoubtedly, it is not surprising that the year 2020 was marked by an immense increase in the number of publications on CoV that addressed SARS-CoV-2 infections. Nevertheless, the scale of increase is surprising: from 289 articles in 2019 to 32,402 articles in 2020 — more than a 100-fold growth.

The global COVID-19 pandemic threatens all countries' societies with high incidence and mortality rates, so the scientific community vehemently sought solutions to contain this viral infection and the resulting symptoms through various approaches. These followed the interests of the national governments, which funded scientific endeavors just as highly.

Nevertheless, the proportion of original articles was relatively low compared to publications on other biomedical topics that are currently the talk of the town, such as climate change, where the share of articles is almost 70%, while the proportion of original research documented in CoV publications is less than 50% with a large proportion of letters and editorial material. A comparison between US-American and Chinese publications showed that the United States followed this result with nearly 70% of the article share, while Chinese articles accounted for over 90% [24]. Under the conditions of the present study, this pattern seems to be overridden in favor of shorter document types — also in the sense of rapid publication.

A closer look at the published document types also revealed that 15 papers had to be retracted although they had not even been published for a year. This demonstrates the immediacy with which results had to be published and the susceptibility to error that this occasionally entails. The pressure to be the first to publish on a new approach and deliver new results, and not be overtaken by colleagues working on the same topic, became very exigent in research on CoV in the 2020s.

The citation speed is similarly rapid. Three high-impact articles with more than 5000 citations were published in 2020 — remarkable because these papers have not yet had a year to be read or cited. The Chinese author groups reported clinical features of COVID-19 patients hospitalized in Wuhan [13-15]. Certainly, these findings have been taken up as background information for many subsequent articles.

Nevertheless, this extremely rapid publication activity, albeit at a much lower level, could also be demonstrated for the SARS and MERS epidemics, with a notable loss of research effort once the threat had subsided [4]. Had this not occurred, the state of knowledge at the onset of the SARS-CoV-2 outbreak would certainly have been sounder and the resulting measures better scientifically validated. These patterns should therefore be considered when the current pandemic has hopefully soon subsided.

It is also not surprising that in early 2020, most articles were submitted by Chinese authors reporting clinical manifestations that occurred during the outbreak in Wuhan, which was also noted in a study using the COVID-19 Open Research Dataset [25]. Subsequently, US-American researchers overtook the Chinese in terms of volume and citations received. Nevertheless, China has received the most citations so far, which is because of early publication. Articles published later have not had the time to receive so many citations. How these numbers will continue to evolve is for later studies to show. It is also not surprising that the share of Italian articles decreased slightly, considering the extreme incidence rates in Italy at the beginning of 2020. As the numbers of COVID-19 cases and mortality rates dropped, so did study numbers in Italy. As a result, Italy's share was overtaken by the British, which is not surprising due to the usually high share of British articles corresponding to the highly developed scientific infrastructure. However, the evaluation of the relationship between research performance and the number of cases showed that Italy was still in second place in the ranking of our study. Here, China was able to present itself best. The United States, which ranks first in absolute numbers in our analysis, together with India and Brazil showed a rather negative deviation due to by far the highest number of cases.

In this context, the success of the Chinese COVID-19 control must be pointed out. Certainly, the centralized epidemic response system functioning and radical surveillance had played its part. In addition, the SARS epidemic, which was accompanied by a huge mortality rate in China in 2013, was not long ago, so the awareness and compliance of the Chinese population were still very high. Both lead to an extremely fast and strict response [26,27]. In particular, in contrast to the policy response of the government of the United States, which was sparse and delayed under the regime of former President Trump, who even refused to wear a mask for a long time, China had a very quick response and was stringent with its containment measures.

Nevertheless, the US research output passed China's during 2020. It is partly assumed that the key position of US-American science will soon be overtaken by Chinese research. However, the scientific infrastructure of the United States is profound and well prepared for rapid adaptation and ramp-up. Currently, both countries are certainly competing for the top spot in global research performance.

The dynamic nature of global CoV publication output led to the similarly dynamic development of its perception in the scientific community, resulting in enormous citation numbers already in the first year. The outbreak location in China and the reporting of hospitalized patients there with the first COVID-19 symptoms explain the high citation rate of Chinese articles, which was not presentable for other research topics.

The recognition that the virus forms a sister clade to SARS-CoV led to the taxonomy of SARS-CoV-2 by the Coronaviridae Study Group (CSG) of the International Committee on Taxonomy of Viruses. The involvement of the Netherlands in the CSG contributes to its high citation rate (rank 2 behind China) because the term is internationally accepted, and subsequent publications naturally are in unison on the designation of SARS-CoV-2 [28]. In addition to the Netherlands, Russia, among others, was also involved in this taxonomy procedure, ranking therewith third in national citation rates. Moreover, Russia, together with Germany (rank 16) and Austria (rank 4), was also involved in the highly cited cell biology study by Hoffman et al [21]. This study identifies angiotensin-converting enzyme 2 as the entry receptor and a cellular serine protease as priming the spike protein with the possibility of its blocking by proven inhibitors. Since the research field "biology and biochemistry" was identified as one of the 4 focus clusters, the importance and the level of its perception are explainable. The other key clusters addressing mental and physical health impairments as well as public health and epidemiology issues are related to the other high-impact studies identified.

When the publication numbers are related to the number of COVID-19 cases, other countries showed up in leading positions. With New Zealand, Vietnam, China, Australia, and Thailand at the top, parts of the eastern and southeastern world showed leading performances due to their very low case numbers in combination with relatively high publication numbers. Countries' awareness of the problem influences not only the containment of the pandemic but also their research efforts and the amount of government spending on research and development. The success of achieving low case numbers was explained by the rapid implementation of nonpharmaceutical measures, determined responses, rigor and brevity of containment measures, and testing strategies by public health authorities [29,30]. Vietnam's effective response is also explained by its early preparation and strict control measures such as contact tracing, isolation, and mass testing combined with border closures. The same is true for other countries that have been successful in reducing the COVID-19 spread [31]. The success of Thailand was furthermore explained by hospitalizing any person with SARS-CoV-2 infection even without symptoms and also by demographic and environmental reasons, such as the high proportion of people living in rural areas and spending much time outdoors [32].

Africa has the lowest confirmed case rate of any continent. Only South Africa reported numbers among the 20 most-affected countries. [3] This is also reflected in the corresponding publication performance and also the related distorted maps. Beneficial demographic and geographic factors have been shown to have a significant negative correlation with the number of COVID-19 cases (eg, population density, temperature) [33].

Nevertheless, the low incidence rates in countries with poor health care systems and the associated low quality of the tests are often questioned. High risk is associated with misdeclaration or misreporting of the COVID-19 development. Therefore, globally transparent and traceable reporting is of immense importance, and the development in these regions must be closely monitored in the future [34].

On the other hand, regions with extremely high case numbers, such as the United States, India, Brazil, and some European countries, were proportionally more involved in CoV research. That is also shown by the correlation between COVID-19 cases and article numbers, which is highly significant. Only Russia, with the fourth highest case numbers, fell slightly behind but still shows an enhanced contribution rate to CoV research compared to other scientific topics.

This is also made possible by the intensive funding of the governments of the severely affected countries. In line with these figures, the high Italian case numbers prompted unusually high levels of cooperation with the United States, the United Kingdom, and Spain.

Saudi Arabia, which ranks in the middle of COVID-19 case numbers, was funded primarily through its universities, resulting in a relatively high ranking in publication numbers. This is likely due to the experience gained during the MERS epidemic, which mainly affected Saudi Arabia, accounting for 77% of cases globally. The results of an earlier analysis looking at CoV research up to the COVID-19 outbreak showed that Saudi Arabia even ranked 11th in overall publication performance [4]. Jordan, where MERS cases are also occurring, also had a prominent position in this analysis when socioeconomic characteristics were included. Other countries still affected by MERS, such as Egypt and Iran [35], are also in the field of CoV research in 2020 and are in the top 30.

Conclusions

The awareness and preparedness of countries affected by previous CoV epidemics also led to high interest in CoV research during the 2020 COVID-19 crisis. Although disease containment may have led to a rapid decline in publication numbers, this experience appears to have been so fresh that the propensity for ad hoc research was high. On the other hand, maintaining a reasonable level of interest would have resulted in a better scientific baseline for all containment efforts at the onset of COVID-19 worldwide. This lesson learned should be kept in mind for all future research planning.

The results of this study demonstrate the extraordinary momentum of CoV research in 2020 due to the ongoing global spread of COVID-19. They also reveal the need for continued interest and dedication by scientists even after pandemics are contained. It is to expect that the next pandemic will come and also become a threat anywhere in the world. Well-prepared and sound scientific support enables decisive measures. The experience of highly developed scientific nations must be linked with that of less developed research structures to be of global benefit.

Authors' Contributions

DK and DAG conceptualized the study and the methodology. DK, MB, and DB performed the investigation. DK created the visualizations of the results. The original manuscript draft was written by DK and MB, and final review and editing were performed by DK, MB, DB, and DAG.

Conflicts of Interest

None declared.

References

1. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun 2020 May;109:102433 [FREE Full text] [doi: 10.1016/j.jaut.2020.102433] [Medline: 32113704]
2. Scarpa R, Caso F, Costa L, Passavanti S, Vitale MG, Trojaniello C, et al. May the analysis of 1918 influenza pandemic give hints to imagine the possible magnitude of Corona Virus Disease-2019 (COVID-19)? J Transl Med 2020 Dec 22;18(1):489 [FREE Full text] [doi: 10.1186/s12967-020-02673-6] [Medline: 33353549]
3. Coronavirus Disease (COVID-19) Dashboard. World Health Organization. URL: https://covid19.who.int [accessed 2021-01-18]
4. Klingelhöfer D, Braun M, Brüggmann D, Groneberg DA. Coronavirus: An insight into global research until outbreak of COVID-19 and its implications for the future. J Glob Health 2020 Dec;10(2):020508 [FREE Full text] [doi: 10.7189/jogh.10.020508] [Medline: 33110591]
5. Pal JK. Visualizing the knowledge outburst in global research on COVID-19. Scientometrics 2021 Mar 06;126(5):1-21 [FREE Full text] [doi: 10.1007/s11192-021-03912-3] [Medline: 33716352]
6. Farooq R, Rehman S, Ashiq M, Siddique N, Ahmad S. Bibliometric analysis of coronavirus disease (COVID-19) literature published in Web of Science 2019–2020. J Fam Community Med 2021;28(1):1. [doi: 10.4103/jfcm.jfcm_332_20]
7. Wang J, Hong N. The COVID-19 research landscape: Measuring topics and collaborations using scientific literature. Medicine (Baltimore) 2020 Oct 23;99(43):e22849 [FREE Full text] [doi: 10.1097/MD.0000000000022849] [Medline: 33120818]
8. Groneberg-Kloft B, Fischer TC, Quarcoo D, Scutaru C. New quality and quantity indices in science (NewQIS): the study protocol of an international project. J Occup Med Toxicol 2009 Jun 26;4:16 [FREE Full text] [doi: 10.1186/1745-6673-4-16] [Medline: 19555514]
9. Gastner MT, Newman MEJ. From The Cover: Diffusion-based method for producing density-equalizing maps. Proc Natl Acad Sci U S A 2004 May 18;101(20):7499-7504 [FREE Full text] [doi: 10.1073/pnas.0400280101] [Medline: 15136719]
10. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 2010 Aug 31;84(2):523-538 [FREE Full text] [doi: 10.1007/s11192-009-0146-3] [Medline: 20585380]
11. Singh VK, Singh P, Karmakar M, Leta J, Mayr P. The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. Scientometrics 2021 Mar 26;126(6):5113-5142. [doi: 10.1007/s11192-021-03948-5]
12. Mongeon P, Paul-Hus A. The journal coverage of Web of Science and Scopus: a comparative analysis. Scientometrics 2015 Oct 19;106(1):213-228. [doi: 10.1007/s11192-015-1765-5]
13. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet 2020 Feb;395(10223):497-506. [doi: 10.1016/s0140-6736(20)30183-5]
14. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
15. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA 2020 Mar 17;323(11):1061-1069 [FREE Full text] [doi: 10.1001/jama.2020.1585] [Medline: 32031570]

16. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar 28;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]
17. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, China Novel Coronavirus InvestigatingResearch Team. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733 [FREE Full text] [doi: 10.1056/NEJMoa2001017] [Medline: 31978945]
18. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The Lancet 2020 Feb;395(10223):507-513. [doi: 10.1016/S0140-6736(20)30211-7]
19. Zhou P, Yang X, Wang X, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020 Feb 03;579(7798):270-273. [doi: 10.1038/s41586-020-2012-7]
20. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]
21. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell 2020 Apr 16;181(2):271-280.e8 [FREE Full text] [doi: 10.1016/j.cell.2020.02.052] [Medline: 32142651]
22. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet 2020 Feb;395(10224):565-574. [doi: 10.1016/S0140-6736(20)30251-8]
23. Data: World Bank Country and Lending Groups. The World Bank. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [accessed 2021-05-01]
24. Zhu J, Liu W. Comparing like with like: China ranks first in SCI-indexed research articles since 2018. Scientometrics 2020 May 25;124(2):1-10 [FREE Full text] [doi: 10.1007/s11192-020-03525-2] [Medline: 32836521]
25. Abd-Alrazaq A, Schneider J, Mifsud B, Alam T, Househ M, Hamdi M, et al. A Comprehensive Overview of the COVID-19 Literature: Machine Learning-Based Bibliometric Analysis. J Med Internet Res 2021 Mar 08;23(3):e23703 [FREE Full text] [doi: 10.2196/23703] [Medline: 33600346]
26. Burki T. China's successful control of COVID-19. The Lancet Infectious Diseases 2020 Nov;20(11):1240-1241. [doi: 10.1016/s1473-3099(20)30800-8]
27. Salzberger B, Glück T, Ehrenstein B. Successful containment of COVID-19: the WHO-Report on the COVID-19 outbreak in China. Infection 2020 Apr 17;48(2):151-153 [FREE Full text] [doi: 10.1007/s15010-020-01409-4] [Medline: 32185635]
28. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 2020 Apr 2;5(4):536-544 [FREE Full text] [doi: 10.1038/s41564-020-0695-z] [Medline: 32123347]
29. Berger E, Reupert A. The COVID-19 pandemic in Australia: Lessons learnt. Psychol Trauma 2020 Jul;12(5):494-496. [doi: 10.1037/tra0000722] [Medline: 32525385]
30. Robert A. Lessons from New Zealand's COVID-19 outbreak response. The Lancet Public Health 2020 Nov;5(11):e569-e570. [doi: 10.1016/s2468-2667(20)30237-1]
31. Van Tan L. COVID-19 control in Vietnam. Nat Immunol 2021 Mar 24;22(3):261-261. [doi: 10.1038/s41590-021-00882-9] [Medline: 33627879]
32. Ratcliffe R. How have Thailand and Cambodia kept Covid cases so low? The Guardian. 2020 Dec 16. URL: https://www.theguardian.com/world/2020/dec/16/thailand-cambodia-covid-19-cases-deaths-low [accessed 2020-04-01]
33. Nguimkeu P, Tadadjeu S. Why is the number of COVID-19 cases lower than expected in Sub-Saharan Africa? A cross-sectional analysis of the role of demographic and geographic factors. World Dev 2021 Feb;138(12):105251 [FREE Full text] [doi: 10.1016/j.worlddev.2020.105251] [Medline: 33106726]
34. Weininger I. Zwischen Ignoranz und Notstand: Der Umgang mit COVID-19 in Südostasien. Konrad Adenauer Striftung. 2020 Apr 07. URL: https://www.kas.de/documents/252038/7938566/Zwischen+Ignoranz+und+Notstand.+Der+Umgang+mit+Covid-19+in+Südostasien.pdf/4a32c797-aa35-6230-dca7-65b29173c7d2?version=1.0&t=1586257037760 [accessed 2021-04-01]
35. Middle East Respiratory Syndrome. World Health Organization. URL: https://www.emro.who.int/health-topics/mers-cov/mers-outbreaks.html [accessed 2021-04-01]

Abbreviations

CoV: coronavirus
CSG: Coronaviridae Study Group
DEMP: density-equalizing map projections
GDP: gross domestic product
MERS: Middle East Respiratory Syndrome

NewQIS: New Quality and Quantity in Science
ppp $: purchasing power parity in current international $
R_{GDP}: ranked gross domestic product
SARS: Severe Acute Respiratory Syndrome
WoS: Web of Science

Edited by C Basch; submitted 25.05.21; peer-reviewed by B Loring, J Wang; comments to author 15.06.21; revised version received 22.06.21; accepted 08.07.21; published 08.09.21.

Please cite as:
Klingelhöfer D, Braun M, Brüggmann D, Groneberg DA
The Pandemic Year 2020: World Map of Coronavirus Research
J Med Internet Res 2021;23(9):e30692
URL: https://www.jmir.org/2021/9/e30692
doi:10.2196/30692
PMID:34346891

Viewpoint

Data Empowerment of Decision-Makers in an Era of a Pandemic: Intersection of "Classic" and Artificial Intelligence in the Service of Medicine

Gil A Geva[1*], MD; Itay Ketko[1,2*], MSc; Maya Nitecki[1,3], MD; Shoham Simon[4], MSc; Barr Inbar[5], MSc; Itay Toledo[5], BSc; Michael Shapiro[1], MD; Barak Vaturi[5], BSc; Yoni Votta[5], BSc; Daniel Filler[5], BSc; Roey Yosef[5], BSc; Sagi A Shpitzer[1], MD; Nabil Hir[1], MSc; Michal Peri Markovich[1,6], VMD; Shachar Shapira[1,3,7], MD, MHA, MA; Noam Fink[1], MD, MHA; Elon Glasberg[1,8], MD, MHA; Ariel Furer[1,3], MD, MBA

[1]Medical Corps, Israel Defense Force, Ramat Gan, Israel

[2]Heller Institute of Medical Research, Sheba Medical Center, Tel-HaShomer, Ramat Gan, Israel

[3]Department of Military Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel

[4]Planning Directorate, Israel Defense Force, Tel Aviv, Israel

[5]Computer and IT Directorate, Israel Defense Force, Tel Aviv, Israel

[6]Israel Veterinary Services, Ministry of Agriculture and Rural Development, Ramat Gan, Israel

[7]Institute for Research in Military Medicine, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel

[8]The Azrieli Faculty of Medicine, Bar-Ilan University, Safed, Israel

[*]these authors contributed equally

Corresponding Author:
Ariel Furer, MD, MBA
Medical Corps
Israel Defense Force
IDF Medical Corps Headquarters
Tel HaShomer
Ramat Gan, 02149
Israel
Phone: 972 529277372
Email: furera@gmail.com

Abstract

Background: The COVID-19 outbreak required prompt action by health authorities around the world in response to a novel threat. With enormous amounts of information originating in sources with uncertain degree of validation and accuracy, it is essential to provide executive-level decision-makers with the most actionable, pertinent, and updated data analysis to enable them to adapt their strategy swiftly and competently.

Objective: We report here the origination of a COVID-19 dedicated response in the Israel Defense Forces with the assembly of an operational Data Center for the Campaign against Coronavirus.

Methods: Spearheaded by directors with clinical, operational, and data analytics orientation, a multidisciplinary team utilized existing and newly developed platforms to collect and analyze large amounts of information on an individual level in the context of SARS-CoV-2 contraction and infection.

Results: Nearly 300,000 responses to daily questionnaires were recorded and were merged with other data sets to form a unified data lake. By using basic as well as advanced analytic tools ranging from simple aggregation and display of trends to data science application, we provided commanders and clinicians with access to trusted, accurate, and personalized information and tools that were designed to foster operational changes and mitigate the propagation of the pandemic. The developed tools aided in the in the identification of high-risk individuals for severe disease and resulted in a 30% decline in their attendance to their units. Moreover, the queue for laboratory examination for COVID-19 was optimized using a predictive model and resulted in a high true-positive rate of 20%, which is more than twice as high as the baseline rate (2.28%, 95% CI 1.63%-3.19%).

Conclusions: In times of ambiguity and uncertainty, along with an unprecedented flux of information, health organizations may find multidisciplinary teams working to provide intelligence from diverse and rich data a key factor in providing executives relevant and actionable support for decision-making.

(J Med Internet Res 2021;23(9):e24295) doi:10.2196/24295

KEYWORDS

COVID-19; medical informatics; decision-making; pandemic; data; policy; validation; accuracy; data analysis

Introduction

SARS-CoV-2 was first reported in Wuhan, China, by early December 2019 [1,2] and by March 11, 2020, COVID-19 was declared a pandemic by the World Health Organization [3]. To date, the global reported death toll of the disease exceeds 1.5 million individuals, while many countries continue to struggle with the pandemic and show an ascending number of cases and mortality rate [4].

Under immense uncertainty, and based on ominous models describing disease exponential expansion [5,6], Israel, like the many other countries, began preparing for the potential threat of the pandemic [7]. The Israel Defense Forces (IDF), one of the largest organizations in the country, was responsible for all health care–related aspects of its own personnel and also took action on the national front and maintained full preparedness for any of its assignments and duties—military and civilian.

One of the key understandings early during response-planning was the paramount importance of reliable data as a substrate for policy changes and decision-making [8,9]. With this mindset, the IDF surgeon general commissioned the establishment of an operational IDF Data Center for the Campaign against Coronavirus (ID3C) in March 2020. The ID3C was ordered to collect and analyze data from all venues and produce concise, data-driven, up-to-date, and relevant recommendations to guide both medical and commanding echelon policy makers.

This paper describes the assembly of the ID3C, the architecture of data utilization, and few samples of the deliverables for which the center was responsible. With much more to learn about the COVID-19 response, worldwide authorities and medical communities strive for quality data as the base for their decision- and policy-making [8-10]. We believe that sharing the lessons from our experience of creating a data center amid a raging pandemic can help other organizations to respond currently an in future to this global event and to other incidents where prompt effective action needs to be based on robust data.

Methods

Methods Overview

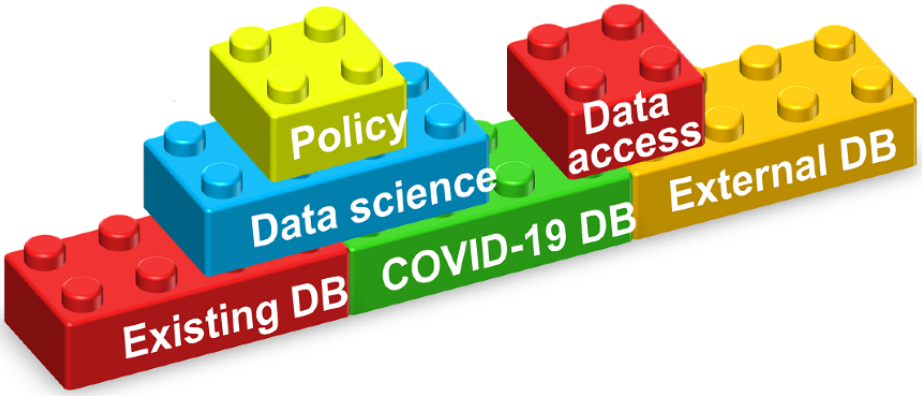
The first step of designing the ID3C was rooted in several key assumptions. We assumed that the disease will have distinct characteristics within our organization compared with the general population, mostly owing to the age of the young population and a relatively low rate of individuals with known risk factors for severe disease [11,12]. Notably, we considered this a dynamic process, as the definition of risk factors kept changing with more evidence accumulating globally at a very high pace [13]. In addition, based on reports suggesting a relatively high rate of asymptomatic individuals among young adults [14,15], we created a digital pathway to shorten the time from patient diagnosis to the completion of contact-tracing. Throughout the process, we weighted the impact of disease spread and policy modifications against the need for preparedness, and assessed the cost-effectiveness of the measures taken by using accurate data.

Modus Operandi

First, we launched an extensive mapping of data sources and assessed the quality of data required to plan a unified data lake diverse enough to establish prediction models to act on. Next, we prioritized mirroring of trends and basic insights from aggregated data to serve as actionable information for executives. The design of a digital platform that will allow them to easily access it followed with the main vectors of action (Figure 1):

1. Harnessing existing IDF data sets: the IDF has a comprehensive digital data set of all service persons derived from decades-long data collection as part of the recruitment process of draftees, as well as electronic medical records of all military personnel. These data were merged with other "nonmedical" parameters on an individual and unit level to deepen the understanding of factors related to the risk of infection and deterioration to severe disease.
2. Construction of new COVID-19–specific information platforms and sensors.
3. Importing and merging of relevant external data provided by the Ministry of Health.
4. Unification of all data sets to 1 "data lake" and extraction of actionable knowledge followed by the use of data science tools to provide a means for pandemic management. The main attempt was to identify trends and patterns to direct the efforts to negate morbidity, especially among units critical to the competence of the organization.
5. Provision of direct data accessibility to executives: a central study objective was to improve executive-level involvement by promotion of data-driven decision-making.

Figure 1. Schematic overview of the rationale for the Israel Defense Forces' Data Center for the Campaign against Coronavirus. The operation was based on the ability to bring together large databases into 1 central COVID-19 data lake. Consequently, 2 parallel efforts were conducted: the first enabled executives and commanders in the organization to gain direct access to actionable data and insights, and the second used advanced analytics to provide additional level of understanding, including using data science–based model to influence policy changes. DB: database.



Team Assembly

The ID3C was led by 3 executives from complementary disciplines: a senior physician commanding officer, a senior intelligence officer, and a senior information technology and data officer. The ID3C comprised 5 small, specified task forces, each consisting of medical personnel, a data specialist, and a project manager. This was a first-of-its-kind military joint venture organization synergizing Medical Corps with the Computer and IT Directorate and intelligence officers. Lean task forces were responsible for gathering relevant published and collected data, providing clear insights and analyzing merged data sets, while other teams oversaw the design of "executive-friendly" dashboards.

Data Collection

To obtain relevant insights, the IDF COVIDataLake (CDL) was created. The CDL is a fusion of existing data sets, existing COVID-19–oriented data sets, and new data sets designed by the ID3C. Existing data sets included medical records, sociodemographic data obtained from a mandatory prerecruitment survey, transportation usage, occupational characteristics, unit characteristics (ie, unit under quarantine conditions and age distribution), service characteristics (ie, daily attendance, close encounter with civilians, etc). These data helped estimate interactions among particular individuals once a suspicion of infection arose. Newly formed COVID-19–oriented data sets included the following: laboratory-based COVID-19 reverse transcription polymerase chain reaction (RT–PCR) test results of the IDF; central command and control platform, which documented all suspected, quarantined, and confirmed cases, as well as all registration of all calls to the call center with structured fields to report on symptoms and exposure to confirmed patients with COVID-19; and special data annotations to highlight quarantined and hospitalized personnel.

Based on the concept of the triple T ("trace, test, treat") [16,17], throughout the ID3C workflow, a system of surveys was developed to collect data deemed essential for early identification of suspected individuals; these included a habits and baseline risk factors survey and a daily survey with a single "yes/no" question regarding exposure and symptoms. For those who had an indication of exposure or symptoms, a more detailed survey was prompted automatically.

To improve the accuracy and efficiency of epidemiological investigations, a designated web-based platform designed by public health officers was introduced where all data from conducted epidemiological investigations were maintained in a structured database.

Lastly, as national-level data collection improved drastically, several sources of data were found to be relevant for the IDF, including daily reports of "hot zones"; a national patients registry elaborating on symptom distribution and disease severity by age; public symptoms surveys conducted through open-source apps, health maintenance organizations, and the Ministry of Health [18]; and risk models for progression to severe disease developed by the "Clalit health services" [19].

Data Security and Ethics

The CDL was treated as a medical database in accordance with national regulations. Data collection and analysis were approved by institutional review board of the IDF Medical Corps (submission 2082-2020).

Data Analysis and Predictive Data Science

The proportions of soldiers treated with respiratory symptoms but without a COVID-19 diagnosis were compared between units. The annotation of theses visits was available using natural language processing techniques to search through medical records and seek signs and symptoms related to COVID-19 (eg, shortness of breath, fever, cough, anosmia, and ageusia). Units were updated once an alarming increase in such visits was evident as an early sign for a possible concealed cluster of COVID-19 cases. Additionally, rates of individuals at risk for severe disease were calculated on the basis of computerized exploration of the medical records for a list for pre-existing medical conditions and treatments, in accordance with updated literature concerning risk factors for severe COVID-19. Shortlists of all at-risk individuals at the unit level were provided to the commanding level daily to reconsider the importance of the individual in terms of the unit's preparedness and functionality.

Machine learning models were used to optimize the IDF's queue for RT–PCR examination for COVID-19. The data points were split into three sets: training, testing, and validation. The validation set comprised 15% of the most recent data points. This set was used to estimate the generalization error of the model. A chronological partition was chosen so as not to underestimate the generalization error: the data were observably chronologically irregular, stemming from the rapid change in the policies, guidelines, and spread of the disease. To deal with these trends, 2 additional sources were incorporated into the data set. First, cities and towns with high contagion counts were recorded daily from public data provided by the Ministry of Health. This allowed us to approximate the suspects' risk of exposure stemming from their city of residence. Second, the number of quarantined personnel in the individual's unit was calculated from the IDF's daily attendance record. This allowed us to account for the current extent of quarantine and policies mandated in the individual's close work environment. Including these 2 variables in the data set led to an observable improvement in the models' performance. The testing set was used by the models' internal optimization procedure to evaluate intermediate trained models and to tune the models' parameters. The training set was used to train the intermediate models. Models were evaluated in accordance with the intended use case: prioritizing the RT–PCR testing queue, with the goal of increasing the proportion of positive cases among those tested. Given the typical daily laboratory capacity and queue length, the score used to evaluate the models was the proportion of true-positive cases among 10% of cases with the highest predicted risk score. This is essentially the pretest probability of the model, with a dynamic threshold (ie, the first predicted high-risk decile of the test queue, in contrast with a constant risk score threshold). Data handling and utilization were conducted using database SQL (Toad for Oracle, version 18c; Quest Software Inc) and SAS (version 9.4, the SAS Institute). Python (version 3.6, Python Software Foundation) was used throughout the study. Preprocessing and exploratory data analysis were conducted with Pandas and Seaborn Python packages. The Scikit-learn Python package was used for training and evaluation of models.

Results

Nearly 300,000 responses for daily questionnaires were recorded by the end of July 2020 with compliance variance from 50% of all personnel in some periods to less than 10% in most units routinely.

Data Sharing Platforms for Executives

One centralized digital platform was designed to provide every unit and senior commanders the following daily updated information about their unit (Figure 2):

1. Identifying individuals living in "hot zones": indeed, analysis conducted using military epidemiological data revealed that 56% of disease contraction was outside of the military base or activity.
2. Identifying individuals attending the unit, who are at risk for severe disease: a decline of ~30% was observed in the attendance of such individuals after the implementation of this tool (Figure 3).
3. Notification on units with alarming clusters of respiratory complaints.

Figure 2. ID3C 360° individual data components. The unique datasets collected during the operation of the ID3C included three layers of data in particular: data set available routinely in the IDF, including medical records, a human-resources data set with demographic and occupational characteristics, and daily attendance and logistic data including transportation usage; a new data set designated to collect COVID-19–related data such as laboratory findings, epidemiological investigations' outputs, designated surveys, and confirmed cases' medical records; and external data sets from government agencies, other health organizations, and open-source publications on broad population surveys. ID3C: Israel Defense Forces' Data Center for the Campaign against Coronavirus; IDF: Israel Defense Forces; HMO: health maintenance organization; MOH: Ministry of Health.

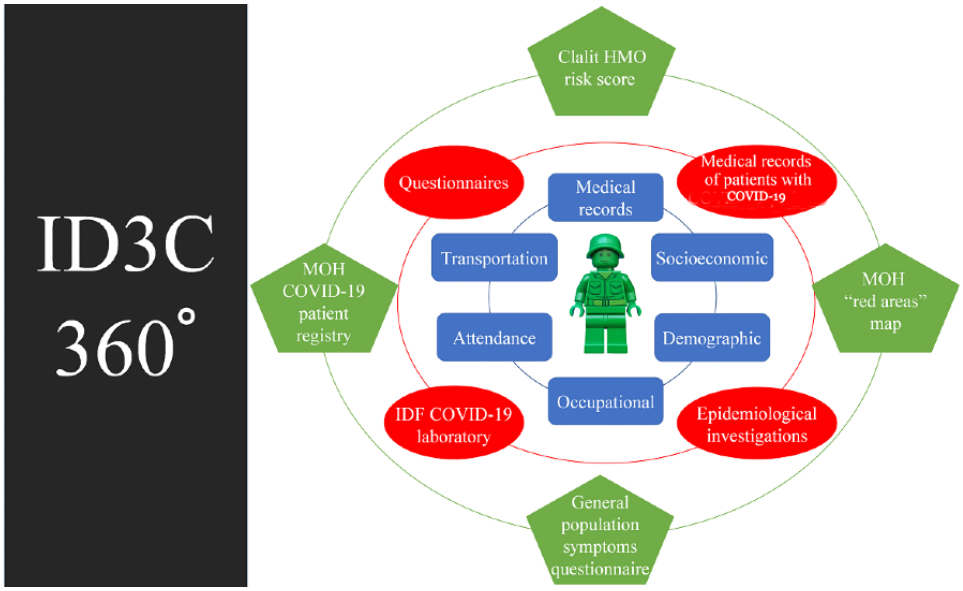
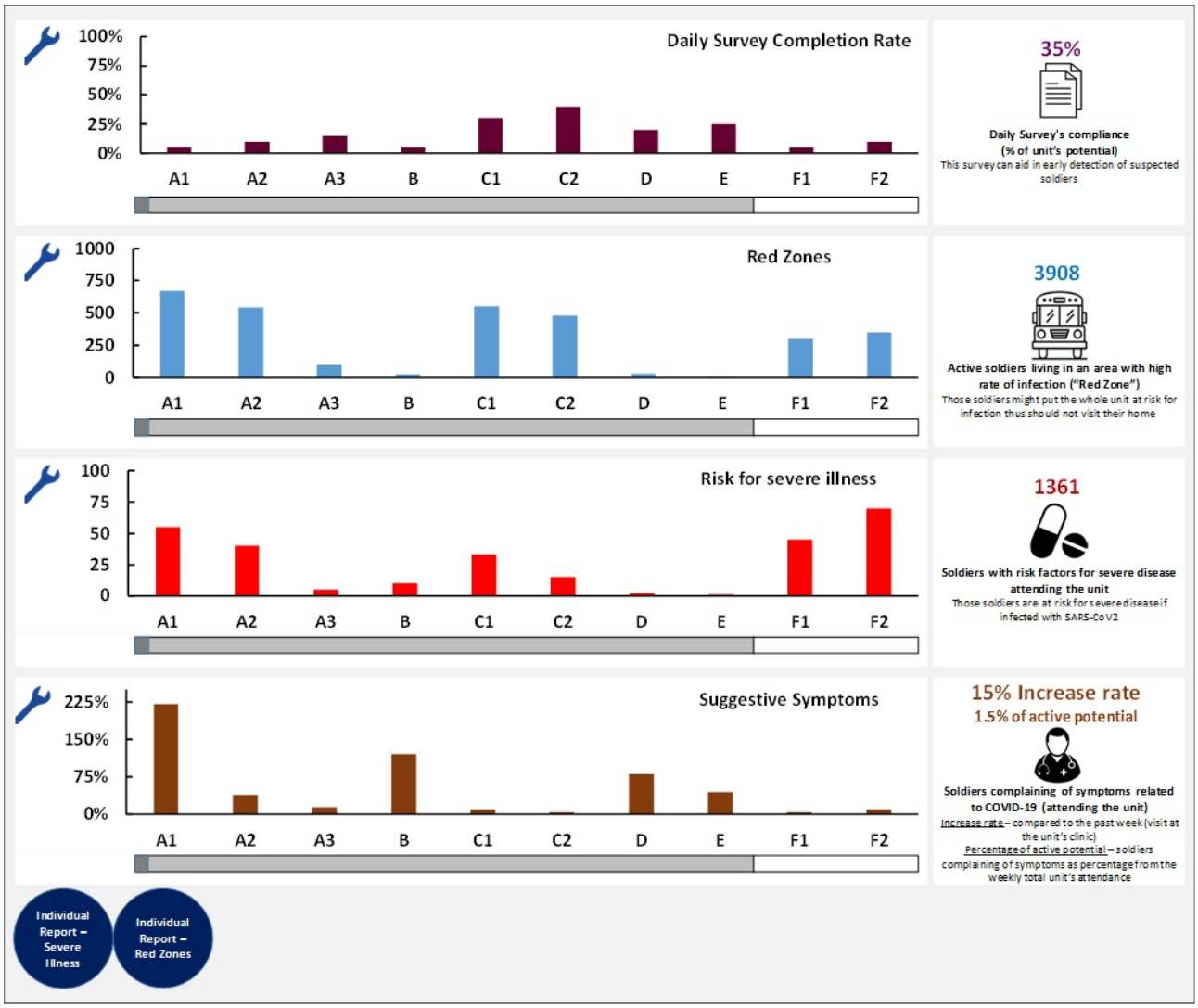


Figure 3. Daily updated main display of the Israel Defense Forces' Data Center for the Campaign against Coronavirus digital platform designed for commanders' use. This includes the daily survey completion rate, which is the proportion of the unit's potential (purple, top panel); red zones, which refers to active soldiers living in an area with a high rate of infection (blue, second panel from the top); risk for severe illness, which refers to soldiers with risk factors for severe disease if infected with SARS-CoV-2, who attend the unit (red, third panel from the top); and suggestive symptoms, which is the proportion of on-duty soldiers presenting to unit's clinic for symptoms related to COVID-19 (increase rate refers to the increase in visits compared to the past week; percentage of active potential refers to soldiers complaining of symptoms as a percentage from the weekly total unit's attendance; brown, bottom panel).



Data Sharing Platforms for Medical Personnel

Unit medical staff had access to a daily updated dashboard, which contained the following information:

1. Notification of individuals at high risk for severe disease: direct communication between the ID3C and units' primary physicians was facilitated with respect to extremely-high-risk individuals by using the Clalit algorithm.
2. Alert on individual cases that form a part of clusters presenting with respiratory symptoms, allowing physicians to revisit cases if an alarming number of patients with suspicious symptoms presented at a specific clinic.
3. Medical indicators from surveys promoted proactive summoning of individuals reporting suspicious symptoms and exposures before they approached the clinic, and those found to be at substantial risk were directed to undergo a nasopharyngeal swab test.

Optimization of the Queue for Laboratory-Based COVID-19 Testing

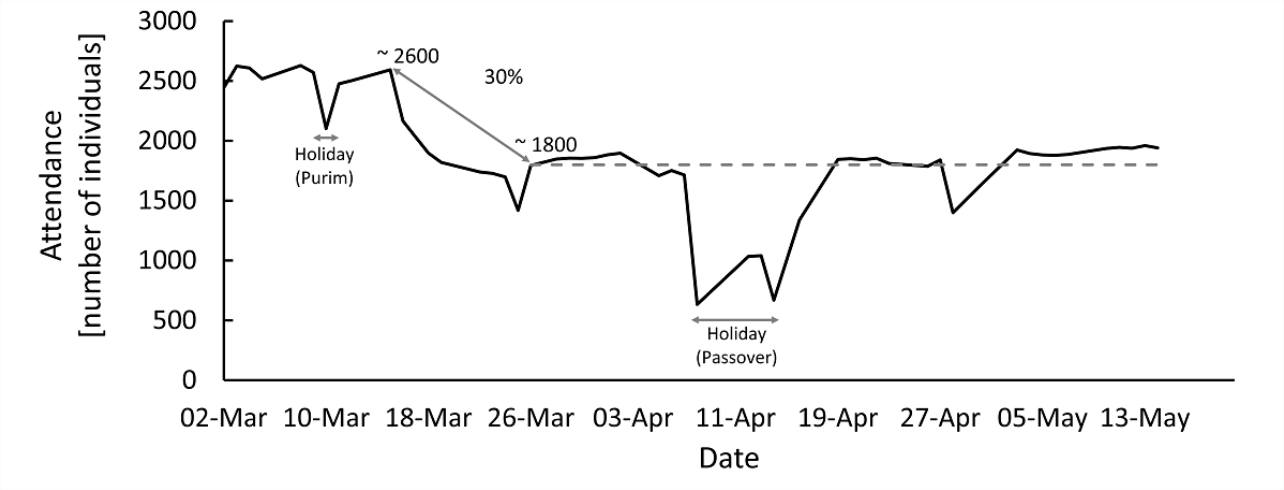
For each individual in the examination queue, which included thousands of individuals at peak periods, the model produced a risk score, which was used to prioritize the predicted high-risk suspect. Based on CDL, a prediction model was trained and tested on 10,428 examinations and validated on additional 1698 examinations. The model indicated daily the potential top 10% high-risk suspects with a very high pretest probability and prioritized the processing of their nasopharyngeal swabs. This optimization process resulted in a high true-positive rate of 20% from among the total examination results processed, which was more than twice as high (2.28%, 95% CI 1.63%-3.19%) as the baseline true-positive rate (8.8%) without using the model recommendation for the same period.

Discussion

Principal Findings

We report here the origination of a COVID-19 dedicated response by the IDF with the assembly of the ID3C. The multidisciplinary team utilized existing and newly developed platforms to collect and analyze large amounts of information for each individual in the context of SARS-CoV-2 contraction and infection. With nearly 300,000 responses for daily questionnaires merged into 1 data lake, we could identify high-risk individuals, and we observed a 30% decline in their attendance to their units (Figure 4). Moreover, we introduced an optimization model for the queue for laboratory-based COVID-19 testing and obtained a high true-positive rate, which was twice as high as the baseline rate (20% vs 8.8%, respectively).

Figure 4. The attendance of individuals deemed at high risk for severe disease in the Israel Defense Forces' units, between March 2 and May 13, 2020. The graph shows a sharp decline in attendance with the implementation of the designated unit-specific tool, allowing to monitor the individuals at risk for severe disease.



Both health care providers and leaders are in the midst of a global crisis, which is characterized by extreme uncertainty and unprecedented flux of information. In this “infodemic” and pandemic, policy makers pursue trustworthy measures to help them make strategic decisions in response to the threat [20,21].

A call for innovative, modern methods of rapidly collecting and processing data has been made by countries and organizations around the world from the beginning of the pandemic. Taiwan—a neighboring country to China, which was one of the first countries to be affected by the pandemic—took rapid action by merging the borders and customs databases with their national medical database to link clinical symptoms with flight and travel history.

As early as January 2020, Sun et al [22] and Qin et al [23] used social media searches to track the trends and spread of COVID-19 through China. Indeed, web-based discussions on social media platforms prompted an international group of rheumatologists to create the COVID-19 Global Rheumatology Alliance [24]. This alliance created a web-based registry in which clinicians and researchers around the world could share their knowledge and findings. The American Academy of Dermatology, American College of Surgeons, American Academy of Orthopedic Surgeons, Spanish Neurological Society, and others have also created specialized registries to track and analyze relevant data in their fields [25,26].

Many other methods have been implemented, such as web-based surveys to detect the spread of the disease in the population [27,28], and the use of machine learning techniques to quickly and reliably test for COVID-19 by private firms such as *Infervision*, *Diagnostic Robotics*, and others [29].

This atmosphere of uncertainty has many similarities to other emergency situations, especially those concerning security, where leaders depend on intelligence officers to provide them with data and estimations. We describe here the prompt establishment of a unique organizational structure set to provide live intelligence, using a synergism of “conventional” military intelligence methods, medical knowledge, and research, as well as data-driven artificial intelligence platforms. The center was based on multidisciplinary teaming, utilization of data, and leaning on digital platforms for providing accurate, relevant, and comprehensive information to executives. The novelty in this process covers multiple aspects and was catalyzed by the intensity of the challenge, and in fact fulfilled many of the perspectives that health care professionals had envisioned in the last decade with respect to the era of information, personalized medicine, and digital health [30-32].

Identifying the engagement of medium-tier commanders as the key to maximize the chance to make an impact dictated our pivotal effort to design tools that would not merely end up with data analysis but would be delivered almost immediately as actionable data, to affect their daily routine. Similar to the approach adopted by others, we found the delivery of accurate and timely knowledge translation an inevitable effort in light of the surging pandemic [33].

The tools used in the ID3C were designed to promote the prevention of contagion, earlier detection of disease clusters, and better control of at-risk individuals. Other data-driven tools were actually developed in response to an emerging challenge with an extreme shortage of RT–PCR testing and provided optimization to make better use of existing testing methods [34],

using computer-aided prioritization of examinations, thus increasing the proportion of positive test results and decreasing the mean diagnosis time of positive cases.

Notwithstanding, the experience of collecting large amounts of data provided valuable insights into the importance of careful planning of the delicate architecture and the quality assurance measures that are a "must-have" in this type of endeavor. Data science and machine learning are very common "buzzwords" in the discussion about the health care revolution [35-37], but one should be very cautious to not overload expectations from these set of tools, especially when the data are gathered at different paces and through different channels. Moreover, this emphasizes the need for local and global medical data sharing among organizations and countries, which would be key for enabling the utilization of emerging computational proficiencies and potentially shed light on clinical interpretation of events, thereby accelerating the discovery and validation of promising interventions [8,38-40].

A centralized medical intelligence arm, where public health officers, physicians, and data experts work together to construct the process of data collection, analysis, and reflection, was found to be very efficient and had a true synergistic value in our experience. The ability to combine the technological arm that designs and broadcasts the different applications for data collection had an additional value and enabled swift changes guided by real-time field feedback. This is further outlined when compared with conventional public health methods primarily based on manually collected data and the tendency to work in isolated professional silos, limiting prompt data analysis and action required to effectively respond to the current COVID-19–imposed challenges [41]. The agility in producing the right set of tools for constantly changing circumstances is crucial for successful efforts in the long term, as the pandemic changes and evolves.

A vital consideration we felt obligated to maintain was the high standard of medical confidentiality, highly ethical standards of data science research, and privacy of data. Despite the urgency and the need to deliver data, we insisted on and maintained 2 separate routes of data flow: one for medical personnel and the other for nonmedical executives. This allowed us to provide meaningful alerts and information at the executive level while keeping the medical staff at the center of the decision-making process, being able to consult judiciously with a very broad clinical overview, and simultaneously not breaching any medical privacy standard.

Limitations

We experienced several difficulties that are worth mentioning in this context. First, we saw very modest compliance rates to the extensive survey effort that we initiated, especially as the timing of broadcasting of the full-scale survey coincided with a sharp decline in cases and an increase in survey compliance with the substantial surge in cases. As with other noncompliance issues in medicine, it is our belief that for the time being, a cultural change is the key to deliver a discernible change. Given its potential effectiveness in the early identification of suspicious domains, a survey should be considered a routine procedure throughout the pandemic. Second, the current pandemic is very dynamic, and the nature of disease spread was different between the 2 main pandemic periods we observed. The knowledge base to understand this specific new infection is lacking and evolves constantly, and this might have a substantial effect on any mathematical model, thus necessitating cautious evaluation of models based on concurrent data. To ensure our tools are valid, we continuously evaluated the yield and accuracy of our models and modified the included parameters to adapt to the changing circumstances. Lastly, despite our extensive effort in building data-driven tools, it is of paramount importance to recall their role as advisory tools and highlight the central role of human judgment in integrating professional, medical, and organizational considerations at all times.

Conclusions

In conclusion, we described the assembly of a specialized multidisciplinary team weighing on diverse, rich data and advanced analytics to assist decision-makers with formulating policies in the context of the unexpected and unfamiliar chain of events throughout the COVID-19 pandemic. With anticipated disease spread, we believe that additional collaborative initiatives are in place to largely utilize accumulating local and global data in an attempt to develop impactful interventions.

Acknowledgments

We acknowledge the contribution of Mrs Dorit Tzur, Mrs Victoria Yoffe, Mrs Maya Simchoni, Mr Itay Klopshtock, Mr Dan Vershov, Mr Doron Stupp, Dr Daniel Gabai, Mr Moti Atkind, Dr Eva Abramovitch, Dr Inbal Galor, and Prof Gilad Twig to this study.

Authors' Contributions

GAG and IK carried out the investigation, curated the data, and drafted the manuscript. MN curated the data, conducted the formal analysis, and drafted the manuscript. S Simon conceptualized the study, was responsible for its administration, interpreted the findings, and reviewed and edited the manuscript. BI conceptualized the study and the methodology, oversaw the data science and the information technology architecture, verified the analytical methods, and reviewed and edited the manuscript. IT developed the software; performed the formal analysis; carried out the investigation; designed the models and the computational framework; handled, processed, and analyzed the data; performed the computations using data science methods; and reviewed and edited the manuscript. MS developed the software, curated the data, analyzed the output and results, and reviewed and edited the manuscript. BV developed the software, handled the data, oversaw the information technology framework design and data architecture, and

reviewed and edited the manuscript. YV developed the software, supervised and verified the information technology design, developed the models and the computational framework, and reviewed and edited the manuscript. DF developed the software; carried out the formal analysis and the investigation; handled, processed, and analyzed the data; performed computations using data science methods, and reviewed and edited the manuscript. RY performed the investigation; handled, processed, and analyzed the data; and performed computations using data science methods. SAS and NH carried out the investigation and data analysis, and reviewed and edited the manuscript. MP designed the methodology, carried out the formal analysis, interpreted the results and output design, and reviewed and edited the manuscript. NF and EG conceptualized the study and oversaw its administration, supervised the study overall, and reviewed and edited the manuscript. AF conceptualized the study and oversaw its administration, analyzed and interpreted the data, and drafted the manuscript and approved its final version for publication.

Conflicts of Interest

None declared.

References

1. Lu H, Stratton CW, Tang Y. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. J Med Virol 2020 Apr;92(4):401-402 [FREE Full text] [doi: 10.1002/jmv.25678] [Medline: 31950516]
2. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
3. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020 Mar 11. URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2021-08-10]
4. WHO Director-General's opening remarks at the media briefing on COVID-19 - 27 July 2020. World Health Organization. 2020 Jul 27. URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---27-july-2020 [accessed 2021-08-10]
5. Li Y, Liang M, Yin X, Liu X, Hao M, Hu Z, et al. COVID-19 epidemic outside China: 34 founders and exponential growth. J Investig Med 2021 Jan;69(1):52-55 [FREE Full text] [doi: 10.1136/jim-2020-001491] [Medline: 33023916]
6. Wu J, Leung K, Leung G. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020 Feb 29;395(10225):689-697 [FREE Full text] [doi: 10.1016/S0140-6736(20)30260-9] [Medline: 32014114]
7. Maor M, Sulitzeanu-Kenan R, Chinitz D. When COVID-19, constitutional crisis, and political deadlock meet: the Israeli case from a disproportionate policy perspective. Policy and Society 2020 Jul 07;39(3):442-457. [doi: 10.1080/14494035.2020.1783792]
8. Segal E, Zhang F, Lin X, King G, Shalem O, Shilo S, et al. Building an international consortium for tracking coronavirus health status. Nat Med 2020 Aug;26(8):1161-1165. [doi: 10.1038/s41591-020-0929-x] [Medline: 32488218]
9. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534 [FREE Full text] [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
10. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing. JAMA 2020 Apr 14;323(14):1341-1342. [doi: 10.1001/jama.2020.3151] [Medline: 32125371]
11. Livingston E, Bucher K. Coronavirus Disease 2019 (COVID-19) in Italy. JAMA 2020 Apr 14;323(14):1335. [doi: 10.1001/jama.2020.4344] [Medline: 32181795]
12. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
13. Zheng Z, Peng F, Xu B, Zhao J, Liu H, Peng J, et al. J Infect 2020 Aug;81(2):e16-e25 [FREE Full text] [doi: 10.1016/j.jinf.2020.04.021] [Medline: 32335169]
14. Huang L, Zhang X, Zhang X, Wei Z, Zhang L, Xu J, et al. Rapid asymptomatic transmission of COVID-19 during the incubation period demonstrating strong infectivity in a cluster of youngsters aged 16-23 years outside Wuhan and characteristics of young patients with COVID-19: A prospective contact-tracing study. J Infect 2020 Jun;80(6):e1-e13 [FREE Full text] [doi: 10.1016/j.jinf.2020.03.006] [Medline: 32283156]
15. Gao Z, Xu Y, Sun C, Wang X, Guo Y, Qiu S, et al. A systematic review of asymptomatic infections with COVID-19. J Microbiol Immunol Infect 2021 Feb;54(1):12-16 [FREE Full text] [doi: 10.1016/j.jmii.2020.05.001] [Medline: 32425996]
16. Timeline: WHO's COVID-19 response. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline#! [accessed 2021-08-10]
17. Quaresima V, Naldini MM, Cirillo DM. The prospects for the SARS-CoV-2 pandemic in Africa. EMBO Mol Med 2020 Jun 08;12(6):e12488 [FREE Full text] [doi: 10.15252/emmm.202012488] [Medline: 32301279]
18. Shoer S, Karady T, Keshet A, Shilo S, Rossman H, Gavrieli A, et al. A Prediction Model to Prioritize Individuals for a SARS-CoV-2 Test Built from National Symptom Surveys. Med (N Y) 2021 Feb 12;2(2):196-208.e4 [FREE Full text] [doi: 10.1016/j.medj.2020.10.002] [Medline: 33073258]

19. Barda N, Riesel D, Akriv A, Levy J, Finkel U, Yona G, et al. Developing a COVID-19 mortality risk prediction model when individual-level data are not available. Nat Commun 2020 Sep 07;11(1):4439 [FREE Full text] [doi: 10.1038/s41467-020-18297-9] [Medline: 32895375]
20. Hua J, Shaw R. Corona Virus (COVID-19) "Infodemic" and Emerging Issues through a Data Lens: The Case of China. Int J Environ Res Public Health 2020 Mar 30;17(7):2309 [FREE Full text] [doi: 10.3390/ijerph17072309] [Medline: 32235433]
21. Zarocostas J. How to fight an infodemic. Lancet 2020 Feb 29;395(10225):676 [FREE Full text] [doi: 10.1016/S0140-6736(20)30461-X] [Medline: 32113495]
22. Sun K, Chen J, Viboud C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. Lancet Digit Health 2020 Apr;2(4):e201-e208 [FREE Full text] [doi: 10.1016/S2589-7500(20)30026-1] [Medline: 32309796]
23. Qin L, Sun Q, Wang Y, Wu K, Chen M, Shia B, et al. Prediction of Number of Cases of 2019 Novel Coronavirus (COVID-19) Using Social Media Search Index. Int J Environ Res Public Health 2020 Mar 31;17(7):2365 [FREE Full text] [doi: 10.3390/ijerph17072365] [Medline: 32244425]
24. Robinson PC, Yazdany J. The COVID-19 Global Rheumatology Alliance: collecting data in a pandemic. Nat Rev Rheumatol 2020 Jun;16(6):293-294 [FREE Full text] [doi: 10.1038/s41584-020-0418-0] [Medline: 32242121]
25. Román GC, Reis J, Spencer PS, Buguet A, Öztürk S, Wasay M, World Federation of Neurology Environmental Neurology Specialty Group. COVID-19 international neurological registries. Lancet Neurol 2020 Jun;19(6):484-485 [FREE Full text] [doi: 10.1016/S1474-4422(20)30148-4] [Medline: 32470417]
26. Freeman EE, McMahon DE, Fitzgerald ME, Fox LP, Rosenbach M, Takeshita J, et al. The American Academy of Dermatology COVID-19 registry: Crowdsourcing dermatology in the age of COVID-19. J Am Acad Dermatol 2020 Aug;83(2):509-510 [FREE Full text] [doi: 10.1016/j.jaad.2020.04.045] [Medline: 32305438]
27. Luo H, Lie Y, Prinzen FW. Surveillance of COVID-19 in the General Population Using an Online Questionnaire: Report From 18,161 Respondents in China. JMIR Public Health Surveill 2020 Apr 27;6(2):e18576 [FREE Full text] [doi: 10.2196/18576] [Medline: 32319956]
28. Campus G, Diaz-Betancourt M, Cagetti M, Carvalho J, Carvalho T, Cortés-Martinicorena JF, et al. Study Protocol for an Online Questionnaire Survey on Symptoms/Signs, Protective Measures, Level of Awareness and Perception Regarding COVID-19 Outbreak among Dentists. A Global Survey. Int J Environ Res Public Health 2020 Aug 03;17(15):5598 [FREE Full text] [doi: 10.3390/ijerph17155598] [Medline: 32756475]
29. McCall B. COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread. Lancet Digit Health 2020 Apr;2(4):e166-e167 [FREE Full text] [doi: 10.1016/S2589-7500(20)30054-6] [Medline: 32289116]
30. Keesara S, Jonas A, Schulman K. Covid-19 and Health Care's Digital Revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
31. Robbins T, Hudson S, Ray P, Sankar S, Patel K, Randeva H, et al. COVID-19: A new digital dawn? Digit Health 2020;6:2055207620920083 [FREE Full text] [doi: 10.1177/2055207620920083] [Medline: 32313668]
32. Heymann DL, Shindo N, WHO ScientificTechnical Advisory Group for Infectious Hazards. COVID-19: what is next for public health? Lancet 2020 Feb 22;395(10224):542-545 [FREE Full text] [doi: 10.1016/S0140-6736(20)30374-3] [Medline: 32061313]
33. Eysenbach G. How to Fight an Infodemic: The Four Pillars of Infodemic Management. J Med Internet Res 2020 Jun 29;22(6):e21820 [FREE Full text] [doi: 10.2196/21820] [Medline: 32589589]
34. Fomsgaard A, Rosenstierne M. An alternative workflow for molecular detection of SARS-CoV-2 - escape from the NA extraction kit-shortage, Copenhagen, Denmark, March 2020. Euro Surveill 2020 Apr;25(14):2000398 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.14.2000398] [Medline: 32290902]
35. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med 2016 Sep 29;375(13):1216-1219 [FREE Full text] [doi: 10.1056/NEJMp1606181] [Medline: 27682033]
36. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. J Big Data 2019 Jun 19;6(1):1-25. [doi: 10.1186/s40537-019-0217-0]
37. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. N Engl J Med 2019 Apr 04;380(14):1347-1358. [doi: 10.1056/nejmra1814259]
38. Peiffer-Smadja N, Maatoug R, Lescure F, D'Ortenzio E, Pineau J, King J. Machine Learning for COVID-19 needs global collaboration and data-sharing. Nat Mach Intell 2020 May 22;2(6):293-294. [doi: 10.1038/s42256-020-0181-6]
39. Latif S, Usman M, Manzoor S, Iqbal W, Qadir J, Tyson G, et al. Leveraging Data Science to Combat COVID-19: A Comprehensive Review. IEEE Trans. Artif. Intell 2020 Aug;1(1):85-103. [doi: 10.1109/TAI.2020.3020521]
40. Alamo T, Reina D, Millán P. Data-driven methods to monitor, model, forecast and control covid-19 pandemic: Leveraging data science, epidemiology and control theory. arXiv. Preprint posted online June 1, 2020.
41. Callaghan S. COVID-19 Is a Data Science Issue. Patterns (N Y) 2020 May 08;1(2):100022 [FREE Full text] [doi: 10.1016/j.patter.2020.100022] [Medline: 32562484]

Abbreviations

CDL: IDF COVIDataLake
ID3C: IDF Data Center for the Campaign against Coronavirus
IDF: Israel Defense Forces
RT–PCR: reverse transcription polymerase chain reaction

Edited by C Basch; submitted 13.09.20; peer-reviewed by H Zhang, M Alvarez de Mon; comments to author 02.11.20; revised version received 14.12.20; accepted 10.04.21; published 10.09.21.

Please cite as:
Geva GA, Ketko I, Nitecki M, Simon S, Inbar B, Toledo I, Shapiro M, Vaturi B, Votta Y, Filler D, Yosef R, Shpitzer SA, Hir N, Peri Markovich M, Shapira S, Fink N, Glasberg E, Furer A
Data Empowerment of Decision-Makers in an Era of a Pandemic: Intersection of "Classic" and Artificial Intelligence in the Service of Medicine
J Med Internet Res 2021;23(9):e24295
URL: https://www.jmir.org/2021/9/e24295
doi:10.2196/24295
PMID:34313589

Original Paper

Efficacy of a Single-Session "Empowered Relief" Zoom-Delivered Group Intervention for Chronic Pain: Randomized Controlled Trial Conducted During the COVID-19 Pandemic

Maisa S Ziadni[1], MS, PhD; Lluvia Gonzalez-Castro[1], BS; Steven Anderson[1], PhD; Parthasarathy Krishnamurthy[2], PhD; Beth D Darnall[1], PhD

[1]Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University School of Medicine, Palo Alto, CA, United States

[2]C.T. Bauer College of Business, University of Houston, Houston, TX, United States

Corresponding Author:
Maisa S Ziadni, MS, PhD
Department of Anesthesiology, Perioperative and Pain Medicine
Stanford University School of Medicine
1070 Arastradero Road
Palo Alto, CA, 94304
United States
Phone: 1 6507365494
Email: mziadni@stanford.edu

Abstract

Background: Cognitive behavioral therapy–pain is an evidence-based treatment for chronic pain that can have significant patient burden, including health care cost, travel, multiple sessions, and lack of access in remote areas.

Objective: The study aims to pilot test the efficacy of a single-session videoconference-delivered *empowered relief* (ER) intervention compared to waitlist control (WLC) conditions among individuals with chronic pain. We hypothesized that ER would be superior to WLC in reducing pain catastrophizing, pain intensity, and other pain-related outcomes at 1-3 months posttreatment.

Methods: We conducted a randomized controlled trial involving a web-based sample of adults (N=104) aged 18-80 years with self-reported chronic pain. Participants were randomized (1:1) to 1 of 2 unblinded study groups: ER (50/104, 48.1%) and WLC (54/104, 51.9%). Participants allocated to ER completed a Zoom-delivered class, and all participants completed follow-up surveys at 2 weeks and 1, 2, and 3 months posttreatment. All the study procedures were performed remotely and electronically. The primary outcome was pain catastrophizing 1-month posttreatment, with pain intensity, pain bothersomeness, and sleep disruption as secondary outcomes. We also report a more rigorous test of the durability of treatment effects at 3 months posttreatment. Data were collected from September 2020 to February 2021 and analyzed using intention-to-treat analysis. The analytic data set included participants (18/101, 17.8% clinic patients; 83/101, 82.1% community) who completed at least one study survey: ER (50/101, 49.5%) and WLC (51/104, 49%).

Results: Participants (N=101) were 69.3% (70/101) female, with a mean age of 49.76 years (SD 13.90; range 24-78); 32.7% (33/101) had an undergraduate degree and self-reported chronic pain for 3 months. Participants reported high engagement (47/50, 94%), high satisfaction with ER (mean 8.26, SD 1.57; range 0-10), and high satisfaction with the Zoom platform (46/50, 92%). For the between-groups factor, ER was superior to WLC for all primary and secondary outcomes at 3 months posttreatment (highest $P<.001$), and between-groups Cohen *d* effect sizes ranged from 0.45 to 0.79, indicating that the superiority was of moderate to substantial clinical importance. At 3 months, clinically meaningful pain catastrophizing scale (PCS) reductions were found for ER but not for WLC (ER: PCS −8.72, 42.25% reduction; WLC: PCS −2.25, 11.13% reduction). ER resulted in significant improvements in pain intensity, sleep disturbance, and clinical improvements in pain bothersomeness.

Conclusions: Zoom-delivered ER had high participant satisfaction and very high engagement. Among adults with chronic pain, this single-session, Zoom-delivered, skills-based pain class resulted in clinically significant improvement across a range of pain-related outcomes that was sustained at 3 months. Web-based delivery of ER could allow greater accessibility of home-based pain treatment and could address the inconveniences and barriers faced by patients when attempting to receive in-person care.

Trial Registration: ClinicalTrials.gov NCT04546685; https://clinicaltrials.gov/ct2/show/NCT04546685

(J Med Internet Res 2021;23(9):e29672) doi:10.2196/29672

KEYWORDS

single-session; empowered relief; Zoom-delivered; pain catastrophizing; pain intensity; randomized-controlled trial; chronic pain

Introduction

Background

Chronic pain is a significant public health concern. It is one of the most common reasons for seeking medical care [1], with considerable societal and financial burden [2,3] in addition to human suffering. Chronic pain treatment is challenging due to limitations of existing pharmacological approaches, particularly intolerability of many medicines [4], health risks [5-9], and lack of sustainable availability. In contrast, behavioral medicine approaches are low risk and have been shown to have low-to-moderate efficacy for a range of chronic pain conditions [5,7,10-12], with cognitive behavioral therapy (CBT)–pain as the established evidence-based treatment for chronic pain [8,9]. However, several barriers prevent broad access to CBT-pain, including out-of-pocket costs, burdensome travel, lack of availability in remote or rural areas, and a lack of adequately trained clinicians [6,13]. The length of treatment (ie, eight sessions or more for CBT-pain) is associated with increased dropout [14], which may perpetuate the overuse of medical services and health disparities. The COVID-19 pandemic has further challenged the feasibility of multisession in-person treatment options [15], and this creates the need for accessible solutions that are low-cost, low-burden, and remove the need for face-to-face meetings [16].

Brief web-based and remotely delivered psychological interventions have been demonstrated to accomplish this goal and have been found to be as effective as face-to-face therapy, while decreasing health care use and barriers to treatment [17]. Effective web-based treatments now exist for various psychological disorders, including depression, panic disorder, and social phobia [18]. Web-based CBT has been effective for a number of chronic health conditions, including irritable bowel syndrome, tinnitus, and headache [19,20]. In a systematic review of web-based interventions for chronic back pain, Garg et al [21] reported that web-based interventions, including CBT, effectively reduced pain catastrophizing scale (PCS) scores. Effect sizes for web-based interventions for chronic pain have been estimated to range from small to moderate (Cohen *d* range 0.04-1.23) [18]. In addition to CBT, existing web-based interventions for patients with pain include compassionate mind training [22], social media–based web-based community intervention [23], pain self-management [24], and hypnosis [25]. Despite the promise of web-based interventions for pain, multisession treatments typically range from 20 days to 12 weeks in duration, which can be costly, burdensome, and limit access to adequate care [26].

Single-session interventions (SSIs), which have been defined as the intentional delivery of a specific, structured program involving a single visit with a clinic, provider, or program [27], have the potential to reduce the burden of traditional multisession treatment approaches and expand access to behavioral medicine. SSIs have been found to be effective for a variety of psychological and health conditions, including severe mental illness [28], anxiety and youth conduct disorder [27], acute insomnia [29], heavy alcohol consumption in college students [30], trauma and adversity [31,32], postsurgical pain [33], and chronic pain [11,12,34-36]. SSIs that have been further optimized via web-based delivery have demonstrated feasibility and efficacy similar to face-to-face interventions [11,12,37] and include interventions for multiple sclerosis pain [38], disordered gambling [39], and adolescent mental health [40-42]. In addition to reducing treatment burdens, web-based SSIs have the vital advantage of ease of scalability, as they can be completed by patients in any location [37]. Effective web and digital-based SSIs would eliminate or reduce many of the existing barriers to treatment, such as cost, lack of trained therapists, and insurance limits on the length of treatment. However, despite the promise of SSIs, rigorously designed and randomized controlled studies are needed to establish their efficacy [43].

Objectives

Building on the limited literature on web-based SSIs for chronic pain, we aim to evaluate a Zoom-delivered version of a previously developed and efficacious, skills-based, behavioral medicine SSI *empowered relief* (ER) [11,36]. Previously, our group conducted a three-arm randomized controlled trial conducted in 263 adults with chronic low back pain comparing in-person ER to a health education class or 8-week CBT [12]. However, web-based delivery of ER has not been tested, particularly in mixed-etiology chronic pain, and the COVID-19 context underscores the importance of providing effective home-based chronic pain care. Extending this work to the digital platform, in this study, we conducted a parallel-group, randomized (1:1) comparative efficacy trial to assess the impact of a single-session videoconference-delivered ER group pain relief skills class versus waitlist control (WLC). We hypothesize that at 3 months posttreatment, (1) ER would be superior to WLC for reductions in pain catastrophizing, an index of pain coping; (2) ER would be superior to WLC for reductions in pain intensity, bothersomeness, and sleep disturbance; and (3) ER would be superior to WLC in reducing anxiety, depression, and physical function.

Methods

Overview and Setting

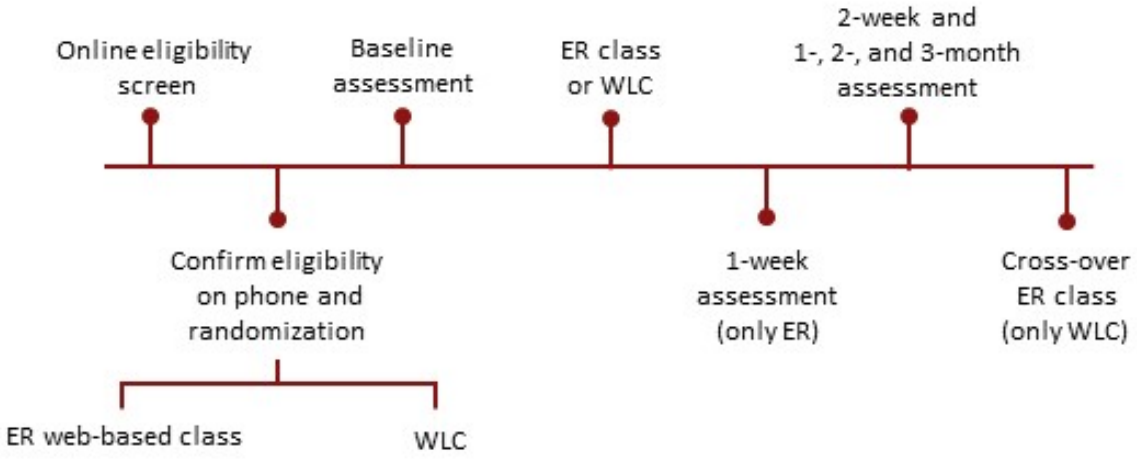
This study was a parallel-group, randomized, clinical trial comparing ER, a 2-hour single-session videoconference-delivered class, to a WLC. Enrolled participants (N=104) aged 18-80 years with mixed chronic pain etiology were randomized (1:1) to one of the two study arms. Participants completed outcome assessments at pretreatment, week 2, and 1, 2, and 3 months posttreatment. The primary outcome was pain catastrophizing levels at the 1-month follow-up [35], with pain intensity, pain, bothersomeness, and sleep disturbance as secondary outcomes. The tertiary outcomes

included PROMIS (Patient-Reported Outcomes Measurement Information System) measures of anxiety, depression, and physical function at the 1-month follow-up time-point. We also report a more rigorous test of the durability of treatment effects at 3 months posttreatment as highlighted in the *Results* section. A crossover intervention class was offered to participants who were initially randomized to the WLC and was conducted in February 2021. Of 48 participants who were offered the treatment, 20 (42%) attended the crossover class at no cost, and no data were collected. This study was conducted between August 2020 and February 2021.

Study Sample, Recruitment, and Participant Compensation

Participants were recruited remotely through targeted emails to lists of patients who agreed to be contacted for research purposes through two databases: (1) the Stanford's Collaborative Health Outcomes Information Registry comprising patients who have received care at the Stanford Pain Management Center (a tertiary referral pain clinic in Redwood City, California) and (2) the Stanford Systems Neuroscience and Pain Lab database of individuals with chronic pain. Interested individuals were directed to a web-based screening form, and those who met the initial eligibility criteria were contacted by research staff by phone to confirm their eligibility. Approximately 18.3% (19/104) of the study sample were clinic patients, and 81.7% (85/104) were from the larger community. Eligible and interested participants were invited to participate in a pain-skills treatment class compared with a WLC and completed an electronic informed consent form (see Multimedia Appendix 1 for the study consent form). After informed consent was obtained, study participants were randomized one-to-one using a REDCap (Research Electronic Data Capture; Vanderbilt University) [44] cloud random number generator and allocated to the study group. All study procedures were completed remotely, and no in-person visits were required. Figure 1 displays the participants' study activities.

Figure 1. Participant activities. ER: empowered relief; WLC: waitlist control.



All the study procedures were performed remotely and electronically. Enrolled participants completed a baseline survey, which was conducted 4 days before the scheduled class, followed by posttreatment surveys at 2 weeks, and 1, 2, and 3 months. The intervention was conducted on Zoom [45], and details of the Zoom session were emailed to participants before the scheduled class time. Overall, in the ER arm, two cohorts were recruited, each consisting of 24-26 participants per class. The class sizes were consistent with studies on web-based interventions [46-51]. Participants received US $10 in the form of an Amazon electronic gift code for each survey they completed and could receive up to US $50 following completion of all study procedures. The research procedures were approved by the Stanford institutional review board.

Inclusion and Exclusion Criteria

All study recruitment efforts directed interested individuals with chronic pain to the web-based screening form to complete an automated eligibility form that screened for initial inclusion and exclusion criteria. Participants who met initial eligibility were then contacted by the research staff to confirm study eligibility according to the inclusion and exclusion criteria provided in Textbox 1. Internet and computer literacy were the implicit de facto eligibility criteria. If interested participants indicated that they had never used the Zoom platform, the research coordinator contacted them by phone to schedule an individual meeting; their skill to participate in the web-based class would increase because they were more familiar with the Zoom platform and navigating the tool in a supervised setting. Of the 9 participants who indicated that they had never used Zoom, none requested to schedule an individual meeting.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Pain >3 months more than half the time
- English fluency
- Ability to attend a one-time web-based class (if assigned) and complete web-based surveys
- Females and males, aged ≥18 years

Exclusion criteria

- Ongoing legal action or disability claim
- Previous participation in the empowered relief class
- Cognitive impairment, non–English-speaking, or psychological factors that would preclude comprehension of material or full participation in the study

Randomization and Blinding

Participants were randomized 1:1 (no blocking applied) and allocated to one of two study arms: (1) ER or (2) WLC. REDCap was used to apply an automatic and blinded randomization program and ensure equal allocation to both groups. Participants were not blinded to the study arm, given the nature of WLC. Statisticians performed blinded analysis of data sets that were randomly labeled as group 1 and group 2, with statistician unblinding occurring only after posttreatment month 3 data were analyzed. The study coordinator (LGC) was unblinded to the individual study arm assignments and was not involved in any data analyses. The study protocol was reviewed and approved [52].

Study Groups

Single-Session Videoconference-Delivered Skills-Based Pain Class (ER)

Overview

ER was developed in 2013, with pilot data [36] showing a reduction in pain catastrophizing 1-month posttreatment, despite comorbid emotional disorders. Notably, a National Institute of Health-funded, three-arm, randomized controlled trial [11,12] conducted in 263 individuals with chronic low back pain demonstrated that 3-month posttreatment, the ER class was equally effective to an 8-week CBT-pain intervention and was superior to a health education control class in reducing pain catastrophizing and pain intensity. ER *compresses* key elements and skills from CBT-pain and mindfulness into a single-session 2-hour class [11,12,34,36]. The class is didactic and delivered with a standardized instructor slide deck and manual. Participants learn self-management skills to decrease physiological hyperarousal, which includes diaphragmatic breathing, cognitive reframing, and self-soothing strategies. During the class, participants self-tailor the skills content by completing a personalized plan for ER. They also received an MP3 audio file with a 20-minute version of the relaxed breathing exercise. The participants can download the audio file from a secure web-based storage platform onto their various devices (mobile or tablet) for convenience. The participants were encouraged to practice the learned skills daily.

Class Platform

The Zoom platform was used to deliver the ER classes, and instructors shared their PowerPoint presentation slides throughout the class. Class participants were encouraged to ask questions at any time through the Zoom chat box and unmute themselves to participate. Zoom is password-protected and hosted within the firewalled Stanford University School of Medicine and Stanford HealthCare systems.

Instructors

The class instructors were doctoral-level clinical psychologists certified by Stanford University [53] to deliver the intervention.

Training and Monitoring of Instructors

The instructors were familiar with the study protocol, were certified in ER delivery, and had extensive experience delivering the class. ER was developed to ensure fidelity (eg, standardized slide deck and instructor manual; instructor certification process), and cohort effects were minimal due to the single session and didactic nature of the intervention, with minimal participant interaction.

Waitlist Control

The study used a WLC group because the ER class web-based delivery is a new approach that has not been previously tested [54,55]. During the screening phone call, participants randomized to the WLC were advised to continue any ongoing clinical care. Although the WLC group did not receive any study intervention, participants were informed that they would be offered the ER treatment class after study completion at 3 months.

Data Collection and Variable Measurement

All surveys were collected through REDCap [44], a web-based electronic data capture platform, which is a secure (password protected), HIPAA (Health Insurance Portability and Accountability Act)–compliant platform, and hosted by the Stanford University School of Medicine.

Data collection consisted of electronically collected participant-reported measures across the four phases of the study: screening, pretreatment, treatment, and posttreatment. For both groups, posttreatment survey collection was conducted at 2 weeks and 1-3 months posttreatment. The ER group

completed an additional brief survey 1-week posttreatment to assess participant satisfaction with the intervention and Zoom platform. Table 1 shows the data collection time-points for both the groups. Baseline surveys included demographic information, pain intensity, pain bothersomeness, pain catastrophizing, and PROMIS measures. Posttreatment surveys included the same measures, excluding the demographics.

Table 1. Timeline of variable assessment.

Variables	Pretreatment	Posttreatment follow-up		
	Baseline	1 week	2 weeks	1, 2, and 3 months
Demographics and pain condition	✓[a]			
PCS[b]	✓		✓	✓
Pain bothersomeness	✓		✓	✓
PROMIS[c]				
Depression	✓		✓	✓
Anxiety	✓		✓	✓
Pain intensity	✓		✓	✓
Physical function	✓		✓	✓
Pain interference	✓		✓	✓
Fatigue	✓		✓	✓
Social isolation	✓		✓	✓
Anger	✓		✓	✓
Sleep disturbance	✓		✓	✓
Satisfaction with treatment		✓		

[a]Variable assessed.

[b]PCS: pain catastrophizing scale.

[c]PROMIS: Patient-Reported Outcomes Measurement Information System.

Study Variables and Measures

Pain Catastrophizing

The 13-item PCS [56] measures negative thoughts and emotional responses to pain. PCS is a 1D measurement with three subscales: helplessness, magnification, and rumination. It is scored by summing all items and generates a total score, with higher scores indicating greater catastrophizing. The PCS is a reliable, validated, and psychometrically trusted instrument [57].

Pain Bothersomeness

Participants rated their pain bothersomeness during the previous 7 days on a 0-10 numeric rating scale anchored by 0 (not at all bothersome) and 10 (extremely bothersome) that is commonly used in chronic low back pain research [58], “How bothersome has your pain been during the past week?”

National Institutes of Health PROMIS Measures

The National Institutes of Health PROMIS short-form measures have been applied in pain research [59-67], and selected domains were identified by the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials [68,69] as core outcomes. Respondents reference the previous 7 days to rate items for pain intensity (version 3a), sleep disturbance (version 6a), physical function (version 6b), depression (version 6a), anxiety (version 6a), and social isolation (4a). Higher scores on PROMIS measures indicate greater symptom severity, except for physical function, wherein higher scores reflect better function. The web-based PROMIS assessment center [70] software [71] was used to calculate the short-form T scores using Item Response Theory scoring algorithms that apply the Bayesian expected A posteriori method [72]. Depression, anxiety, and physical function are major correlates of chronic pain [73-76], and as such, changes in these variables were examined as a function of treatment. Social isolation was included because of the COVID-19 pandemic. We did not expect it to change, per se, but given the dynamic nature of COVID-19 and lockdowns, we measured this variable throughout the study.

Sample Size and Statistical Analysis

The planned enrollment accounted for 4% attrition in each study arm. The enrolled sample size ensured significant statistical power (80%) to detect medium-to-large treatment effects between the two treatment groups. Of the 104 enrolled participants (ER: 50/104, 48.1%; WLC: 54/104, 51.9%), intent-to-treat analyses were used to analyze the data of 101 (97.1%) participants (ER: 50/101, 49.5%; WLC: 51/101, 50.5%) who met the study criteria and completed at least one survey.

Baseline scores on continuous variables were summarized using mean (SD), and categorical variable scores were summarized by count and proportions. The differences between conditions at baseline for continuous and categorical variables for the treatment groups were assessed using one-way analysis of

variance and Fisher exact test, respectively. For primary analyses, we used a between-within mixed design to assess the effect of treatment, time, and treatment × time, where time represented the within-subjects factor and treatment was specified as a between-subjects factor. Within-subjects' dependency was modeled by specifying participants as a random effect with a heterogeneous compound symmetry covariance structure. Similar analyses were used to examine the treatment effects for all secondary and tertiary outcomes at each posttreatment month. Intention-to-treat analysis was used to investigate the causal effects of the treatment. All statistical significances were applied at a two-sided level of 0.05, with Benjamini-Hochberg adjustment for multiple comparisons. Cohen *d* effect sizes for the between-subjects' factor were calculated from the between-subjects' *F* statistic adjusting for sample size. Responder analyses were conducted by calculating the proportion of participants with 15% (minimal), 30% (moderate), and 50% (substantial) improvement from baseline for pain-related outcomes. All data processing and statistical analyses were performed using the SAS version 9.4 (SAS Institute, Inc).

Results

Study Participants

Figure 2 shows the CONSORT (Consolidated Standards of Reporting Trials) diagram for the study. In total, 196 individuals were assessed for initial eligibility, and 92 were excluded because they did not meet the study criteria. In total, 104 participants were enrolled and randomized, and 101 (97.1%) participants completed the baseline surveys (ER: 50/101, 49.5%; WLC: 51/101, 50.5%). A total of 49 participants received the ER intervention, of whom 2 (4%) were lost to follow-up due to feeling unwell or being too busy to complete follow-up surveys. A final sample of 94 participants provided complete follow-up data. Nearly 94% (47/50) of participants in the ER group and 86% (44/51) of participants in the WLC group completed their 3-month follow-up surveys (end of study).

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow diagram. ER: empowered relief; ITT: intention-to-treat; WLC: waitlist control.

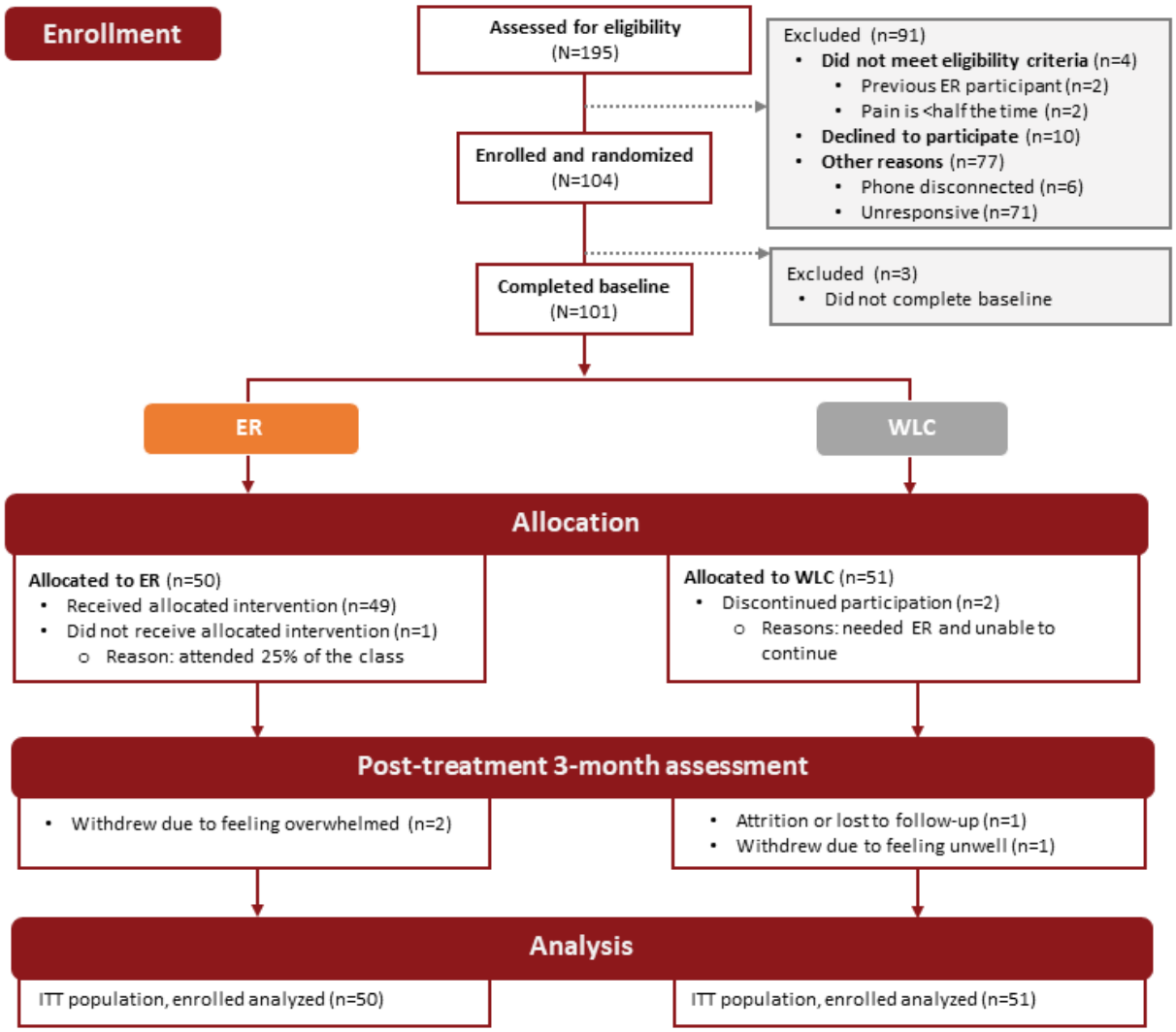


Table 2 shows the baseline demographic characteristics of the study arm. The sample included 101 participants recruited primarily from northern California. The sample was predominantly female (70/101, 69.3%), White (75/101, 74.3%), and with at least some college education (66/101, 65.3%), and a mean age of 49.76 years (SD 13.90; range 24.00-78.00). No significant between-group differences were observed for any demographic variables, demonstrating that randomization was effective.

Table 2. Baseline demographic characteristics by treatment group.

Demographics	Values		*P* value
	Treatment group with empowered relief (n=50)	WLC[a] (n=51)	
Gender, n (%)			.15[b]
Male	12 (24)	19 (37)	
Female	38 (76)	32 (63)	
Age (years)			.34[c]
Mean (SD; range)	48.6 (14.1; 26.0-78.0)	50.9 (13.7; 24.0-74.0)	
Median (IQR)	48.5 (35.0-59.0)	51.0 (40.0-67.0)	
Race, n (%)			.90[b]
Asian	6 (12)	7 (14)	
White	37 (74)	38 (75)	
Black or African American	1 (2)	0 (0)	
More than one race	3 (6)	3 (6)	
Other or unknown	3 (6)	3 (6)	
Education, n (%)			.20[b]
High school graduate	1 (2)	1 (2)	
Some college	7 (14)	7 (14)	
Associate	8 (16)	11 (22)	
Undergraduate	14 (28)	19 (37)	
Master's degree	13 (26)	10 (20)	
Professional degree	1 (2)	3 (6)	
Doctoral degree	6 (12)	0 (0)	
Employment, n (%)			.72[b]
Part-time	10 (20)	10 (20)	
Full-time	15 (30)	18 (35)	
Full-time homemaker	0 (0)	4 (8)	
Temporarily laid off	3 (6)	1 (2)	
Unemployed	3 (6)	2 (4)	
Looking for work, unemployed	3 (6)	2 (4)	
Disabled or not working due to pain	5 (10)	4 (8)	
Retired	5 (10)	6 (12)	
Student, currently employed	1 (2)	1 (2)	
Student, not currently employed	1 (2)	0 (0)	
Other	4 (8)	3 (6)	
Income (US $), n (%)			.64[b]
<10,000	1 (2)	2 (4)	
>25,000	4 (8)	7 (14)	
25,000-44,999	6 (12)	3 (6)	
45,000-64,999	2 (4)	3 (6)	
65,000-84,999	2 (4)	4 (8)	
85,000-104,999	7 (14)	3 (6)	
105,000-124,999	7 (14)	10 (20)	

Demographics	Values		*P* value
	Treatment group with empowered relief (n=50)	WLC[a] (n=51)	
>125,000	15 (30)	14 (27)	
Missing	6 (12)	5 (10)	
Relationship, n (%)			.82[b]
Never married	10 (20)	9 (18)	
Married	26 (52)	26 (51)	
Divorced	6 (12)	7 (14)	
Separated	1 (2)	2 (4)	
Widowed	0 (0)	2 (4)	
Partnered and living together	5 (10)	4 (8)	
In a relationship but not living together	2 (4)	1 (2)	

[a]WLC: waitlist control.

[b]Chi-square *P* value.

[c]Kruskal-Wallis *P* value.

Table 3 presents the baseline pain and clinical characteristics of the sample by the study arm. The mean pain intensity T score was 62.2 (mean 50, SD 10). No significant differences were observed between the two study groups for any of the variables assessed, except for sleep disturbance (*P*=.04), which was controlled for in the analysis.

Table 3. Baseline clinical variables by treatment group.

Variables	Values		*P* value[b]
	Treatment group with empowered relief (n=50)	WLC[a] (n=51)	
Pain catastrophizing			.86
Mean (SD; range)	20.6 (10.1; 2.0-47.0)	20.2 (10.7; 2.0-43.0)	
Median (IQR)	19.0 (13.0-28.0)	19.0 (12.0-29.0)	
Average pain intensity			.60
Mean (SD; range)	62.2 (5.7; 48.0-81.9)	62.5 (7.7; 36.3-81.9)	
Median (IQR)	62.6 (58.9-64.1)	62.6 (60.2-67.8)	
Pain bothersomeness			.28
Mean (SD; range)	5.8 (2.0; 0.0-10.0)	6.3 (2.0; 2.0-10.0)	
Median (IQR)	6.0 (4.0-7.0)	6.0 (5.0-7.0)	
PROMIS[c]			
Depression			.95
Mean (SD; range)	57.3 (9.6; 38.4-73.0)	57.8 (9.3; 38.4-80.2)	
Median (IQR)	58.2 (50.6-65.4)	56.4 (52.3-64.7)	
Anxiety			.36
Mean (SD; range)	59.6 (8.3; 39.1-82.4)	60.4 (8.9; 39.1-79.5)	
Median (IQR)	60.7 (52.9-64.5)	61.8 (54.4-66.1)	
Physical function			.99
Mean (SD; range)	40.8 (6.3; 28.7-59.0)	41.1 (7.2; 24.8-59.0)	
Median (IQR)	40.4 (36.0-45.1)	40.5 (36.7-44.1)	
Interference			.82
Mean (SD; range)	59.7 (6.6; 41.1-76.2)	60.3 (5.4; 51.0-76.2)	
Median (IQR)	59.5 (55.7-64.8)	60.0 (56.2-63.8)	
Fatigue			.79
Mean (SD; range)	61.6 (8.4; 33.7-75.8)	62.4 (9.2; 33.7-75.8)	
Median (IQR)	62.8 (57.1-66.8)	62.7 (57.1-68.8)	
Social isolation			.23
Mean (SD; range)	53.5 (11.0; 34.8-74.2)	52.3 (9.4; 34.8-74.2)	
Median (IQR)	56.2 (47.9-62.1)	51.7 (47.9-56.2)	
Anger			.23
Mean (SD; range)	54.6 (9.1; 32.9-82.9)	55.5 (12.0; 32.9-77.1)	
Median (IQR)	54.7 (47.9-60.7)	57.8 (49.8-63.8)	
Sleep disturbance			.04
Mean (SD; range)	56.3 (9.0; 36.4-76.1)	59.5 (8.5; 36.6-76.1)	
Median (IQR)	54.5 (51.3-61.8)	60.0 (53.9-64.7)	

[a]WLC: waitlist control.

[b]Kruskal-Wallis *P* value.

[c]PROMIS: Patient-Reported Outcomes Measurement Information System.

Treatment Satisfaction (ER Group Only)

For satisfaction with ER treatment on a scale of 0 (completely dissatisfied) to 10 (completely satisfied), participants (n=47) reported high satisfaction with all items related to overall satisfaction with the class (mean 8.26, SD 1.57), likelihood of recommending the class (mean 8.77, SD 1.43), class relevance (mean 7.94, SD 2.03), the usefulness of the presented

information (mean 8.37, SD 1.79), and likelihood of using the skills and information learned (mean 8.66, SD 1.58).

Primary Outcomes

Table 4 reports the between-group comparisons at 2 weeks and 1-3 months posttreatment for all outcomes. Figures 3-12 compare group effects over time for pain-related outcomes. For all figures, the trend of the pain-related variables is displayed over time for participants in both groups. Values in the x-axis refer to the number of weeks in the study, where 0 represents the baseline (roughly 4 days preintervention) time-point. The color band represents the 95% CI for the mean outcome variable. Overlapping bands indicate nonsignificant treatment group differences (*P* value) of simple main effects within each time-point. The corresponding model effects for each outcome are displayed in Figures 3-12.

We observed a significant treatment effect (*P*=.005) on pain catastrophizing; on average, the ER group had lower PCS scores than the WLC group (Cohen *d*=0.50). Separately, we observed a time effect; the average PCS significantly decreased over time for both study groups (time effect; *P*<.001). Most importantly, the decrease was greater for ER versus WLC (group×time effect; *P*<.001). At 3-month posttreatment, clinically meaningful reductions in PCS were found for ER but not for WLC (ER: PCS −8.72, 42.25% reduction; WLC: PCS −2.25, 11.13% reduction). Figure 3 displays the average PCS at 3 months by study arm, and Table 4 displays the between-group comparisons for PCS at baseline and posttreatment months. ER was superior to WLC (difference −6.05, 95% CI −9.92 to −2.18; *P*=.002 at 3-month posttreatment.

The ER group effect size was 0.89, with combined results showing a large effect size and moderate clinical importance. As much as 62% (31/50) of ER and 24% (12/51) of WLC participants achieved a 30% or more reduction in PCS. For ER, 46% (23/50) achieved ≥50% in PCS reduction, whereas for WLC, 12% (6/51) reached that threshold.

Table 4. Outcome measures from baseline to 3 months by treatment group with between-group comparisons.

Outcome measure time-point	Empowered relief, mean (SE)	WLC,[a] mean (SE)	Between-group differences		Effect size[b] (Cohen *d*)
			Mean difference (SE; 95% CI)	*P* value	
Pain catastrophizing					
Baseline	20.64 (1.53)	20.22 (1.52)	0.42 (2.16; −3.81 to 4.66)	.84	0.04
2 weeks posttreatment	13.85 (1.37)	19.52 (1.38)	−5.67 (1.94; −9.49 to −1.85)	.004	0.57
1 month posttreatment	12.13 (1.37)	18.7 (1.35)	−6.57 (1.93; −10.36 to −2.78)	<.001	0.62
2 months posttreatment	11.16 (1.3)	18.5 (1.31)	−7.34 (1.85; −10.97 to −3.71)	<.001	0.76
3 months posttreatment	11.92 (1.39)	17.97 (1.39)	−6.05 (1.97; −9.92 to −2.18)	.002	0.64
Pain bothersomeness					
Baseline	5.78 (0.29)	6.3 (0.29)	−0.52 (0.40; −1.31 to 0.27)	.20	0.26
2 weeks posttreatment	5.13 (0.33)	5.66 (0.34)	−0.53 (0.47; −1.46 to 0.4)	.26	0.27
1 month posttreatment	4.81 (0.32)	5.92 (0.31)	−1.10 (0.44; −1.98 to −0.23)	.01	0.57
2 months posttreatment	4.71 (0.34)	5.84 (0.34)	−1.13 (0.48; −2.07 to −0.19)	.02	0.60
3 months posttreatment	4.62 (0.35)	5.92 (0.36)	−1.30 (0.50; −2.29 to −0.31)	.01	0.67
PROMIS[c]					
Pain intensity					
Baseline	62.24 (0.98)	62.53 (0.97)	−0.28 (1.39; −3.01 to 2.44)	.84	0.04
2 weeks posttreatment	59.81 (0.98)	61.61 (0.99)	−1.8 (1.40; −4.54 to 0.94)	.20	0.27
1 month posttreatment	57.96 (0.95)	62.06 (0.94)	−4.1 (1.34; −6.73 to −1.47)	.002	0.62
2 months posttreatment	59.17 (1)	62.55 (1.01)	−3.38 (1.42; −6.17 to −0.59)	.02	0.54
3 months posttreatment	57.26 (1.18)	61.07 (1.2)	−3.81 (1.68; −7.12 to −0.49)	.02	0.60
Physical function					
Baseline	40.76 (0.94)	41.12 (0.94)	−0.37 (1.33; −2.98 to 2.24)	.78	0.05
2 weeks posttreatment	41.65 (0.91)	40.42 (0.92)	1.23 (1.29; −1.31 to 3.77)	.34	0.24
1 month posttreatment	42.13 (0.93)	40.77 (0.93)	1.35 (1.31; −1.23 to 3.94)	.30	0.26
2 months posttreatment	42.18 (0.94)	39.97 (0.95)	2.22 (1.33; −0.41 to 4.84)	.01	0.41
3 months posttreatment	42.55 (1.15)	41.27 (1.16)	1.28 (1.63; −1.93 to 4.5)	.43	0.25
Pain interference					
Baseline	59.71 (0.86)	60.27 (0.86)	−0.56 (1.22; −2.96 to 1.84)	.65	0.09
2 weeks posttreatment	58.4 (0.92)	60.41 (0.93)	−2 (1.31; −4.59 to 0.58)	.13	0.32
1 month posttreatment	57.56 (0.89)	60.02 (0.88)	−2.46 (1.25; −4.92 to 0)	.05	0.43
2 months posttreatment	57.47 (0.93)	59.79 (0.94)	−2.32 (1.33; −4.94 to 0.29)	.08	0.43
3 months posttreatment	57.51 (1.12)	59.51 (1.14)	−1.99 (1.60; −5.14 to 1.15)	.21	0.33
Anxiety					
Baseline	59.57 (1.23)	60.41 (1.22)	−0.84 (1.73; −4.24 to 2.56)	.63	0.10
2 weeks posttreatment	58.12 (1.33)	60.4 (1.33)	−2.28 (1.88; −5.97 to 1.42)	.23	0.28
1 month posttreatment	56.06 (1.25)	59.79 (1.23)	−3.72 (1.75; −7.17 to −0.28)	.03	0.41
2 months posttreatment	56.18 (1.28)	59.98 (1.28)	−3.8 (1.81; −7.36 to −0.25)	.04	0.49
3 months posttreatment	54.85 (1.28)	59.99 (1.28)	−5.14 (1.81; −8.7 to −1.58)	.005	0.68
Depression					
Baseline	57.35 (1.4)	57.78 (1.38)	−0.43 (1.97; −4.3 to 3.44)	.83	0.05
2 weeks posttreatment	54.68 (1.25)	57.71 (1.25)	−3.03 (1.77; −6.51 to 0.45)	.09	0.32

Outcome measure time-point	Empowered relief, mean (SE)	WLC,[a] mean (SE)	Between-group differences		Effect size[b] (Cohen *d*)
			Mean difference (SE; 95% CI)	*P* value	
1 month posttreatment	53.58 (1.25)	57.24 (1.24)	−3.65 (1.76; −7.11 to −0.19)	.04	0.37
2 months posttreatment	53.5 (1.21)	57.55 (1.21)	−4.05 (1.71; −7.41 to −0.7)	.02	0.45
3 months posttreatment	53.89 (1.29)	57.26 (1.29)	−3.37 (1.82; −6.95 to 0.21)	.06	0.40
Social isolation					
Baseline	53.49 (1.47)	52.34 (1.47)	1.15 (2.08; −2.95 to 5.25)	.58	0.11
2 weeks posttreatment	51.59 (1.45)	52.75 (1.46)	−1.16 (2.06; −5.21 to 2.89)	.57	0.19
1 month posttreatment	51.06 (1.35)	52.48 (1.34)	−1.42 (1.90; −5.16 to 2.32)	.46	0.23
2 months posttreatment	50.12 (1.48)	52.49 (1.49)	−2.37 (2.10; −6.5 to 1.75)	.26	0.33
3 months posttreatment	49.5 (1.42)	52.37 (1.43)	−2.87 (2.02; −6.84 to 1.1)	.16	0.38
Anger					
Baseline	54.64 (1.57)	55.48 (1.57)	−0.84 (2.22; −5.21 to 3.53)	.70	0.08
2 weeks posttreatment	52.84 (1.45)	56.46 (1.47)	−3.62 (2.07; −7.68 to 0.45)	.08	0.34
1 month posttreatment	52.59 (1.28)	56.24 (1.27)	−3.65 (1.8; −7.18 to −0.11)	.04	0.35
2 months posttreatment	50.88 (1.37)	55.2 (1.39)	−4.31 (1.95; −8.15 to −0.48)	.03	0.44
3 months posttreatment	50.54 (1.33)	56 (1.35)	−5.47 (1.89; −9.19 to −1.75)	.004	0.59

[a]WLC: waitlist control.

[b]These are between-subjects effect size within each time frame.

[c]PROMIS: Patient-Reported Outcomes Measurement Information System.

Figure 3. Pain catastrophizing over time.

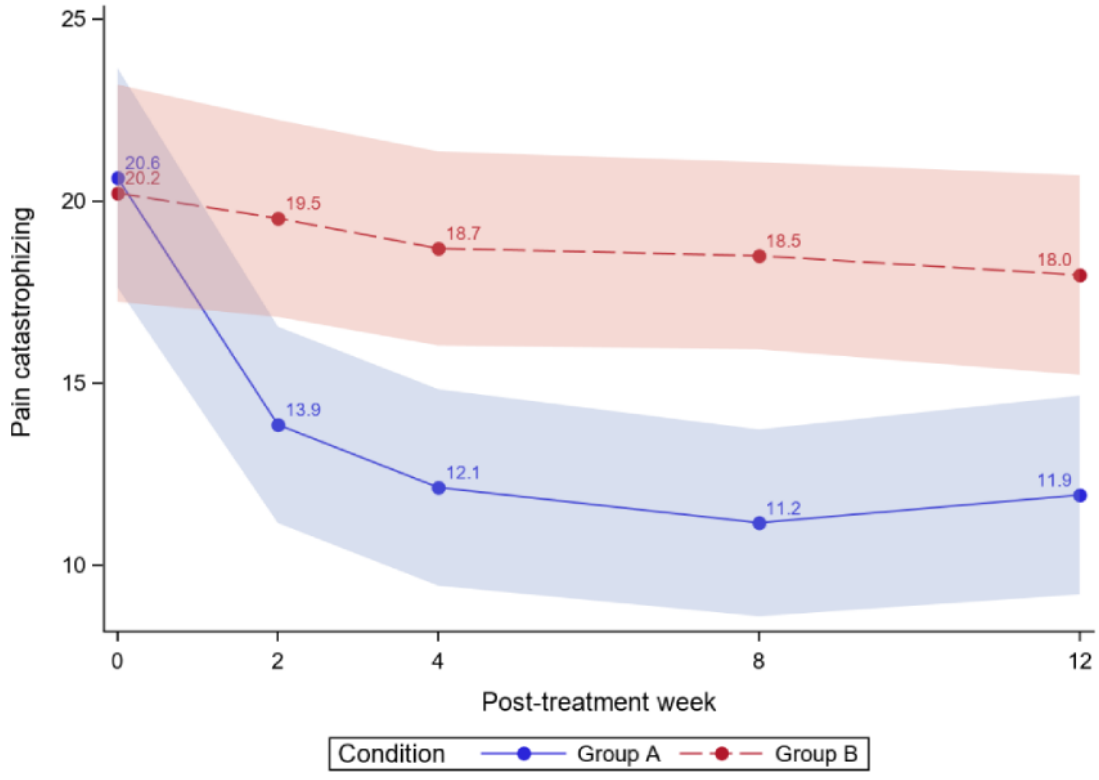


Figure 4. Pain intensity over time.

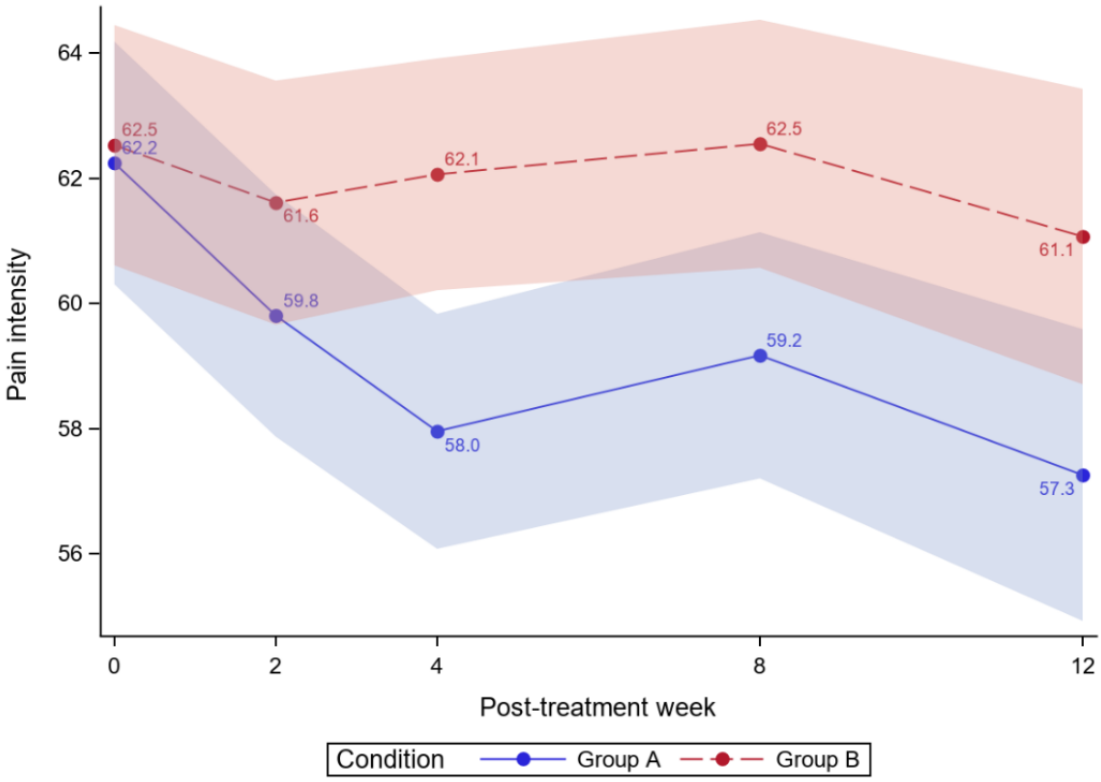


Figure 5. Pain bothersomeness over time.

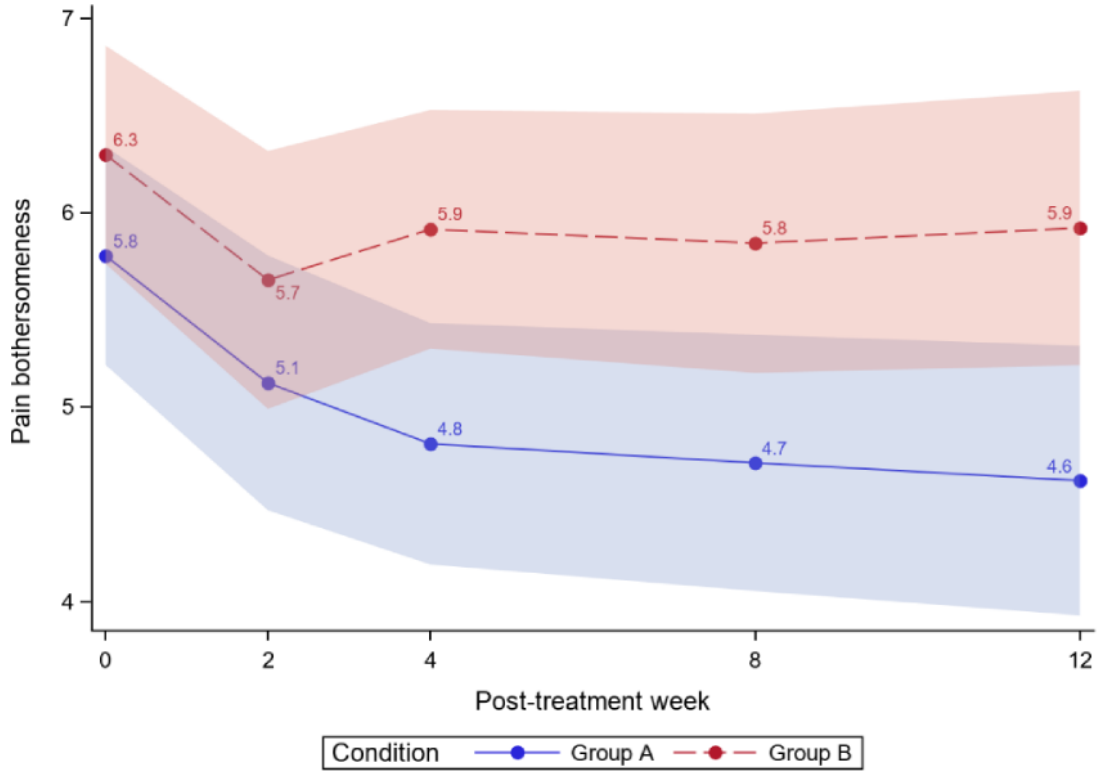


Figure 6. Sleep disturbance over time. PROMIS: Patient-Reported Outcomes Measurement Information System.

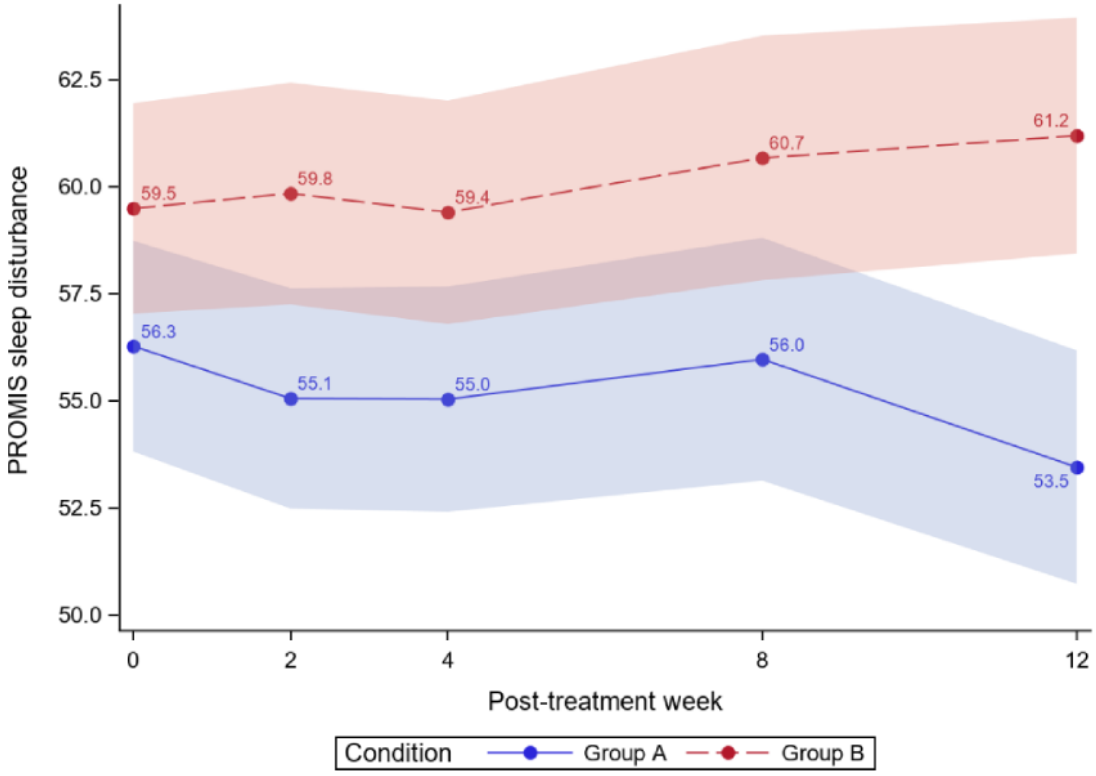


Figure 7. Anxiety over time. PROMIS: Patient-Reported Outcomes Measurement Information System.

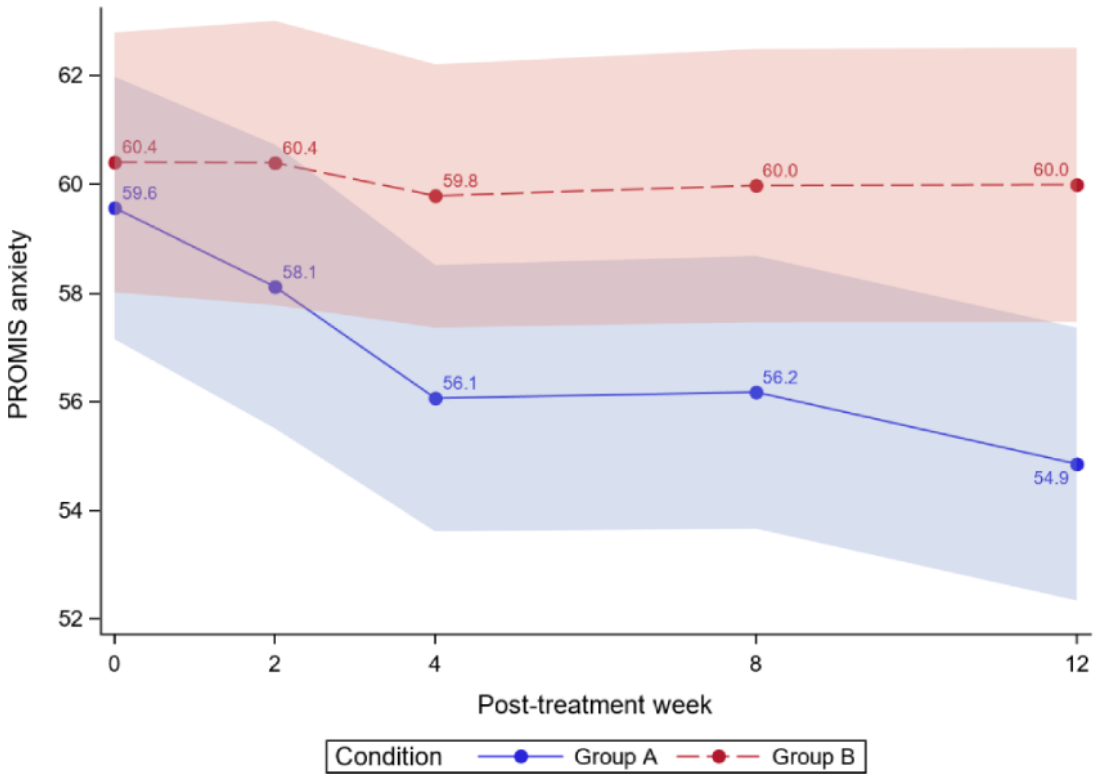


Figure 8. Depression over time. PROMIS: Patient-Reported Outcomes Measurement Information System.

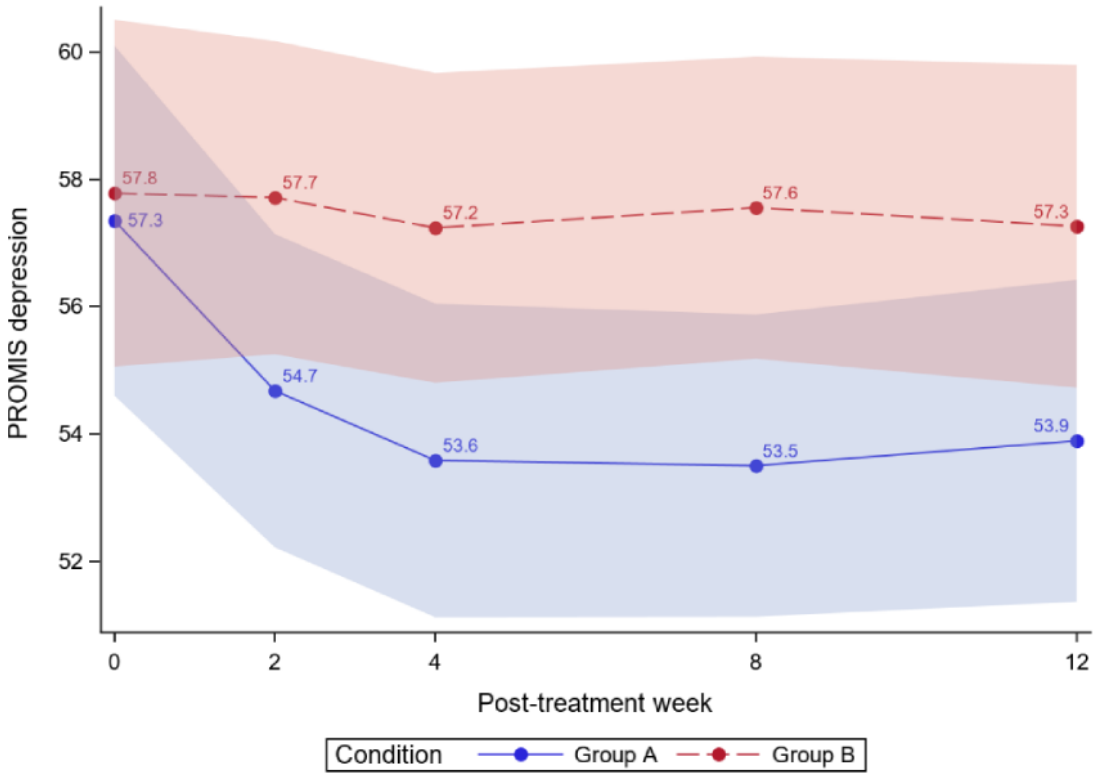


Figure 9. Physical function over time. PROMIS: Patient-Reported Outcomes Measurement Information System.

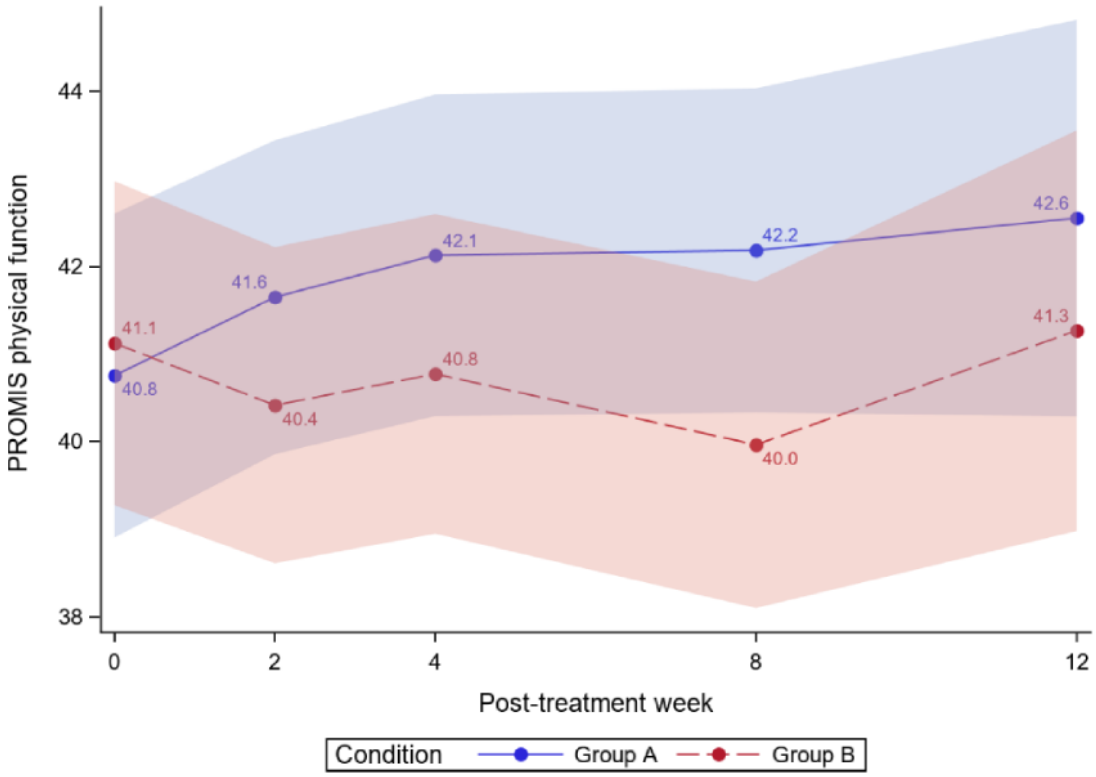


Figure 10. Pain interference over time. PROMIS: Patient-Reported Outcomes Measurement Information System.

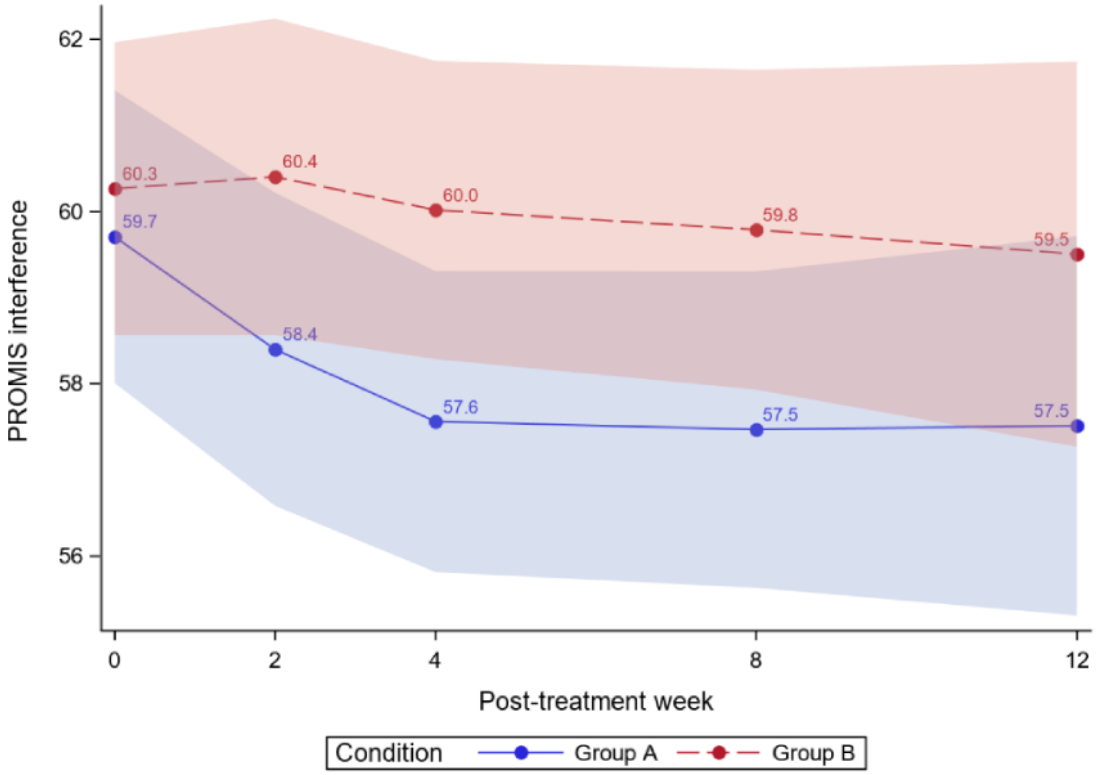


Figure 11. Social isolation over time. PROMIS: Patient-Reported Outcomes Measurement Information System.

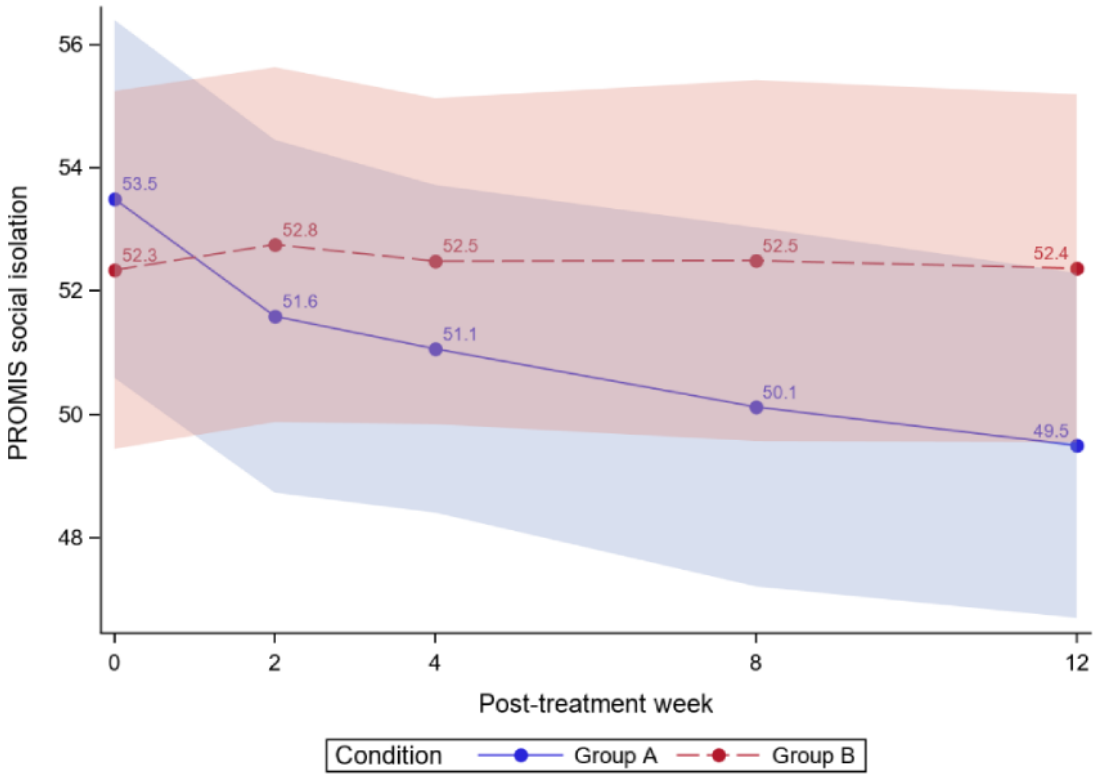
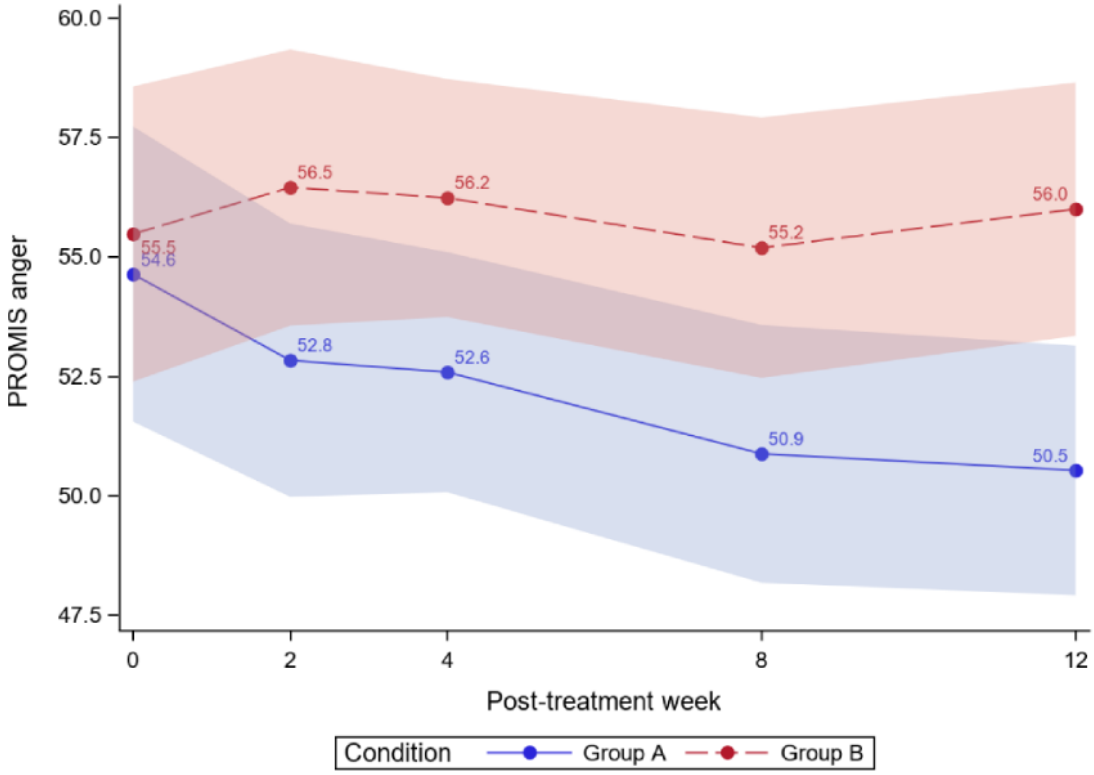


Figure 12. Anger over time. PROMIS: Patient-Reported Outcomes Measurement Information System.



Secondary Outcomes

Pain Intensity

Analyzing all data from baseline to 3 months, we observed a significant treatment effect (*P*=.02) on pain intensity; on average, the ER group had lower pain intensity scores than the WLC group (Cohen *d*=0.41). Separately, we observed a time effect; the average PCS significantly decreased over time for both study groups (time effect; *P*<.001). Most importantly, the decrease was greater for ER versus WLC (group×time effect; *P*=.01).

At 3 months, significant reductions in pain intensity were found for ER but not for WLC (ER: pain intensity −4.98, 8% reduction; WLC: pain intensity −1.46, 2.34% reduction). Figure 4 displays the average pain intensity during the 3-month follow-up by the study arm, and Table 4 displays the between-group comparisons for pain intensity at baseline and posttreatment months. At 3 months, ER was superior to WLC (difference −6.05, 95% CI −9.92 to −2.18; *P*<.001). Notably, when applying the 15% threshold for clinical meaningfulness, the reduction in pain intensity was not clinically meaningful for either group.

The ER group effect size was 0.76, with combined results showing a moderate-to-large effect size but no clinical importance. As much as 30% (15/50) of ER and 6% (3/51) of WLC participants achieved a 15% or greater reduction in pain intensity. For ER, 4% (2/50) achieved ≥30% reduction in pain intensity, whereas for WLC, 2% (1/51) reached that threshold.

Pain Bothersomeness

We observed a time effect; average pain bothersomeness significantly decreased over time for both study groups (time effect; *P*<.001); however, during the 1- to 3-month follow-ups, the between-group difference was significant (*P*=.01 to *P*=.02) per Table 4. This significance is largely driven by the initial decline in pain bothersomeness, which was steeper in the ER group and eventually plateaued for both groups (Figure 5). At 3 months, ER was superior to WLC (difference −1.30, 95% CI −2.29 to −0.31; *P*=.01).

At 3 months, clinically meaningful reductions in pain bothersomeness were found for ER but not for WLC (ER: pain bothersomeness −1.16, 20.07% reduction; WLC: pain bothersomeness −0.38, 6.03% reduction). This meets the threshold for moderate clinical meaningfulness. Figure 6 displays the average pain bothersomeness during the 3-month follow-up by the study arm, and Table 4 displays the between-group comparisons for pain intensity at baseline and posttreatment months.

The ER group effect size was 0.61, with combined results showing a moderate effect size and minimal clinical importance. As much as 34% (17/50) of ER and 12% (6/51) of WLC participants achieved ≥30% reduction in pain bothersomeness. For ER, 22% (11/50) achieved ≥50% in pain bothersomeness reduction, whereas for WLC, 3.9% (2/51) reached that threshold.

Sleep Disturbance

At 3 months, significant reductions in sleep disturbance were found for ER but not for WLC (ER: 2.82, 5.01% reduction; WLC: 1.7, 2.86% increase; Table 4). Separately, we did not observe a time effect (*P*=.52), and we observed a significant interaction effect (*P*=.02). Sleep disturbance was lower in the ER group than in the WLC group (*P*=.02). At 3 months, ER was superior to WLC (difference −7.74, 95% CI −11.61 to

−3.87; *P*<.001). The ER group effect size was 0.37, with combined results showing a small-to-moderate effect size and no clinical importance. As much as 12% (6/50) of ER and 8% (4/51) of WLC participants achieved a 15% or more reduction in sleep disturbance.

The Benjamini-Hochberg adjustment for multiple comparisons was applied across primary and secondary variables, and statistical significance was maintained (*P*=.02 to *P*<.001).

Tertiary Outcomes

Anxiety

At 3 months, significant reductions in anxiety were found for ER but not for WLC (ER: −4.71, 7.91% reduction; WLC: −0.42, 0.7% reduction; Table 4). Anxiety decreased over time in both groups, and the decline was significantly steeper in the ER group (*P*=.009). ER was superior to WLC (difference −5.14, 95% CI −8.7 to −1.58; *P*<.001). The ER group effect size was 0.59, with combined results showing a moderate effect size and no clinical importance.

Depression

We did not observe a treatment effect (*P*=.08) on depression. Separately, we observed a time effect; average depression significantly decreased over time for both study groups (time effect; *P*=.002). Most importantly, the decrease was greater for ER versus WLC (group×time effect; *P*=.03). However, between-group differences showed that ER was superior to WLC only during months 1 and 2 (*P*=.04 and *P*=.02, respectively). The overall interaction effect appears primarily driven by these two time-points and maintains the trend and directionality, but not statistically significant, at the 3-month follow-up (*P*=.07).

Physical Function

We did not observe a treatment effect (*P*=.36) or a time effect (*P*=.36) on physical function. We observed a trend for the interaction effect (*P*=.09). Between-subjects analysis revealed no significant differences at any of the follow-up time-points (*P*=.30 to *P*=.43), except at 2-months posttreatment (*P*=.01).

Exploratory Outcomes

For social isolation, we observed a time effect; average social isolation scores significantly decreased over time for both study groups (time effect; *P*=.02). Most importantly, this was qualified by a significant group×time interaction (*P*=.02); there was a decrease in social isolation in the ER group but not in the WLC group. No time or interaction effects were observed for pain interference or anger.

Zoom Platform Satisfaction

Participants in the ER arm provided feedback on their experience using the Zoom platform on a scale from 1 (strongly disagree) to 7 (strongly agree), with higher scores indicating a higher endorsement of each item. Overall, participants reported high satisfaction with the platform, including the ease of operating Zoom (46/50, 92% satisfaction: mean 6.45, SD 1.43), engaging with class material (mean 6.02, SD 1.45), comfort in engaging with Zoom instructor and participants (mean 5.68, SD 1.45), and feeling connected to the instructor (mean 5.43, SD 1.64).

Discussion

Principal Findings

We conducted the first Zoom-delivered randomized controlled trial of the group-based, SSI *ER* in a sample of individuals with mixed-etiology chronic pain. The primary goal of this study was to evaluate the feasibility, acceptability, and engagement of a single-session Zoom-delivered intervention for chronic pain. We also aimed to determine the preliminary efficacy of Zoom-delivered ER for reducing average pain intensity, pain catastrophizing, pain bothersomeness, and sleep disruption at 1-month posttreatment. Secondary and tertiary pain-related outcomes included physical function, anxiety, depression, and physical function. We also report a more rigorous test of the durability of treatment effects at 3-month posttreatment, as well as report the clinical meaningfulness of effects.

ER demonstrated high feasibility as indexed by high enrollment rate and excellent engagement; 94% (47/50) of participants in the ER group attended the class and completed all surveys through the 3-month duration of the study. The enrollment rate was 53.1% (104/196), which is five times higher than that observed in other studies [12,77] and stands as an index of both the inclusivity of the study protocol (few exclusionary criteria; all types of chronic pain) and the convenience of a single-session pain treatment study. Importantly, participants reported high overall satisfaction with the class (mean 8.26-8.77), the usefulness of presented information, and the likelihood of using the skills and information learned. In addition, participants reported high satisfaction with operating the Zoom platform (46/50, 92% satisfaction), in addition to high satisfaction with engaging with class material, other attendees, and the class instructor. Importantly, of the 9 participants who indicated they had never used the Zoom platform during screening, none requested additional or individual training with the platform. These metrics indicate superior feasibility, accessibility, and engagement with the Zoom-delivered ER intervention.

With respect to preliminary efficacy, time effects showed reductions in pain intensity and most pain-related indices, except for sleep disruption. ER demonstrated superior treatment effects on average pain intensity and pain-related catastrophizing, pain bothersomeness, and sleep disruption over the 3 months following treatment. The between-groups Cohen *d* effect sizes at 3-month posttreatment ranged from 0.60 to 0.91 for primary and secondary variables. For ER, pre- or posttreatment Cohen *d* effect sizes ranged from 0.37 to 0.89, demonstrating no clinical importance to moderate clinical importance for reducing pain-related catastrophizing, bothersomeness, and sleep disruption at the end of treatment. The treatment effect sizes across the primary and secondary outcomes suggested that a single-session, digital, skills-based intervention may have clinically meaningful effects on patient-reported outcomes. The durability of the treatment effects was demonstrated at the 3-month follow-up time-point. A greater proportion of participants in the ER group exceeded the thresholds for the clinical importance of effects at 3 months posttreatment.

First, for moderate importance in PCS reduction (≥30% reduction in PCS), 62% (31/50) in ER versus 24% (12/51) in

WLC met this threshold. For substantial clinical importance in PCS reduction (≥50% reduction in pain), 46% (23/50) in ER versus 12% (6/51) in WLC met this threshold. PCS reductions for ER (mean −8.72, SD 1.03) exceeded the clinically meaningful threshold of 6.8 reported in the literature [78] and is notably greater than that reported in other studies examining in-person 8-week CBT [7,10,79]. Our findings on PCS are expected and aligned with the literature suggesting that pain catastrophizing, an index of pain coping, is highly responsive to behavioral treatments [7,10,79], including SSI [11,12]. For context, multidisciplinary pain rehabilitation research has shown that PCS reductions of 38% are clinically important and associated with less disability and work status at one-year follow-up [80].

Second, for minimal clinical importance in pain reduction (≥15% reduction in pain), 30% (15/50) in ER versus 6% (3/51) in WLC met this threshold. The moderate clinical importance of pain reduction was low in both groups. Importantly, this study used a PROMIS pain measure with a stricter range (5-point scale) versus a traditional 11-point Numeric Pain Rating Scale. As such, the lack of variability in this measure may have negatively impacted our ability to detect the effects of pain intensity.

The tertiary outcomes yielded mixed findings. Reductions in anxiety and depression over time were found in both groups, and the decline was significantly greater for ER than WLC. However, these reductions did not reach the threshold for clinical significance. Notably, for depression, we observed regression to the mean at 3 months with the treatment effect lost, thus suggesting that more intensive treatment might be needed for durable reductions in depressive symptoms. This may be contextualized within widespread social isolation due to COVID-19, which has been shown to worsen depression symptoms over time [81,82], or other stressors related to COVID-19 or the seasonal context of this study.

Interestingly, no treatment or time effects were observed for physical function, only a trend toward significance (P=.09), which suggests that the study may be underpowered to detect differences or identify subgroups where effect may exist. In addition, due to a coding error, the short form used in this study did not capture upper extremity function, and scale items focused on lower mobility. As this was a mixed-etiology chronic pain study, comprehensive assessment of physical function is warranted to definitively explore its effects. Exploratory analyses revealed interaction effects for reductions in social isolation, which is a notable and important therapeutic target during the COVID-19 pandemic social restrictions. Although the two groups did not differ at 3 months, the interaction effect revealed a decrease in social isolation in the ER group but not in the WLC group. These findings lend support to ER serving as a buffer against the social threat perpetuated by the pandemic [16,83], which can ameliorate its impact on people's overall health status [65,66].

Although a growing body of research exists on the efficacy of ER for chronic and acute pain management [12,33,36,52], the use of Zoom as a platform to deliver behavioral medicine for chronic pain, particularly single-session treatment, remains novel and understudied. In addition, the extant literature on digital behavioral health research has reported participant treatment engagement rates ranging from 20% to 60% [33,84-86]. Strikingly, the current trial demonstrated a 94% (47/50) engagement rate in the treatment arm. These results also highlight the public interest in the web-based delivery of behavioral medicine interventions as a home-based chronic pain treatment modality. Combined with high participant engagement data and high satisfaction with using the Zoom platform, these data extend prior work [18-25] supporting the utility, user satisfaction, and efficacy of videoconference-delivered interventions for chronic pain.

Strengths and Limitations

Several findings should be considered within a number of limitations. First, given that the participant population was on average of higher socioeconomic status, as evidenced by their average education and income levels, the results may not be generalizable to populations of different demographics. No data were available on medical diagnoses, current prescriptions, over-the-counter medication use, or concurrent treatments for either study group. In addition, all data were self-reported, and we did not control for receipt of medical care. In light of COVID-19 restrictions, participants were likely to receive their care via web-based delivery as well. In addition, the pain intensity measure used was a 5-point Numeric Pain Rating Scale, as opposed to the 11-point Numeric Pain Rating Scale [87] typically used in research, limiting variability and may have obscured treatment effects and clinically meaningful findings. Finally, the participants were unblinded to the study group assignment. However, given that ER is low risk and low burden, placebo effects are less concerning. Replication studies were indicated with participants with more demographic variability over a longer follow-up period. Finally, future studies should characterize pain diagnoses and pain types.

Despite these limitations, this web-based study evidenced strong interest and participant engagement with single-session skills-based pain treatment, thus supporting further research and extension of web-based ER into clinical care. Second, the COVID-19 context and web-based delivery modality support the ecological validity of the study findings. Notably, the study was conducted remotely and did not benefit from any in-person contact that occurred when research was conducted in medical treatment settings (ie, halo effects). Additional aspects of methodological rigor included analyst blinding, intention-to-treat analyses, and randomization. Finally, our sample included minimal exclusions, which rendered it highly generalizable to other *real-world* individuals.

This study examined class cohorts comprising approximately 25 participants, and in-person class sizes ranged between 20 and 85, thus illustrating the promise of scalable pain care. Indeed, ER may help address pain disparities by ensuring rapid and more equitable access to pain treatment. Our findings provide initial evidence that the web-based delivery of ER may efficiently reduce the burden of chronic pain and improve symptom management. Additional research is needed to test the pragmatic comparative effectiveness of web-based ER to other treatments, such as gold-standard 8-week CBT, and to determine the heterogeneity of treatment effects.

Conclusions

In 2019, the US Health and Human Services cited ER as a promising scalable behavioral pain treatment [88]. For the first time, this study determined the preliminary efficacy of the Zoom-delivered class in reducing the burden of chronic pain and improving symptom management. Importantly, the web-based single-session class addresses the rapidly expanding need for alternatives to face-to-face encounters due to the COVID-19 pandemic. Web-based delivery of ER stands to largely improve patient access and engagement because it is adaptable to medical or community settings, is readily extendable to underserved populations (eg, rural locations), and may be offered at a low cost.

Acknowledgments

The authors thank the National Institute on Drug Abuse for supporting this work: K23DA047473 (MSZ), K23DA048972 (DSY), and NCCIH R01AT008561 (BDD).

Conflicts of Interest

Stanford University receives revenue for empowered relief, continuing medical education and instructor certification training provided to health care clinicians. BDD is Chief Science Advisor at AppliedVR and her consulting role with this company (personal fees) is unrelated to the current research. BDD receives royalties for four pain treatment books she has authored or coauthored. BDD is also the principal investigator for a pain research award from the Patient-Centered Research Outcomes Research Institute and pain research grants from the National Institutes of Health. BDD received consultant fees from Axial Healthcare related to physician education for opioid prescribing and deprescribing (2019; unrelated to the current work). BDD serves on the Board of Directors for the American Academy of Pain Medicine and is on the Board of Directors for the Institute for Brain Potential. BDD is a scientific member of the National Institutes of Health Interagency Pain Research Coordinating Committee, the Centers for Disease Control and Prevention Opioid Workgroup and is a member of the Pain Advisory Group of the American Psychological Association.

Multimedia Appendix 1
Consent form.
[DOC File , 62 KB - jmir_v23i9e29672_app1.doc]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 277 KB - jmir_v23i9e29672_app2.pdf]

References

1. Schappert SM, Burt CW. Ambulatory care visits to physician offices, hospital outpatient departments, and emergency departments: United States, 2001-02. Vital Health Stat 13 2006 Feb(159):1-66 [FREE Full text] [Medline: 16471269]
2. Institute of Medicine. Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education and Research. Washington, DC: The National Academies Press; 2011:55-112.
3. Becker N, Thomsen A, Olsen A, Sjøgren P, Bech P, Eriksen J. Pain epidemiology and health related quality of life in chronic non-malignant pain patients referred to a Danish multidisciplinary pain center. Pain 1997;73(3):393-400. [doi: 10.1016/S0304-3959(97)00126-7]
4. Shaheed CA, Maher CG, Williams KA, Day R, McLachlan AJ. Efficacy, tolerability, and dose-dependent effects of opioid analgesics for low back pain: a systematic review and meta-analysis. JAMA Intern Med 2016 Jul 01;176(7):958-968. [doi: 10.1001/jamainternmed.2016.1251] [Medline: 27213267]
5. Bernardy K, Füber N, Köllner V, Häuser W. Efficacy of cognitive-behavioral therapies in fibromyalgia syndrome - a systematic review and metaanalysis of randomized controlled trials. J Rheumatol 2010 Oct 03;37(10):1991-2005. [doi: 10.3899/jrheum.100104] [Medline: 20682676]
6. Darnall BD, Scheman J, Davin S, Burns JW, Murphy JL, Wilson AC, et al. Pain Psychology: A global needs assessment and national call to action. Pain Med 2016 Feb 23;17(2):250-263 [FREE Full text] [doi: 10.1093/pm/pnv095] [Medline: 26803844]
7. Williams ACC, Fisher E, Hearn L, Eccleston C. Psychological therapies for the management of chronic pain (excluding headache) in adults. Cochrane Database Syst Rev 2020 Aug 12;8(8):CD007407. [doi: 10.1002/14651858.CD007407.pub4] [Medline: 32794606]
8. Mariano TY, Urman RD, Hutchison CA, Jamison RN, Edwards RR. Cognitive behavioral therapy (CBT) for subacute low back pain: a systematic review. Curr Pain Headache Rep 2018 Feb 23;22(3):15. [doi: 10.1007/s11916-018-0669-5] [Medline: 29476270]

9. Morley S, Eccleston C, Williams A. Systematic review and meta-analysis of randomized controlled trials of cognitive behaviour therapy and behaviour therapy for chronic pain in adults, excluding headache. Pain 1999;80(1):1-13. [doi: 10.1016/S0304-3959(98)00255-3]
10. Cherkin DC, Sherman KJ, Balderson BH, Cook AJ, Anderson ML, Hawkes RJ, et al. Effect of mindfulness-based stress reduction vs cognitive behavioral therapy or usual care on back pain and functional limitations in adults with chronic low back pain: a randomized clinical trial. J Am Med Assoc 2016 Mar 22;315(12):1240-1249 [FREE Full text] [doi: 10.1001/jama.2016.2323] [Medline: 27002445]
11. Darnall BD, Ziadni MS, Roy A, Kao MC, Sturgeon JA, Cook KF, et al. Comparative efficacy and mechanisms of a single-session pain psychology class in chronic low back pain: study protocol for a randomized controlled trial. Trials 2018 Mar 06;19(1):165 [FREE Full text] [doi: 10.1186/s13063-018-2537-3] [Medline: 29510735]
12. Darnall BD, Roy A, Chen AL, Ziadni MS, Keane RT, You DS, et al. Comparison of a single-session pain management skills intervention with a single-session health education intervention and 8 sessions of cognitive behavioral therapy in adults with chronic low back pain: a randomized clinical trial. JAMA Network Open 2021 Aug 16;4(8):e2113401 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.13401]
13. Earnshaw VA, Quinn DM. The impact of stigma in healthcare on people living with chronic illnesses. J Health Psychol 2012 Mar 28;17(2):157-168. [doi: 10.1177/1359105311414952] [Medline: 21799078]
14. Sharf J, Primavera LH, Diener MJ. Dropout and therapeutic alliance: a meta-analysis of adult individual psychotherapy. Psychotherapy (Chic) 2010 Dec;47(4):637-645. [doi: 10.1037/a0021175] [Medline: 21198249]
15. Eccleston C, Blyth FM, Dear BF, Fisher EA, Keefe FJ, Lynch ME, et al. Managing patients with chronic pain during the COVID-19 outbreak: considerations for the rapid introduction of remotely supported (eHealth) pain management services. Pain 2020 May;161(5):889-893 [FREE Full text] [doi: 10.1097/j.pain.0000000000001885] [Medline: 32251203]
16. Kazdin AE. Annual Research Review: Expanding mental health services through novel models of intervention delivery. J Child Psychol Psychiatry 2019 Apr 13;60(4):455-472. [doi: 10.1111/jcpp.12937] [Medline: 29900543]
17. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. J Technol Hum Serv 2008 Jul 03;26(2-4):109-160. [doi: 10.1080/15228830802094429]
18. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res 2012 Dec;12(6):745-764. [doi: 10.1586/erp.12.67] [Medline: 23252357]
19. Beatty L, Lambert S. A systematic review of internet-based self-help therapeutic interventions to improve distress and disease-control among adults with chronic health conditions. Clin Psychol Rev 2013 Jun;33(4):609-622. [doi: 10.1016/j.cpr.2013.03.004] [Medline: 23603521]
20. Cuijpers P, van Straten A, Andersson G. Internet-administered cognitive behavior therapy for health problems: a systematic review. J Behav Med 2008 Apr;31(2):169-177 [FREE Full text] [doi: 10.1007/s10865-007-9144-1] [Medline: 18165893]
21. Garg S, Garg D, Turin TC, Chowdhury MF. Web-based interventions for chronic back pain: a systematic review. J Med Internet Res 2016 Jul 26;18(7):e139 [FREE Full text] [doi: 10.2196/jmir.4932] [Medline: 27460413]
22. Dhokia M, Elander J, Clements K, Gilbert P. A randomized-controlled pilot trial of an online compassionate mind training intervention to help people with chronic pain avoid analgesic misuse. Psychol Addict Behav 2020 Nov;34(7):726-733. [doi: 10.1037/adb0000579] [Medline: 32271055]
23. Young SD, Koussa M, Lee S, Perez H, Gill N, Gelberg L, et al. Feasibility of a social media/online community support group intervention among chronic pain patients on opioid therapy. J Addict Dis 2018;37(1-2):96-101 [FREE Full text] [doi: 10.1080/10550887.2018.1557992] [Medline: 30614403]
24. Wilson M, Finlay M, Orr M, Barbosa-Leiker C, Sherazi N, Roberts ML, et al. Engagement in online pain self-management improves pain in adults on medication-assisted behavioral treatment for opioid use disorders. Addict Behav 2018 Nov;86:130-137. [doi: 10.1016/j.addbeh.2018.04.019] [Medline: 29731244]
25. Flynn N. Effect of an online hypnosis intervention in reducing migraine symptoms: a randomized controlled trial. Int J Clin Exp Hypn 2019 Jun 28;67(3):313-335. [doi: 10.1080/00207144.2019.1612674] [Medline: 31251706]
26. Kazdin AE, Blase SL. Interventions and models of their delivery to reduce the burden of mental illness: reply to commentaries. Perspect Psychol Sci 2011 Sep;6(5):507-510. [doi: 10.1177/1745691611418241] [Medline: 26168204]
27. Schleider JL, Weisz JR. Little treatments, promising effects? Meta-analysis of single-session interventions for youth psychiatric problems. J Am Acad Child Adolesc Psychiatry 2017 Feb;56(2):107-115. [doi: 10.1016/j.jaac.2016.11.007] [Medline: 28117056]
28. Depp C, Perivoliotis D, Holden J, Dorr J, Granholm E. Single-session mobile-augmented intervention in serious mental illness: a three-arm randomized controlled trial. Schizophr Bull 2019 Jun 18;45(4):752-762 [FREE Full text] [doi: 10.1093/schbul/sby135] [Medline: 30281086]
29. Ellis J, Cushing T, Germain A. Treating Acute Insomnia: A randomized controlled trial of a "Single-Shot" of cognitive behavioral therapy for insomnia. Sleep 2015 Jun 01;38(6):971-978 [FREE Full text] [doi: 10.5665/sleep.4752] [Medline: 25515106]

30. Samson JE, Tanner-Smith EE. Single-session alcohol interventions for heavy drinking college students: a systematic review and meta-analysis. J Stud Alcohol Drugs 2015 Jul;76(4):530-543 [FREE Full text] [doi: 10.15288/jsad.2015.76.530] [Medline: 26098028]
31. Ziadni MS, Carty JN, Doherty HK, Porcerelli JH, Rapport LJ, Schubiner H, et al. A life-stress, emotional awareness, and expression interview for primary care patients with medically unexplained symptoms: a randomized controlled trial. Health Psychol 2018 Mar;37(3):282-290 [FREE Full text] [doi: 10.1037/hea0000566] [Medline: 29154608]
32. Carty JN, Ziadni MS, Holmes HJ, Tomakowsky J, Peters K, Schubiner H, et al. The effects of a life stress emotional awareness and expression interview for women with chronic urogenital pain: a randomized controlled trial. Pain Med 2019 Jul 01;20(7):1321-1329 [FREE Full text] [doi: 10.1093/pm/pny182] [Medline: 30252113]
33. Darnall BD, Ziadni MS, Krishnamurthy P, Flood P, Heathcote LC, Mackey IG, et al. "My Surgical Success": Effect of a digital behavioral pain medicine intervention on time to opioid cessation after breast cancer surgery-a pilot randomized controlled clinical trial. Pain Med 2019 Nov 01;20(11):2228-2237 [FREE Full text] [doi: 10.1093/pm/pnz094] [Medline: 31087093]
34. Ziadni MS, Chen AL, Winslow T, Mackey SC, Darnall BD. Efficacy and mechanisms of a single-session behavioral medicine class among patients with chronic pain taking prescription opioids: study protocol for a randomized controlled trial. Trials 2020 Jun 12;21(1):521 [FREE Full text] [doi: 10.1186/s13063-020-04415-x] [Medline: 32532346]
35. Ziadni MS, Gonzalez-Castro LA, Darnall BD. Virtual single-session “Empowered Relief” group intervention for chronic pain. ClinicalTrials.gov. 2020 Sep 14. URL: https://clinicaltrials.gov/ct2/show/NCT04546685?id=NCT04546685&draw=2&rank=1 [accessed 2021-08-09]
36. Darnall BD, Sturgeon JA, Kao MC, Hah JM, Mackey SC. From Catastrophizing to Recovery: a pilot study of a single-session treatment for pain catastrophizing. J Pain Res 2014;7:219-226 [FREE Full text] [doi: 10.2147/JPR.S62329] [Medline: 24851056]
37. Schleider JL, Dobias ML, Sung JY, Mullarkey MC. Future directions in single-session youth mental health interventions. J Clin Child Adolesc Psychol 2020 Dec 04;49(2):264-278 [FREE Full text] [doi: 10.1080/15374416.2019.1683852] [Medline: 31799863]
38. Alschuler KN, Altman JK, Ehde DM. Feasibility and acceptability of a single-session, videoconference-delivered group intervention for pain in multiple sclerosis. Rehabil Psychol 2021 Feb;66(1):22-30. [doi: 10.1037/rep0000360] [Medline: 32790454]
39. Rodda SN, Lubman DI, Jackson AC, Dowling NA. Improved outcomes following a single session web-based intervention for problem gambling. J Gambl Stud 2017 Mar;33(1):283-299. [doi: 10.1007/s10899-016-9638-2] [Medline: 27562272]
40. Schleider JL, Dobias M, Sung J, Mumper E, Mullarkey MC. Acceptability and utility of an open-access, online single-session intervention platform for adolescent mental health. JMIR Ment Health 2020 Jun 30;7(6):e20513 [FREE Full text] [doi: 10.2196/20513] [Medline: 32602846]
41. Wasil AR, Park SJ, Gillespie S, Shingleton R, Shinde S, Natu S, et al. Harnessing single-session interventions to improve adolescent mental health and well-being in India: Development, adaptation, and pilot testing of online single-session interventions in Indian secondary schools. Asian J Psychiatr 2020 Apr;50:101980. [doi: 10.1016/j.ajp.2020.101980] [Medline: 32146337]
42. Cardamone-Breen MC, Jorm AF, Lawrence KA, Rapee RM, Mackinnon AJ, Yap MB. A single-session, web-based parenting intervention to prevent adolescent depression and anxiety disorders: randomized controlled trial. J Med Internet Res 2018 Dec 26;20(4):e148 [FREE Full text] [doi: 10.2196/jmir.9499] [Medline: 29699964]
43. Hymmen P, Stalker CA, Cait C. The case for single-session therapy: Does the empirical evidence support the increased prevalence of this service delivery model? J Ment Health 2013 Feb 23;22(1):60-71. [doi: 10.3109/09638237.2012.670880] [Medline: 23343048]
44. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
45. Zoom video communications. Zoom. URL: https://stanford.zoom.us [accessed 2021-08-09]
46. Kaldo V, Levin S, Widarsson J, Buhrman M, Larsen H, Andersson G. Internet versus group cognitive-behavioral treatment of distress associated with tinnitus: a randomized controlled trial. Behav Ther 2008 Dec;39(4):348-359. [doi: 10.1016/j.beth.2007.10.003] [Medline: 19027431]
47. Riva S, Camerini A, Allam A, Schulz PJ. Interactive sections of an internet-based intervention increase empowerment of chronic back pain patients: randomized controlled trial. J Med Internet Res 2014 Aug;16(8):e180 [FREE Full text] [doi: 10.2196/jmir.3474] [Medline: 25119374]
48. Buhrman M, Fältenhag S, Ström L, Andersson G. Controlled trial of internet-based treatment with telephone support for chronic back pain. Pain 2004 Oct;111(3):368-377. [doi: 10.1016/j.pain.2004.07.021] [Medline: 15363881]
49. Brattberg G. Internet-based rehabilitation for individuals with chronic pain and burnout: a randomized trial. Int J Rehab Res 2006;29(3):221-227. [doi: 10.1097/01.mrr.0000210055.17291.f5] [Medline: 16900043]

50. Ferguson RJ, Sigmon ST, Pritchard AJ, LaBrie SL, Goetze RE, Fink CM, et al. A randomized trial of videoconference-delivered cognitive behavioral therapy for survivors of breast cancer with self-reported cognitive dysfunction. Cancer 2016 Jun 01;122(11):1782-1791 [FREE Full text] [doi: 10.1002/cncr.29891] [Medline: 27135464]
51. Owen J, Klapow J, Roth D, Shuster J, Bellis J, Meredith R, et al. Randomized pilot of a self-guided internet coping group for women with early-stage breast cancer. Ann Behav Med 2005 Aug;30(1):54-64. [doi: 10.1207/s15324796abm3001_7] [Medline: 16097906]
52. Ziadni MS, Anderson SR, Gonzalez-Castro L, Darnall BD. Comparative efficacy of a single-session "Empowered Relief" videoconference-delivered group intervention for chronic pain: study protocol for a randomized controlled trial. Trials 2021 May 22;22(1):358 [FREE Full text] [doi: 10.1186/s13063-021-05303-8] [Medline: 34022930]
53. Empowered Relief (ER) Certification. Stanford University. URL: https://empoweredrelief.stanford.edu [accessed 2021-08-09]
54. Hart T, Fann JR, Novack TA. The dilemma of the control condition in experience-based cognitive and behavioural treatment research. Neuropsychol Rehabil 2008 Jan;18(1):1-21. [doi: 10.1080/09602010601082359] [Medline: 17852761]
55. Mohr DC, Spring B, Freedland KE, Beckner V, Arean P, Hollon SD, et al. The selection and design of control conditions for randomized controlled trials of psychological interventions. Psychother Psychosom 2009;78(5):275-284. [doi: 10.1159/000228248] [Medline: 19602916]
56. Sullivan MJ, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and validation. Psychol Assess 1995;7(4):524-532. [doi: 10.1037/1040-3590.7.4.524]
57. Osman A, Barrios F, Gutierrez P, Kopper B, Merrifield T, Grittmann L. The pain catastrophizing scale: further psychometric evaluation with adult samples. J Behav Med 2000;23(4):351-365. [doi: 10.1023/A:1005548801037]
58. Turner JA, Anderson ML, Balderson BH, Cook AJ, Sherman KJ, Cherkin DC. Mindfulness-based stress reduction and cognitive behavioral therapy for chronic low back pain: similar effects on mindfulness, catastrophizing, self-efficacy, and acceptance in a randomized controlled trial. Pain 2016 Nov;157(11):2434-2444 [FREE Full text] [doi: 10.1097/j.pain.0000000000000635] [Medline: 27257859]
59. Hung M, Hon SD, Franklin JD, Kendall RW, Lawrence BD, Neese A, et al. Psychometric properties of the PROMIS physical function item bank in patients with spinal disorders. Spine (Phila Pa 1976) 2014 Jan 15;39(2):158-163. [doi: 10.1097/BRS.0000000000000097] [Medline: 24173018]
60. Karayannis N, Baumann I, Sturgeon J, Melloh M, Mackey S. The impact of social isolation on pain interference: a longitudinal study. Ann Behav Med 2019 Jan 01;53(1):65-74 [FREE Full text] [doi: 10.1093/abm/kay017] [Medline: 29668841]
61. Karayannis NV, Sturgeon JA, Chih-Kao M, Cooley C, Mackey SC. Pain interference and physical function demonstrate poor longitudinal association in people living with pain: a PROMIS investigation. Pain 2017 Jun;158(6):1063-1068 [FREE Full text] [doi: 10.1097/j.pain.0000000000000881] [Medline: 28221284]
62. Lai J, Cella D, Choi S, Junghaenel DU, Christodoulou C, Gershon R, et al. How item banks and their application can influence measurement practice in rehabilitation medicine: a PROMIS fatigue item bank example. Arch Phys Med Rehabil 2011 Oct;92(10 Suppl):20-27 [FREE Full text] [doi: 10.1016/j.apmr.2010.08.033] [Medline: 21958919]
63. Revicki D, Chen W, Harnam N, Cook K, Amtmann D, Callahan L, et al. Development and psychometric analysis of the PROMIS pain behavior item bank. Pain 2009 Nov;146(1-2):158-169 [FREE Full text] [doi: 10.1016/j.pain.2009.07.029] [Medline: 19683873]
64. Revicki DA, Cook KF, Amtmann D, Harnam N, Chen W, Keefe FJ. Exploratory and confirmatory factor analysis of the PROMIS pain quality item bank. Qual Life Res 2013 Jul 9;23(1):245-255. [doi: 10.1007/s11136-013-0467-9] [Medline: 23836435]
65. Sturgeon JA, Carriere JS, Kao MC, Rico T, Darnall BD, Mackey SC. Social disruption mediates the relationship between perceived injustice and anger in chronic pain: a collaborative health outcomes information registry study. Ann Behav Med 2016 Dec 20;50(6):802-812 [FREE Full text] [doi: 10.1007/s12160-016-9808-6] [Medline: 27325314]
66. Sturgeon JA, Dixon EA, Darnall BD, Mackey SC. Contributions of physical function and satisfaction with social roles to emotional distress in chronic pain: a Collaborative Health Outcomes Information Registry (CHOIR) study. Pain 2015 Dec;156(12):2627-2633 [FREE Full text] [doi: 10.1097/j.pain.0000000000000313] [Medline: 26230739]
67. Sturgeon JA, Darnall BD, Kao MC, Mackey SC. Physical and psychological correlates of fatigue and physical function: a Collaborative Health Outcomes Information Registry (CHOIR) study. J Pain 2015 Mar;16(3):291-298 [FREE Full text] [doi: 10.1016/j.jpain.2014.12.004] [Medline: 25536536]
68. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, IMMPACT. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005 Jan;113(1-2):9-19. [doi: 10.1016/j.pain.2004.09.012] [Medline: 15621359]
69. Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. Pain 2003 Dec;106(3):337-345. [doi: 10.1016/j.pain.2003.08.001] [Medline: 14659516]
70. PROMIS-Assessment. Assessment Center. URL: https://www.assessmentcenter.net/ [accessed 2021-08-09]
71. Gershon RC, Rothrock N, Hanrahan R, Bass M, Cella D. The use of PROMIS and assessment center to deliver patient-reported outcome measures in clinical research. J Appl Meas 2010;11(3):304-314 [FREE Full text] [Medline: 20847477]

72. Fischer HF, Rose M. Scoring depression on a common metric: a comparison of EAP estimation, plausible value imputation, and full Bayesian IRT modeling. Multivariate Behav Res 2019 Sep 20;54(1):85-99. [doi: 10.1080/00273171.2018.1491381] [Medline: 30235003]
73. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain 2006 May;10(4):287-333. [doi: 10.1016/j.ejpain.2005.06.009] [Medline: 16095934]
74. Demyttenaere K, Bruffaerts R, Lee S, Posada-Villa J, Kovess V, Angermeyer M, et al. Mental disorders among persons with chronic back or neck pain: results from the World Mental Health Surveys. Pain 2007 Jun;129(3):332-342. [doi: 10.1016/j.pain.2007.01.022] [Medline: 17350169]
75. Gore M, Sadosky A, Stacey BR, Tai K, Leslie D. The burden of chronic low back pain: clinical comorbidities, treatment patterns, and health care costs in usual care settings. Spine 2012;37(11):668-677. [doi: 10.1097/brs.0b013e318241e5de]
76. Tsang A, Von Korff M, Lee S, Alonso J, Karam E, Angermeyer MC, et al. Common chronic pain conditions in developed and developing countries: gender and age differences and comorbidity with depression-anxiety disorders. J Pain 2008 Oct;9(10):883-891. [doi: 10.1016/j.jpain.2008.05.005] [Medline: 18602869]
77. Garcia LM, Birckhead BJ, Krishnamurthy P, Sackman J, Mackey IG, Louis RG, et al. An 8-week self-administered at-home behavioral skills-based virtual reality program for chronic low back pain: double-blind, randomized, placebo-controlled trial conducted during COVID-19. J Med Internet Res 2021 Feb 22;23(2):e26292 [FREE Full text] [doi: 10.2196/26292] [Medline: 33484240]
78. Suzuki H, Aono S, Inoue S, Imajo Y, Nishida N, Funaba M, et al. Clinically significant changes in pain along the Pain Intensity Numerical Rating Scale in patients with chronic low back pain. PLoS One 2020 Mar 3;15(3):e0229228 [FREE Full text] [doi: 10.1371/journal.pone.0229228] [Medline: 32126108]
79. Cherkin DC, Sherman KJ, Balderson BH, Turner JA, Cook AJ, Stoelb B, et al. Comparison of complementary and alternative medicine with conventional mind-body therapies for chronic back pain: protocol for the Mind-body Approaches to Pain (MAP) randomized controlled trial. Trials 2014 Jun 07;15(1):211 [FREE Full text] [doi: 10.1186/1745-6215-15-211] [Medline: 24906419]
80. Scott W, Wideman T, Sullivan M. Clinically meaningful scores on pain catastrophizing before and after multidisciplinary rehabilitation: a prospective study of individuals with subacute pain after whiplash injury. Clin J Pain 2014 Mar;30(3):183-190. [doi: 10.1097/AJP.0b013e31828eee6c] [Medline: 23552561]
81. Ettman CK, Abdalla SM, Cohen GH, Sampson L, Vivier PM, Galea S. Prevalence of depression symptoms in US adults before and during the COVID-19 pandemic. JAMA Netw Open 2020 Sep 01;3(9):e2019686 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.19686] [Medline: 32876685]
82. Twenge JM, Joiner TE. U.S. Census Bureau-assessed prevalence of anxiety and depressive symptoms in 2019 and during the 2020 COVID-19 pandemic. Depress Anxiety 2020 Oct;37(10):954-956 [FREE Full text] [doi: 10.1002/da.23077] [Medline: 32667081]
83. Karos K, McParland JL, Bunzli S, Devan H, Hirsh A, Kapos FP, et al. The social threats of COVID-19 for people with chronic pain. Pain 2020 Oct;161(10):2229-2235 [FREE Full text] [doi: 10.1097/j.pain.0000000000002004] [Medline: 32694381]
84. Geraghty AW, Torres LD, Leykin Y, Pérez-Stable EJ, Muñoz RF. Understanding attrition from international internet health interventions: a step towards global eHealth. Health Promot Int 2013 Sep 10;28(3):442-452 [FREE Full text] [doi: 10.1093/heapro/das029] [Medline: 22786673]
85. Geraghty AW, Wood AM, Hyland ME. Attrition from self-directed interventions: investigating the relationship between psychological predictors, intervention content and dropout from a body dissatisfaction intervention. Soc Sci Med 2010 Jul;71(1):30-37. [doi: 10.1016/j.socscimed.2010.03.007] [Medline: 20400220]
86. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
87. Farrar JT, Young JP, LaMoreaux L, Werth JL, Poole MR. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001;94(2):149-158. [doi: 10.1016/S0304-3959(01)00349-9]
88. U.S. Department of Health and Human Services. Pain Management Best Practices Inter-Agency Task Force Report: Updates, Gaps, Inconsistencies, and Recommendations. Washington, D.C: U. S. Department of Health and Human Services; 2019. URL: https://www.hhs.gov/sites/default/files/pmtf-final-report-2019-05-23.pdf [accessed 2021-08-09]

Abbreviations

CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
ER: empowered relief
HIPAA: Health Insurance Portability and Accountability Act
PCS: pain catastrophizing scale
PROMIS: Patient-Reported Outcomes Measurement Information System
REDCap: Research Electronic Data Capture

SSI: single-session intervention
WLC: waitlist control

Edited by G Eysenbach; submitted 20.04.21; peer-reviewed by L Martire, C Ashton-James; comments to author 11.05.21; revised version received 08.06.21; accepted 01.08.21; published 10.09.21.

Please cite as:
Ziadni MS, Gonzalez-Castro L, Anderson S, Krishnamurthy P, Darnall BD
Efficacy of a Single-Session "Empowered Relief" Zoom-Delivered Group Intervention for Chronic Pain: Randomized Controlled Trial Conducted During the COVID-19 Pandemic
J Med Internet Res 2021;23(9):e29672
URL: https://www.jmir.org/2021/9/e29672
doi:10.2196/29672
PMID:34505832

Original Paper

Framing of and Attention to COVID-19 on Twitter: Thematic Analysis of Hashtags

Iman Tahamtan[1], PhD; Devendra Potnis[1], PhD; Ehsan Mohammadi[2], PhD; Laura E Miller[3], PhD; Vandana Singh[1], PhD

[1]School of Information Sciences, The University of Tennessee, Knoxville, TN, United States

[2]School of Information Science, The University of South Carolina, Columbia, SC, United States

[3]School of Communication Studies, The University of Tennessee, Knoxville, TN, United States

Corresponding Author:
Iman Tahamtan, PhD
School of Information Sciences
The University of Tennessee
1345 Circle Park Drive 451 Communications Building
Knoxville, TN, 37996-0332
United States
Phone: 1 865 974 2148
Email: iman.tahamtan@gmail.com

Abstract

Background: Although past research has focused on COVID-19–related frames in the news media, such research may not accurately capture and represent the perspectives of people from diverse backgrounds. Additionally, research on the public attention to COVID-19 as reflected through frames on social media is scarce.

Objective: This study identified the frames about the COVID-19 pandemic in the public discourse on Twitter, which voices diverse opinions. This study also investigated the amount of public attention to those frames on Twitter.

Methods: We collected 22 trending hashtags related to COVID-19 in the United States and 694,582 tweets written in English containing these hashtags in March 2020 and analyzed them via thematic analysis. Public attention to these frames was measured by evaluating the amount of public engagement with frames and public adoption of those frames.

Results: We identified 9 frames including “public health guidelines,” “quarantine life,” “solidarity,” “evidence and facts,” “call for action,” “politics,” “post-pandemic life,” “shortage panic,” and “conflict.” Results showed that some frames such as “call for action” are more appealing than others during a global pandemic, receiving greater public adoption and engagement. The “call for action” frame had the highest engagement score, followed by “conflict” and “evidence and facts.” Additionally, “post-pandemic life” had the highest adoption score, followed by “call for action” and “shortage panic.” The findings indicated that the frequency of a frame on social media does not necessarily mean greater public adoption of or engagement with the frame.

Conclusions: This study contributes to framing theory and research by demonstrating how trending hashtags can be used as new user-generated data to identify frames on social media. This study concludes that the identified frames such as “quarantine life” and “conflict” and themes such as “isolation” and “toilet paper panic” represent the consequences of the COVID-19 pandemic. The consequences could be (1) exclusively related to COVID-19, such as hand hygiene or isolation; (2) related to any health crisis such as social support of vulnerable groups; and (3) generic that are irrespective of COVID-19, such as homeschooling or remote working.

(J Med Internet Res 2021;23(9):e30800) doi:10.2196/30800

KEYWORDS

COVID-19; framing; Twitter; social media; public opinion; engagement; public attention; thematic analysis; public health

Introduction

Public Opinion on Twitter

COVID-19 is a global public health pandemic threatening millions of lives worldwide, leading to approximately 188,655,968 confirmed cases and 4,067,517 deaths across the globe as of July 16, 2021 [1]. Twitter was used as one of the major platforms for disseminating information and knowledge about the COVID-19 pandemic, resulting in massive data generated by the public about various aspects of the virus [2]. Twitter has become the most frequently used communication medium for disseminating health information since the outbreak of H1N1 in 2009 [3] and later during the outbreak of the H7N9 virus, or bird flu, in 2013 [4]. It was also a platform for discussing the Zika virus epidemic in 2015 and 2016 [5].

People use Twitter and other social media platforms to interact with, share their opinion about, and engage with public health messages in real time. Twitter allows public health gatekeepers to interact with the public directly. For instance, the Centers for Disease Control and Prevention, the World Health Organization (WHO), health care officials, and organizations used Twitter regularly to share public health messages about the pandemic and communicate its risks to the public [2].

The data shared on Twitter can be used to analyze and study public opinion. Understanding public opinion could help researchers and authorities identify the public's needs, priorities, preferences, and behavior in real time. In turn, these data could impact public policy by encouraging governments and health care officials to distribute proper resources, take actions, and plan accordingly [6-8]. For instance, Avery [9] reported that monitoring public opinion about the Zika virus crisis on social media helped public information officers have a higher level of preparedness for managing the crisis.

Frame Analysis of Public Opinion

Public opinion refers to people's collective opinion about an issue such as COVID-19. The public consists of all the groups and subgroups in society, such as workers, doctors, officials, politicians, journalists, and students [10]. An effective strategy to understanding public opinion and attention to an issue is analyzing how people perceive and frame the issue [11,12]. Framing refers to selecting some aspects of an issue, promoting them, and making them salient [13]. Journalists and news media often use framing to conceptualize an issue, bring public attention to some aspects of the issue, and minimize attention to other aspects of it [14,15]. People also use framing to make sense of complex information, interpret and organize these ideas into comprehensible concepts, and present them to others [16].

Analyzing public opinion by frame analysis provides insights into the content and sentiment-based aspects of an issue [17]. Frames can serve as a starting point for designing effective messages to address people's needs and concerns during public health crises [18]. Journalists, policymakers, professionals, and scientists use frames to communicate their messages in a more effective way that is easily understandable by the public [19]. Frames can also be used by governments, advocacy groups, and authorities to design public education materials, present more engaging and effective public dialogue, write more relevant stories for the public, and expand their audience, reach, and impact [20]. For instance, framing of messages related to COVID-19 can be used to design and target messages in an effective way to enhance public engagement with public health guidelines [21].

Effect of Frames on Public Attention

An important aspect of framing research is to study the effect of frames on public attention. "Public attention refers to the general acknowledgement of a subject in the public sphere and subsequent civic discourse on the subject" [22]. Different frames about an issue can cause varied effects on public attention. A small change in how a topic is presented can sometimes cause a butterfly effect on public attention [15]. Framing or reframing an issue can shift how people understand the story, consequently changing how people respond to it [16]. Once the public turns its attention towards a series of frames about an issue and adopts those frames, it is likely that they collectively agree on the best decision and course of action [15]. The study by Krishnamurthy et al [23] showed that, when discussing the performance of medical treatments, the messages that are framed positively (ie, the chance of treatment success rather than its failure) had a greater impact on health treatment decisions. Almashat et al [24] indicated that survival frames (ie, the likelihood of surviving a certain procedure) would lead to making more informed medical decisions than mortality frames (ie, the possibility of dying from a certain procedure). Analyzing the effect of COVID-19 frames on public attention can be used to determine which frames are more effective in receiving greater public attention and impacting public opinion.

In summary, Twitter provides valuable user-generated data that can be beneficial for different stakeholders to respond to health crises. Although some studies have analyzed the frames about COVID-19 with a focus on the news media of politicians [18,25-28], there is no systematic study about the frames in public discourse on Twitter for this global pandemic. Therefore, this study aimed to fill this gap with the following research objectives.

Research Objectives

The aim of this research was to (1) identify the frames associated with the COVID-19 pandemic in public discourse on Twitter and (2) analyze which of the COVID-19 frames in public discourse have received greater public attention on Twitter.

Literature Review

Framing of the COVID-19 Pandemic

During public health emergencies, it is critical to communicate public health messages and guidelines to the public effectively. One way to make public health messages more effective is to frame them in a way that helps people understand health crises and positively impact their decision making and behavior. Previous empirical research shows that the news media play an important role in framing public health crises [25]. As evident in Multimedia Appendix 1, Ogbodo et al [26] analyzed the frames in 8 leading global media outlets globally, such as the BBC, News York Times, CNN, and People's Daily. Poirier et

al [27] investigated the frames used on the front pages of 12 well-known Canadian news media sources. Park et al [18] analyzed Korean COVID-19–related tweets to identify medical and nonmedical news frames, and Yu et al [25] studied the frames in the tweets of 2 major newspapers in Spain: El Mundo and El País.

The literature on the framing of COVID-19 shows that most past studies have focused on how COVID-19 has been framed by political figures and the news media [28]. These studies reflect the frames promoted by established gatekeepers, namely the news media, in society who may have some hidden agenda. Unlike previous framing studies, this study attempts to fill this gap by focusing on the frames related to COVID-19 in public discourse on Twitter, which voices the opinions of diverse groups of people.

Hashtags for Framing Messages on Twitter

Framing studies uses the features of communicating text to identify frames. The linguistic features include headlines or subheads of newspapers; photographs and photo captions; paragraphs of articles [29]; social media posts, such as tweets [30]; and other visual images, like icons [31]. For instance, Hellmueller and Zhang [32] used photographs to identify frames regarding the European refugee crisis in CNN International and Der Spiegel online news sites. Benziman [28] used US President and British Prime Minister speeches to identify how they framed issues related to COVID-19.

Hashtags are a linguistic feature that can be used to identify frames on Twitter and other social media platforms. Hashtags serve as catchphrases that can identify frames in tweets [33]. They represent a shared meaning or context [34] that, similar to frames, can highlight the most salient aspects of an issue (see [35]) in such a way as to promote it (see [13]).

In such circumstances as breaking or emerging events, a specific hashtag (eg, #BlackLivesMatter) or a set of hashtags (eg, #StayHome, #WashYourHands) become the main channel to represent an issue in online conversations by social media users. Hashtags that gain community-wide adoption and popularity can be used to identify the ad hoc framing of an event. Widely adopted hashtags show how the public frames a topic and can be used to determine the thematic frames of the issue [17]. Nevertheless, limited studies have acknowledged the possibility of using hashtags to identify frames (eg, [17,33]). Hence, this research explored hashtags in frame analysis as a new approach for framing theory and research.

Trending Hashtags in Frame Analysis

Among the numerous widely adopted hashtags on social media, some may become trendy and viral. Popular hashtags emerge in response to breaking news and other unexpected events, such as when an important, nationwide, or global issue happens [36,37].

Hashtags may go viral on Twitter as more and more people begin to use them [38]. Many users engage with trending topics and hashtags on social media [36]. Trending hashtags or topics often do not last long on Twitter [39]; however, they receive an initial increase of public attention, and then the focus of the public shifts elsewhere [34]. Trending topics and hashtags represent which issues have drawn the most public attention [36]. Trending hashtags also represent a community of online users who attend a unique topic or event for a limited period [34]. The mechanisms by which Twitter identifies the top trending topics and hashtags [36] or what causes some topics to become widely popular are not clear [39]. Nonetheless, such popularity peaks are of great relevance for identifying the issues that are the focal point of the public. As such, trending hashtags are suitable tools to be used in frame analysis of social media posts.

Benefits of Using Hashtags in Frame Analysis

Using hashtags in frame analysis has 2 benefits: It facilitates framing public opinion on social media, and it mitigates the subjectivity issue in frame analysis. Hashtags, similar to photographs [40], make it easy for researchers to understand what the tweet's content is about because they are the commonly accepted public signals for framing and presenting an event or topic among all social media platforms. While framing research on social media has focused mostly on analyzing social media posts, the question of how issues are framed through hashtags has remained relatively underexamined.

Content analysis is a dominant method for analyzing and identifying frames on social media (eg, [41]) and in non-social media contexts (eg, [42]). Matthes and Kohring [43] noted that a major issue with identifying frames through content analysis is that it involves researchers' subjective biases in analyzing and coding the text differently. Researchers' subjective biases question the validity and reliability of content analysis and the results in frame analysis [43]. Subjectivity is nearly impossible to avoid [19]; however, using hashtags to identify frames mitigates the subjective role of the researcher in frame analysis because hashtags have a "classificatory function" of indicating what a social media post is about [44]. Therefore, hashtags enhance the researcher's capability to describe online information [45] and identify "textual aboutness" [46] of social media posts. In some cases, hashtags may not stand alone in understanding textual aboutness because textual aboutness can be sensitive to context. For instance, a hashtag might reflect 2 different meanings simultaneously [44].

The Effect of Frames on Public Attention on Social Media

There is no standard way for measuring public attention on social media [47]. However, some previous studies have used tweet activity, such as the retweet frequency, to measure public attention [47,48]. Ripberger et al [48] studied public attention by assessing the number of tornado-related tweets posted on Twitter. Ripberger et al [48] assumed that the increase in the number of tornado warnings and number of tornado watches (issued by 2 different institutes) led to an increase in the number of tweets (representing public attention). Chew and Eysenbach [3] also indicated a coincidence between major H1N1 news and the frequency of H1N1 tweets in April 2009. The presence of a relationship between tweet activity and public attention has been shown in other studies about swine flu [49] or influenza [50].

None of the framing studies on COVID-19 [18,21,25-28] have identified the effect of frames on public attention. This study investigated the effect of COVID-19 frames in public discourse on Twitter on public attention. This study took a new approach that measures public attention more accurately than the methods used in previous studies because it measured not only tweet activity (ie, number of likes and retweets a frame receives) but also how many unique Twitter users have adopted a frame. In this new approach, public attention was measured by evaluating the score of public engagement with frames and the score of public adoption of frames (see Methods, Measuring Public Attention). Public engagement refers to the number of likes and retweets each frame has received. Public adoption is assessed by the number of unique users in each frame category [51].

Methods

Data Collection

Twitter was the primary source of data collection (ie, hashtags and tweets) in this study. The *get_trend* function from the *rtweet* package in the R software was used to collect trending hashtags related to COVID-19 on Twitter in March 2020. Additionally, the trending feature on the Twitter website was observed at least 3 times a day for the entire month of March 2020 to track and record trending hashtags. The trending hashtags identified by the *get_trend* function were exactly similar to those identified on Twitter. In March 2020, 22 trending hashtags related to COVID-19 were identified on Twitter. Additionally, the trending hashtags from each day were searched in the search box of the Twitter website to find other relevant hashtags and achieve a greater degree of reliability in data collection. For instance, by searching #QuarantineLife, a trending hashtag on March 16, 2020, additional hashtags, including #QuarantineDogs and #QuarantineCats, were identified.

This study then collected the tweets that contained at least one of the trending or associated hashtags. The *search_tweets* function in the *rtweet* package was used to collect tweets by the R software and the streaming application programming interface (API). R, Salesforce Social Studio, and Brandwatch were used to collect 694,582 tweets written in English from March 1, 2020 to March 31, 2020.

This study removed the retweets in the dataset to analyze initial tweets posted by people on Twitter. Additionally, quote tweets (ie, the retweets with comments), representing the original comments posted by Twitter users, were analyzed.

Bot Account Removal

Bots are automated accounts that can manipulate and impact public attention on social media [52]. Ferrara [52] indicated that bots were active during the COVID-19 pandemic and found that hashtags like #bitcoin, #smartnew, and #grreatwakening were mostly posted by bot accounts on Twitter. This study removed 269,854 (269,854/694,582, 38.85%) tweets posted by bot accounts to increase the validity of results, leaving 424,728 (424,728/694,582, 61.15%) tweets in the dataset.

The default model of the *tweetbotornot* package in R software was used to remove bot accounts [53]. The default model is 93.53% accurate in classifying bots and 95.32% accurate in classifying nonbots. It uses user-level features (eg, bio, location, number of followers and friends) and tweet-level features (eg, number of hashtags, mentions, capital letters) of Twitter accounts to identify bot accounts [53]. Accounts that received a score of at least 50% or a probability of 0.5 were considered bots and were removed [54].

Thematic Analysis to Identify Frames

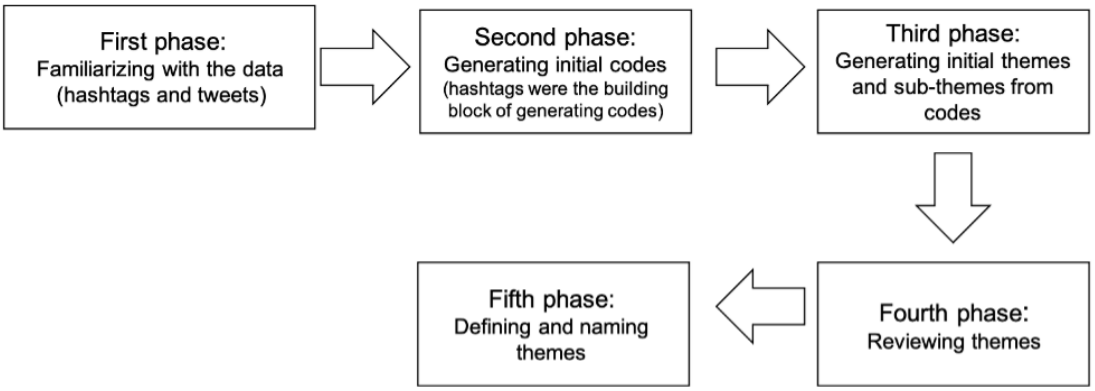
Hashtags were the main unit of analysis in this research and the building blocks of identifying frames. The frames identified in this study are primarily informed by the classification of trending and associated hashtags to different categories. Additionally, to identify the frames about COVID-19 in the public discourse on Twitter, thematic analysis was used. Thematic analysis is a method that provides a detailed description of the textual data [55]. In this research, public opinion includes all Twitter users such as workers, doctors, officials, journalists, politicians, teachers, and any other group on Twitter.

In thematic analysis, the judgment of the researcher or analyst is adequate for identifying themes [55]. The researcher analyzes the text and identifies themes within the “surface meanings of the data” without looking for anything beyond the text [55].

The thematic analysis helped us provide a more detailed and nuanced description of hashtags and find repeated patterns of meaning [55]. The contents of up to 100 randomly selected tweets were checked to ensure alignment between the hashtags and the contents of tweets. For instance, #WashYourHands, #HandWashing, #SafeHand, #SafeHands, and #HandWashChallenge were classified as one category named “hand hygiene.” Using thematic analysis, this study analyzed the content of the tweets associated with these hashtags to ensure they were related to hand hygiene. The thematic analysis also helped provide a more accurate description of this category. In some cases, the thematic analysis helped us redefine the theme assigned with a set of grouped hashtags.

We followed 5 phases to analyze the themes in the tweets. Figure 1 presents the workflow visualization for the 5 phases of the thematic analysis. In the first phase of thematic analysis, researchers familiarized themselves with the data [56] by reading the hashtags and the tweets associated with them several times before coding. It helped researchers obtain some knowledge of the hashtags, related tweets, and the initial list of ideas and themes.

Figure 1. Workflow visualization for the 5 phases of the thematic analysis.



In the second phase, initial codes were generated from trending and associated hashtags [56]. Hashtags were used to describe what a particular tweet was about and to create codes. For instance, #ToiletPaperPanic is a hashtag in the data referring to the public's panic for the possible shortage of toilet paper during the early days of the COVID-19 pandemic. Additionally, #WashYourHands is a hashtag that emphasizes the importance of hand hygiene to prevent contracting the COVID-19 virus.

In the third phase, codes were analyzed and classified into potential themes and subthemes [56]. For instance, the following set of hashtags — #SocialDistancing, #SocialDistancing, #KeepYourDistance, #PhysicalDistancing, #SocialDistance, #SocialDistancingNow, #YouAreTooCloseIf, #Dont BeASpreader, and #StoptheSpread — and the tweets associated with them helped the researchers to identify and propose the "social distancing" theme.

In phase 4, the researchers reviewed the themes and, when necessary, combined, refined, or separated them [56] to generate the final themes. In the fifth phase, the themes were clustered into frames. Concise names were assigned to the themes to help readers comprehend the theme's meaning [56].

Measuring Public Attention

This study used the scores of public engagement with and public adoption of frames to measure public attention to frames on Twitter. Following DeMasi et al [51], this study defined public engagement as the extent to which users engage with a frame. Public engagement was measured by the number of times a given frame had received at least one like or retweet divided by the frequency of that frame in the entire dataset. The lowest public engagement score that each frame could receive was 0, with a maximum of 1.

Public adoption refers to how broadly Twitter users adopt a frame. High adoption indicates a diverse community with many unique users who have posted tweets with hashtags in each frame. Low diversity shows a tight community with only a few users posting tweets with the hashtags in a given frame multiple times [51]. Public adoption was measured by the number of unique Twitter user accounts in a given frame divided by the frequency of the frame [51]. The public adoption score ranges from 0 to 1 as the minimum and maximum scores, respectively.

Results

Framing of COVID-19 on Twitter

We identified 9 categories of frames through thematic analysis (see Multimedia Appendix 2). Each frame is defined by a set of consequences of COVID-19 along with hashtags, characterizing what each frame category describes. The frames are "public health guidelines," "quarantine life," "solidarity," "evidence and facts," "call for action," "politics," "post-pandemic life," "shortage panic," and "conflict."

Public Health Guidelines

Public health guidelines consisted of 5 main themes: isolation, social distancing, hand hygiene, face hygiene, preventive tips, and awareness. Isolation included hashtags that encourage and advise people to isolate themselves to prevent the spread of the virus. Some of the hashtags in this theme, such as #StayHomeStayHealthy or #StayHomeSaveLives, positively motivate people to self-isolate by staying home. This frame also included the recommendations people made to call others' attention to the precautionary measures required to reduce the chance of being infected by the virus, such as practicing social distancing and face and hand hygiene (eg, #SocialDistancing, #WashYourHands, and #WearAMask). Sometimes tweets contained hashtags with a sense of humor and positive sentiments, such as #YouAreTooCloseIf or #How toKeepPeopleHome.

Other hashtags in this frame, such as #CoronaVirusTips, were used to provide guidelines and tips about various aspects of the virus, including how to clean smartphones, wash hands, and do contactless delivery of food. Additionally, this frame included tips for preventing potential exposure to the virus. For instance, #CoronaVirusPrevention was used to show how police departments should prevent potential disclosure and provide coronavirus prevention tips for people with disabilities.

Quarantine Life

This frame contained several themes that focused on people's daily lives during the quarantine. People used various hashtags to share stories, pictures, or videos of their home offices, the homeschooling of kids, and their pets. Pet owners used hashtags such as #QuarantineCats or #QuarantineDogs to share humorous content about their living experiences in isolation. #After3WeeksWithMyFamily and #SideEffectsofQuarantineLife

were used to post amusing messages along with photos or videos, often to make fun of the lived experience and actual difficulties people were dealing with during the pandemic. Quarantine life also focused on the impact of the pandemic on how people worked and learned during the quarantine. People posted videos or pictures of themselves using hashtags such as #RemoteLearning, #DistanceLearning, #RemoteWorking, or #OnlineLearning while working or learning from home.

Solidarity

Solidarity focused on inspiring, encouraging, and giving hope to each other; social support; providing voluntary services to communities; acknowledging health care professionals; and emphasizing unity. People used hashtags such as #StayStrong or #TogetherApart to inspire "staying home together" during the pandemic. The hashtags #CoronaWarriors, #ClapForCarers, #HealthCareHeroes, and #ClapForOurCarers were used to show appreciation to the frontline workers in fighting the pandemic, namely health care professionals.

A major theme in solidarity was social support, which consisted of hashtags used to assist various (vulnerable) populations or services. For instance, #ProtectOurSeniors was used to emphasize the importance of supporting seniors medically, financially, and emotionally. Other hashtags such as #OpenForTakeout, #OpenForDelivery, #SupportSmallBusiness, and #SupportLocal were used to support businesses and acknowledge the stores open for takeout or delivery during the pandemic.

Call for Action

This frame refers to a lack of accountability and responsibility in governments for not taking necessary actions to close public places. This frame also consisted of hashtags that show that the public is also responsible for their lack of compliance with public health guidelines. For instance, #CloseTheBeaches was used in tweets to emphasize that a public space like beaches should be closed as many people went to Florida's beaches during the pandemic.

Evidence and Facts

This category focused on updates, reality, and truths about various aspects of the COVID-19 virus, such as the total number of deaths, new cases, discharged patients, and further evidence about how the virus spreads. For instance, #CoronaVirusUpdates was used on March 29, 2020 to refer to emerging hotspots of coronavirus and emphasized that New York remained the worst-hit US state. Another hashtag in this category was #CoronaVirusTruth, which was used to provide people with the facts and truths about COVID-19, such as the increasing number of deaths across the globe.

Politics

This frame refers to the accountability of politicians for their national policies and actions during the pandemic. Some hashtags were used to hold political officials accountable for their policies, decisions, and actions regarding the COVID-19 pandemic. The tweets in this category often had a negative sentiment. For instance, #TrumpLiedPeopleDied was used to show a lack of transparency and timely action to control the virus, not taking the pandemic seriously, and not following public health guidelines and policies. #ChinaLiedPeopleDied was, on the other hand, used to make China responsible for the transmission of the COVID-19 virus to other countries.

Post-Pandemic Life

This frame focused on the positive sentiment surrounding people's plans, feelings, and life after the pandemic is over. #WhenCoronaVirusIsOver and #WhenILeaveMyHouseAgain are the hashtags in this category that sometimes were posted with tweets with a sense of humor. For example, a tweet stated, "people are going to stay home even when the pandemic is over." People used the hashtags in this category to refer to the food they like to eat; to describe the places or people they enjoy visiting; to recommend things people should avoid, like eating animals; to emphasize that the world would be a cleaner place after the pandemic; and to suggest that people would love and respect nature more than in the past.

Shortage Panic

This frame showed panic about the shortage of products as a top priority of people when the pandemic had just started in March 2020. Shortage panic refers to the public reaction to and anxiety about the shortage of resources, including bottled water, toilet paper, hand sanitizer, and food. Most tweets in this category contained an image showing empty shelves in stores or people lined up in stores trying to purchase products such as toilet paper.

This frame demonstrated uncertainty in the public about how much supplies were needed because the future was unpredictable [57]. The uncertainty about the pandemic led to an "exaggerated sense of urgency and a fear of scarcity," which resulted in actual scarcity [57].

Conflict

Conflicts refers to arguments and disagreements among people. Hashtags such as #FilmYourHospital or #EmptyHospitals were used to imply that COVID-19 was not as severe as stated by the media, governments, or health officials. Some people used these hashtags to frame their disapproval of the lockdown and social distancing or to suggest that COVID-19 is not real by filming empty hospitals. Another hashtag in this category was #FakeNews, which people used to share opposing views about various aspects of COVID-19. It was also used to show disapproval with the information and news spread about the virus, mostly by the news media or politicians from both the Democratic and Republican parties. #FakeNews was sometimes used to argue in favor of issues related to COVID-19, and other times, it was used to argue against those issues (see [19]).

Public Attention to Frames

Public attention to frames was assessed using the 2 measures of "public engagement" and "public adoption." Public engagement shows how engaging a frame has been on Twitter, counted by the number of retweets and likes a frame had received. Public adoption refers to the number of unique Twitter user accounts in a given frame category, divided by the frequency of frames in that frame category [58].

Table 1 shows public engagement with each frame. "Call for action" had the highest engagement (0.68), followed by "conflict" (0.67) and "evidence and facts" (0.59). Table 2 shows the public adoption score for each frame. "Post-pandemic life" (0.99) had the highest adoption score, followed by "call for action" (0.92) and "shortage panic" (0.91).

Table 1. Public engagement with frames.

Frame number	Frame	Frequency of the frame, n	Labeled Yes[a], n	Labeled No[a], n	Public engagement score (Yes/Yes+No)
1	Public health guidelines	171,009	94,545	76,464	0.55
2	Quarantine life	86,983	48,701	38,282	0.56
3	Solidarity	46,164	25,253	20,911	0.54
4	Evidence and facts	31,819	18,874	12,945	0.59
5	Call for action	31,809	21,881	9928	0.68
6	Politics	24,346	11,903	12,443	0.49
7	Post-pandemic life	20,442	11,439	9003	0.56
8	Shortage panic	16,815	9321	7494	0.55
9	Conflict	3550	2384	1166	0.67

[a]Each frame receiving at least a retweet or like was labeled as "Yes"; otherwise, it was labeled as "No."

Table 2. Public adoption of frames.

Frame number	Frame	Frequency of unique users (U), n	Frequency of frame (F), n	Public adoption score (U/F)
1	Public health guidelines	142,105	171,009	0.83
2	Quarantine life	75,439	86,983	0.87
3	Solidarity	37,424	46,164	0.80
4	Call for action	29,134	31,809	0.92
5	Evidence and facts	27,957	31,819	0.88
6	Politics	21,724	24,346	0.89
7	Post-pandemic life	20,149	20,442	0.99
8	Shortage panic	15,255	16,815	0.91
9	Conflict	2095	3550	0.60

The distribution of public engagement and public adoption is illustrated in Figure 2. According to this figure, although conflict had the lowest frequency (n=3550), it received the second-highest engagement score (0.67). Public health guidelines had the highest frequency (n=171,009), followed by quarantine life (n=86,983); however, their public adoption and engagement scores were not high. Pearson correlation did not show any correlation between public adoption and public engagement: r_8=–0.46, P=.18.

Figure 3 presents the distribution of public engagement with and public adoption of frames in terms of frame frequency. No meaningful patterns can be observed in frame engagement and adoption based on frequency. Pearson correlation showed no correlation between engagement and frequency (r_8=–0.26, P=.46) nor between adoption and frequency (r_8=0.13, P=.70). The insignificant results could mean that, if a frame receives high adoption by the people, it does not indicate that the frame would also achieve a high engagement (or vice versa).

Figure 2. Distribution of the frames "public engagement" and "public adoption".

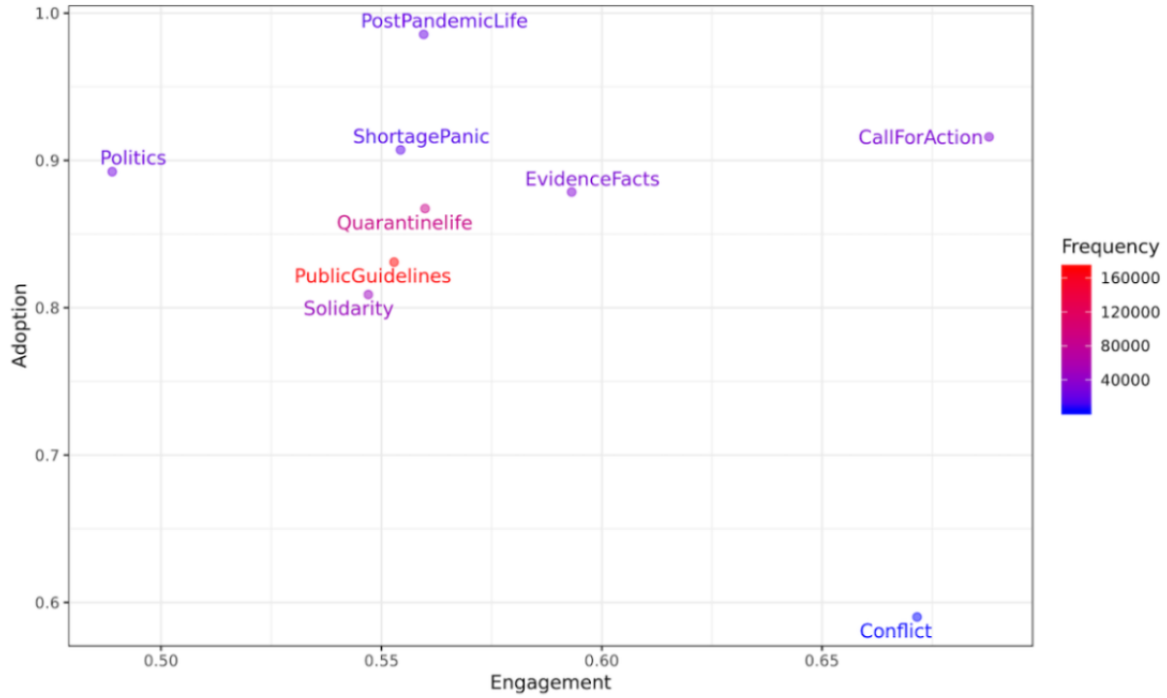
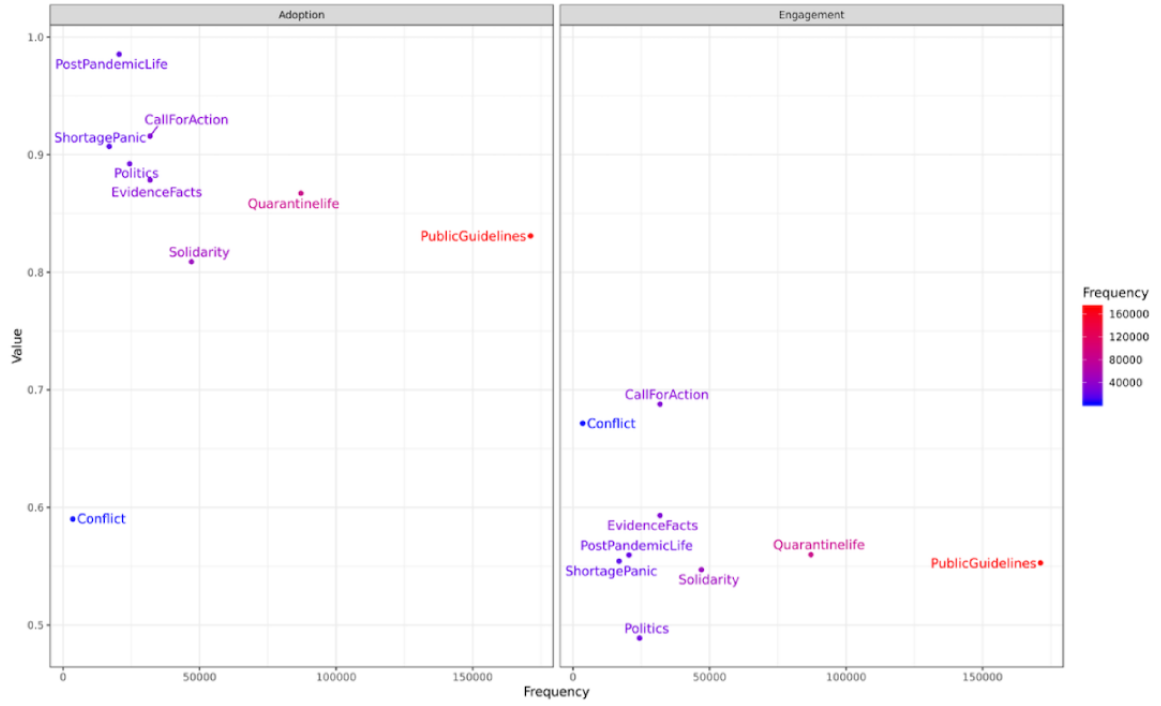


Figure 3. Distribution of the frames "public adoption" and "public engagement" in terms of frequency.



Discussion

Principal Findings

This study analyzed the trending and associated hashtags in March 2020 during the initial phases of the COVID-19 pandemic. We found 9 frames in the public discourse on Twitter: “public health guidelines,” “quarantine life,” “solidarity,” “evidence and facts,” “call for action,” “politics,” “post-pandemic life,” “shortage panic,” and “conflict.” These frames and the hashtags within them such as #SupportSmallBusinesses and #StayPositive can be used to identify the types of information that should be delivered during any public health crisis, specifically the initial phase of the crisis. Furthermore, some of the frames such as solidarity, evidence and facts, call for action, politics, shortage panic, and conflict can be used in any local or global crisis such as social movements, earthquakes, tornadoes, hurricanes, and floods. For instance, the solidarity frame can be used to design messages to encourage people during an earthquake to show solidarity to those impacted by the crisis.

This study also evaluated public attention to the identified frames by assessing the amount of public engagement with and public adoption of frames. Among all frames, "call for action" had the highest engagement score, followed by "conflict" and "evidence and facts." Additionally, "post-pandemic life" had the highest adoption score, followed by "call for action" and "shortage panic."

According to Price and Tewksbury [59], framing is more effective when it is relevant to people, such as when the frames resonate with the audience's beliefs or ideology [20]. The effects of frames on public opinion could be more substantial when the quality or logic of the argument and source credibility are more reliable [60]. The frames with greater public attention could be used strategically to design messages that affect public opinion more efficiently during public health emergencies. Additionally, these frames could be used to increase the performance of social media posts and encourage public compliance with public health messages.

Theoretical Contributions and Implications

This study informs framing theory and research in several ways, mentioned in the following sections.

Frames as the Manifestation of COVID-19 Consequences

This research informs framing theory by indicating that frame analysis of public discourse on social media via hashtags is valuable not only in understanding public opinion about various aspects of the pandemic but also in recognizing the consequences of the pandemic on people's lives, such as panic over the shortage of products.

The findings indicated that frames are the manifestation of COVID-19 consequences. The consequences discussed by people on Twitter about COVID-19 could be classified and analyzed based on whether (1) they were exclusively related to COVID-19, such as hand hygiene or isolation (ie, people are isolated or wash their hands because of COVID-19); (2) consequences could be related to any health crisis such as social support of vulnerable groups (ie, vulnerable groups require social support in any pandemic); and (3) consequences were generic irrespective of COVID-19, such as homeschooling or remote working (ie, some people homeschooled or worked remotely even before the pandemic).

Motivations and Sentiments in Frames

The findings showed that hashtags can promote intrinsic or extrinsic motivations among social media users. Hashtags can also convey a positive or negative sentiment. For instance, #StayHomeStayHealthy promotes an extrinsic motivation in people because this hashtag motivates people to stay home for external factors, such as prevention of contracting or spreading the virus. Another example could be #StayHomeStayHealthy, which has a positive sentiment because it indicates that staying healthy is a positive aspect of staying home.

Frames Promote the Collective Interest

Most hashtags promoted collective interest among people. For instance, #SocialDistancingNow was used by Twitter users to motivate people to prevent the potential transmission of COVID-19 for the public good. Another example could be #WearAMask in public, which promotes a collective interest to prevent the spread of the virus rather than merely preventing oneself from contracting the virus [61].

Conflict and Disagreement With Evidence and Facts

People used the "conflict" frame to discredit the evidence, facts, and updates about the COVID-19 virus, such as its fatality rates. Previous studies indicate that the news media sometimes use the "conflict" frame to report disagreements and arguments among people or groups to capture more audiences [42]. Using the conflict frame in social media posts can "increase the perceived seriousness and news value of an event" [62].

The conflict frame seems to be important in both the news media and public discourse during public health emergencies. Shih et al [14] stated that the conflict frame (defined as the arguments and disagreements among news sources) is the main frame in the media coverage of public health pandemics. Park et al [18] also found that conflict was a main frame in the Korean news media tweets about COVID-19. The popularity of the conflict frame demonstrates that there are always groups of people with disagreements about facts and evidence during public health emergencies.

This frame had the lowest frequency and lowest public adoption score among all frames identified in this study. A few previous studies have also indicated that news with conflict frames has been shared less on social media than news without conflict frames [62]. However, in this study, the conflict frame received the second-highest score for public engagement. This high score could be an indication that hashtags with conflicting information, such as #FilmYourHospital, are highly engaging on social media, despite their low frequency. A low adoption score with a high engagement score could indicate a focused community structure [51].

The "conflict" and "evidence and facts" frames indicate the presence of a chain of reactions, in that in response to the people who post information about facts and updates related to COVID-19 (ie, the evidence and facts frame), another group of people used the conflict frame to challenge those facts and updates (ie, the conflict frame). Some people would then, in turn, react to the conflict frame by sharing evidence and facts about the virus.

Vulnerable Groups

Our results indicated that frame analysis could be used to identify groups of vulnerable populations (eg, seniors, workers, and local and small businesses) and how the people frame their concerns and opinions about the needs of such populations on Twitter. It was found that vulnerable populations have been in need of prompt attention and support during the COVID-19 pandemic. For instance, seniors require social support, workers need safe workplaces, and local and small businesses need financial support. The findings did not show any hashtag referring to race or ethnic minority groups and health disparity populations, such as refugees, immigrants, homeless individuals, countries with fewer resources, or other marginalized communities. Overall, frame analysis through hashtags can be

used to identify the populations that need more attention and support during global crises.

Accountability of Gatekeepers

Frame analysis can also identify who or which groups are responsible for managing a public crisis in the public view. The "call to action" frame indicated that Twitter users held federal and local governments accountable for their lack of actions regarding the closure of public places like beaches or restaurants. The "politics" frame also demonstrated people's reaction to politicians' decisions, policies, and insufficient actions during the pandemic. The politics frame includes hashtags that show those responsible for the situation in the public view. For instance, many hashtags in this category were used to make the US President accountable for the increasing number of COVID-19 cases and deaths. This frame could also show a lack of trust between politicians and the public, which often occurs during the initial phases of a pandemic [63].

The Use of Humor in Public Health Messages

The results in this study inform framing research by showing that people used hashtags such as #QuarantineCat or #HowToKeepPeopleHome to share information with humorous content. These hashtags were not directly related to COVID-19 but were generated and used by Twitter users due to COVID-19. Additionally, people used these hashtags to share their experiences of living in isolation. Some tweets with these hashtags were used to make fun out of the lived experience and actual difficulties people were dealing with during the pandemic. The use of humor in tweets has been previously reported in some past studies [64,65]. For instance, Kopper [65] noted that the use of humor in diplomatic tweets has a conflict-mediating role. In this study, the use of humor in public health tweets had a role in mitigating public anxiety or fear as the consequences of the pandemic. Perhaps people use humor in tweets to convey complex public health messages that are more attractive to audiences.

Although humorous hashtags may simply not seem appropriate for use in public health communication, they can make up part of a broader information-sharing strategy. For example, to communicate safety guidelines, social media posts could be labeled with humorous hashtags that are appealing. The safety-related messages may then be transferred to and adopted by the public through such hashtags. Additionally, practitioners, activists, organizations, and authorities can actively embed relevant and meaningful hashtags into their posts that are framed more generally within appealing concepts.

Practical Implications

Designing Health Messages

The frames found in this study can also facilitate the communication on social media of various aspects of COVID-19 or any future health crisis. They can be used as a starting point for designing effective messages to address people's needs and concerns during future public health crises [18]. For instance, based on the current research, it can be suggested that health messages should be communicated with empathy, as also proposed in the study by Hyland-Wood et al [66]. These frames can support the public in the expression of their feelings and opinions during global crises [21]. The frames can also be used to influence public opinion, behavior, and actions about critical and complex issues related to COVID-19 (see [33]).

Governments, practitioners, organizations, activists, and authorities can take advantage of hashtags to frame their messages to inform the public about the consequences of lack of adherence to public health guidelines. For instance, governments and public health officials need to share actionable guidelines with citizens during public crises and inform them about what is happening. These activities require understanding of how messages should be framed and presented so that risk communications are accepted and adopted by the public [67]. If health communication officers, political officials, and other authorities wish to encourage citizens to follow public health guidelines, stay optimistic, and be supportive during a health crisis, they can frame their messages using the hashtags identified in this research. For instance, a message by the WHO could be framed as follows: "#WeAreInThisTogether: #StayStrong at home and post a video or picture of your dogs or cats in #Quarantine. Use #QuarantineCats or #QuarantineDogs to help others see your posts."

Issues and Their Roots

The results indicated that framing can also be used to identify the roots of the issues discussed by people on social media during public health crises. For instance, framing could be used to understand when, how, and about which topics false information is being distributed. The results indicated that #FilmYourHospital was a hashtag used by people to frame disagreements with evidence and facts about the COVID-19 pandemic. Once such problems and their roots are discovered through frame analysis, officials can find solutions, which, in this case, would be to neutralize and thwart false information dissemination.

Frame analysis can be used to identify what people are panicked about during a public health crisis. Another issue discussed by people was the "shortage panic," referring to people's reactions to the shortage and unavailability of products in the early days of the pandemic in March 2020. This panic could possibly be due to uncertainty in the public about what will happen in the future, lack of real-time communication with the public, or lack of trust between people and government officials. The study by Naeem [68] also indicated that uncertainty about COVID-19 triggered people to buy extra necessary food items, which resulted in the buying panic during the COVID-19 pandemic.

Methodological Contributions

This research contributes to the methodology of framing research on social media in several ways, as explained in the following sections.

Accuracy in Data Collection

In this research, the trending and associated hashtags were used to collect tweets. Using hashtags can increase the accuracy of collecting relevant tweets. The hashtag search approach is more accurate than the keyword-based search approach used in some past studies, such as that by Jang and Hart [69], mainly because

people use hashtags to describe their tweets [70]. Jang and Hart [69] searched "climate change AND real" to find tweets associated with the "real frames" or searched "climate change AND action" to find tweets about "action frames." A major issue with the keyword-based search and analysis is that it may result in retrieving extraneous tweets or excluding relevant ones, as stated by Jang and Hart [69].

Using Trending Hashtags to Identify Frames

Unlike most previous framing studies on Twitter that have solely analyzed tweets to identify frames, this study analyzed trending and associated hashtags along with the tweets containing those hashtags to identify frames. Jang and Hart [69] stated that an effective approach for analyzing frames is "to identify unique components of public rhetoric that clearly represent single frames of a more complex issue." Hashtags are unique components that represent frames of complex issues on social media. Moy and Bosch [30] noted that frames provide "meaning about social phenomena through the highlighting and packaging of information." Hashtags can also be used to give meaning to and highlight social phenomena. In the Literature Review section, the rationale for using (trending) hashtags for identifying frames was discussed.

Measuring Public Attention on Twitter

There is no standard way for measuring public attention on social media [47]. This study contributes to the methodology of framing research by examining the amount of public attention to frames by evaluating public adoption of frames and public engagement with frames [51]. Rarely any previous framing research has used this approach to measure the effect of frames on public attention on social media. This approach can also be used to identify the framing effect of the news media on public attention on social media in future studies.

Study Limitations

The study limitations were mostly related to data collection, some of which were beyond our control. For instance, due to the Twitter API restrictions in collecting historical data, all tweets associated with the trending and associated hashtags could not be collected. We tried to mitigate this limitation by using 2 other software for collecting historical data, as explained in the Data Collection section.

Another limitation is that sampling tweets based on trending hashtags may not represent all Twitter communication. Hashtags may only represent a specific subset of Twitter communication. All Twitter users do not use hashtags in their online communications on social media [71,72].

As mentioned, to identify the frames, the trending hashtags on Twitter in March 2020 were collected. Collecting data in other months may lead to identifying different frames.

Additionally, this research may have missed collecting all trending hashtags about COVID-19 in March 2020 for several reasons. First, Twitter determines emerging and trending topics (topics that are popular now) based on who the users (researchers who collect the data) follow, their interests, their locations, and other criteria (see Twitter trends FAQs). Second, a hashtag may be trending for a couple of hours, a day, or in rare cases, more than a day. A hashtag that is trending today may not be trending tomorrow, but people may still use the hashtag for a short or long period of time. The hashtag may disappear temporarily after a few days, may gain attention again in the future, or may disappear permanently [36].

Future Research

This study answers important questions about the frames in public discourse during the initial phase of the COVID-19 pandemic, but also raises questions to investigate in future research. This research identified 9 frames, which provide useful orientation for future empirical and theoretical research that aims to investigate the frames on social media during global (health) crises. Future studies can expand these frames and apply them to other global crises.

This study did not investigate how different groups of people that create the public, like citizens, officials, researchers, journalists, and organizations, frame COVID-19 differently, which is worth studying in future studies.

Experimental designs should be designed to manipulate the type and category of frames (while keeping other variables constant) to understand to what degree hashtags affect public collective attention (see [73]) on social media.

Future research could also investigate (1) how frames and hashtags emerge, evolve, and operate and (2) how they succeed in achieving sustainability during different phases of a pandemic, such as (1) precrisis, (2) the initial event, (3) maintenance, (4) resolution, and (5) evolution [74].

Additionally, public compliance to health messages and guidelines is different than public attention to frames. Public compliance to COVID-19 can be evaluated by different measures such as "compliance to public health and social measures in preventing the spread of COVID-19" [75]. Future research should investigate how health messages should be designed and framed to increase public compliance to health guidelines during pandemics.

Conclusion

This is one of the first studies to use trending hashtags to analyze and identify frames on social media. It contributes to framing theory and research by showing that frames represent the consequences of a public health emergency such as COVID-19. Additionally, the findings inform framing theory and research by showing that the methodological advantages of using trending and associated hashtags lie not only in their ability to understanding the frames that are the focal point of the public but also in their potential to allow researchers to measure public attention to those frames. The results indicated that some frames are more appealing during a global pandemic than others, such as "call for action," therefore receiving greater public adoption and engagement. It was found that the frequency of a frame on social media does not necessarily mean greater public adoption of or engagement with the frame.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Framing of COVID-19 by the news media.
[DOCX File , 19 KB - jmir_v23i9e30800_app1.docx]

Multimedia Appendix 2
Framing of the COVID-19 pandemic in the public discourse on Twitter.
[DOCX File , 24 KB - jmir_v23i9e30800_app2.docx]

References

1. WHO Coronavirus (COVID-19) Dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2021-08-27]
2. Merchant RM, Lurie N. Social Media and Emergency Preparedness in Response to Novel Coronavirus. JAMA 2020 May 26;323(20):2011-2012. [doi: 10.1001/jama.2020.4469] [Medline: 32202611]
3. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
4. Vos SC, Buckner MM. Social Media Messages in an Emerging Health Crisis: Tweeting Bird Flu. J Health Commun 2016 Dec 31;21(3):301-308. [doi: 10.1080/10810730.2015.1064495] [Medline: 26192209]
5. Fu K, Liang H, Saroha N, Tse ZTH, Ip P, Fung IC. How people react to Zika virus outbreaks on Twitter? A computational content analysis. Am J Infect Control 2016 Dec 01;44(12):1700-1702. [doi: 10.1016/j.ajic.2016.04.253] [Medline: 27566874]
6. Burstein P. The Impact of Public Opinion on Public Policy: A Review and an Agenda. Political Research Quarterly 2016 Jul 02;56(1):29-40. [doi: 10.1177/106591290305600103]
7. Page BI, Shapiro RY. Effects of Public Opinion on Policy. Am Polit Sci Rev 2014 Aug 01;77(1):175-190. [doi: 10.2307/1956018]
8. Han X, Wang J, Zhang M, Wang X. Using Social Media to Mine and Analyze Public Opinion Related to COVID-19 in China. Int J Environ Res Public Health 2020 Apr 17;17(8):2788 [FREE Full text] [doi: 10.3390/ijerph17082788] [Medline: 32316647]
9. Avery EJ. Public information officers' social media monitoring during the Zika virus crisis, a global health threat surrounded by public uncertainty. Public Relations Review 2017 Sep;43(3):468-476. [doi: 10.1016/j.pubrev.2017.02.018]
10. Hannay A. On the Public. Milton Park, UK: Routledge; 2005.
11. Russell Neuman W, Guggenheim L, Mo Jang S, Bae SY. The Dynamics of Public Attention: Agenda-Setting Theory Meets Big Data. J Commun 2014 Mar 26;64(2):193-214. [doi: 10.1111/jcom.12088]
12. Goffman E. Frame Analysis: An Essay on the Organization of Experience. Cambridge, MA: Harvard University Press; 1974.
13. Entman R. Framing: Toward clarification of a fractured paradigm. Journal of Communication 1993:51. [doi: 10.1111/j.1460-2466.1993.tb01304.x]
14. Shih T, Wijaya R, Brossard D. Media Coverage of Public Health Epidemics: Linking Framing and Issue Attention Cycle Toward an Integrated Theory of Print News Coverage of Epidemics. Mass Communication and Society 2008 Apr 07;11(2):141-160. [doi: 10.1080/15205430701668121]
15. Chong D, Druckman JN. Framing Theory. Annu. Rev. Polit. Sci 2007 Jun;10(1):103-126. [doi: 10.1146/annurev.polisci.10.072805.103054]
16. Kaufman S, Elliott M, Shmueli D. Frames, framing and reframing. Beyond intractability. 2013. URL: http://rachelgrantcoaching.com/course/03-frames-and-brains.pdf [accessed 2021-08-27]
17. Meraz S, Papacharissi Z. Networked Gatekeeping and Networked Framing on #Egypt. The International Journal of Press/Politics 2013 Jan 27;18(2):138-166. [doi: 10.1177/1940161212474472]
18. Park HW, Park S, Chong M. Conversations and Medical News Frames on Twitter: Infodemiological Study on COVID-19 in South Korea. J Med Internet Res 2020 May 05;22(5):e18897 [FREE Full text] [doi: 10.2196/18897] [Medline: 32325426]
19. Van Gorp B. Strategies to Take Subjectivity Out of Framing Analysis. In: D'Angelo P, Kuypers JA, editors. Doing news framing analysis: Empirical and theoretical perspectives. New York, NY: Routledge; 2009.
20. Nisbet MC. Knowledge Into Action: Framing the Debates Over Climate Change and Poverty. In: D'Angelo P, Kuypers JA, editors. Doing news framing analysis: Empirical and theoretical perspectives. New York, NY: Routledge; 2010.
21. Wicke P, Bolognesi MM. Framing COVID-19: How we conceptualize and discuss the pandemic on Twitter. PLoS One 2020 Sep 30;15(9):e0240010 [FREE Full text] [doi: 10.1371/journal.pone.0240010] [Medline: 32997720]
22. Stoltz DS, Taylor MA. Paying with change: The purposeful enunciation of material culture. Poetics 2017 Oct;64:26-39 p 31. [doi: 10.1016/j.poetic.2017.07.003]
23. Krishnamurthy P, Carter P, Blair E. Attribute Framing and Goal Framing Effects in Health Decisions. Organ Behav Hum Decis Process 2001 Jul;85(2):382-399. [doi: 10.1006/obhd.2001.2962] [Medline: 11461207]

24. Almashat S, Ayotte B, Edelstein B, Margrett J. Framing effect debiasing in medical decision making. Patient Educ Couns 2008 Apr;71(1):102-107. [doi: 10.1016/j.pec.2007.11.004] [Medline: 18164168]
25. Yu J, Lu Y, Muñoz-Justicia J. Analyzing Spanish News Frames on Twitter during COVID-19-A Network Study of El País and El Mundo. Int J Environ Res Public Health 2020 Jul 28;17(15):5414 [FREE Full text] [doi: 10.3390/ijerph17155414] [Medline: 32731359]
26. Ogbodo JN, Onwe EC, Chukwu J, Nwasum CJ, Nwakpu ES, Nwankwo SU, et al. Communicating health crisis: a content analysis of global media framing of COVID-19. Health Promot Perspect 2020 Jul 12;10(3):257-269 [FREE Full text] [doi: 10.34172/hpp.2020.40] [Medline: 32802763]
27. Poirier W, Ouellet C, Rancourt M, Béchard J, Dufresne Y. (Un)Covering the COVID-19 Pandemic: Framing Analysis of the Crisis in Canada. Can J Pol Sci 2020 Apr 29;53(2):365-371. [doi: 10.1017/s0008423920000372]
28. Benziman Y. "Winning" the "battle" and "beating" the COVID-19 "enemy": Leaders' use of war frames to define the pandemic. Peace and Conflict: Journal of Peace Psychology 2020 Aug;26(3):247-256. [doi: 10.1037/pac0000494]
29. Reese SD, Gandy JO, Grant AE, editors. Framing public life: Perspectives on media and our understanding of the social world. New York, NY: Routledge; 2001.
30. Moy P, Bosch B. Theories of public opinion. In: Cobley P, Schultz PJ, editors. Handbook of Communication Science. Vol. 1: Theories and Models of Communication. Berlin, Germany: De Gruyter Mouton; 2013:289-308.
31. Gamson WA, Modigliani A. Media Discourse and Public Opinion on Nuclear Power: A Constructionist Approach. American Journal of Sociology 1989 Jul;95(1):1-37. [doi: 10.1086/229213]
32. Hellmueller L, Zhang X. Shifting toward a humanized perspective? Visual framing analysis of the coverage of refugees on and before and after the iconic photo publication of Alan Kurdi. Visual Communication 2019 Mar 18:147035721983279. [doi: 10.1177/1470357219832790]
33. Hemphill L, Culotta A, Heston M. Framing in Social Media: How the US Congress Uses Twitter Hashtags to Frame Political Issues. SSRN Journal 2013:2317335. [doi: 10.2139/ssrn.2317335]
34. Lehmann J, Gonçalves B, Ramasco JJ, Cattuto C. Dynamical classes of collective attention in twitter. 2012 Presented at: 21st international conference on World Wide Web; April 16-20, 2012; Lyon, France. [doi: 10.1145/2187836.2187871]
35. Rodriguez L, Dimitrova DV. The levels of visual framing. Journal of Visual Literacy 2016 Feb 29;30(1):48-65. [doi: 10.1080/23796529.2011.11674684]
36. Kwak H, Lee C, Park H, Moon S. What is Twitter, a social network or a news media? 2010 Presented at: 19th international conference on World wide web; April 26-30, 2010; Raleigh, NC. [doi: 10.1145/1772690.1772751]
37. Bruns A, Stieglitz S. Quantitative Approaches to Comparing Communication Patterns on Twitter. Journal of Technology in Human Services 2012 Jul;30(3-4):160-185. [doi: 10.1080/15228835.2012.744249]
38. Bruns A, Burgess J. The use of Twitter hashtags in the formation of ad hoc publics. Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference 2011 2011:1-9.
39. Asur S, Huberman BA, Szabo G, Wang C. Trends in Social Media: Persistence and Decay. SSRN Journal 2011:1. [doi: 10.2139/ssrn.1755748]
40. Schwalbe C, Dougherty S. Visual coverage of the 2006 Lebanon War: Framing conflict in three US news magazines. Media, War & Conflict 2015 Feb 26;8(1):141-162. [doi: 10.1177/1750635215571204]
41. van der Meer TG, Verhoeven P. Public framing organizational crisis situations: Social media versus news media. Public Relations Review 2013 Sep;39(3):229-231. [doi: 10.1016/j.pubrev.2012.12.001]
42. Semetko H, Valkenburg P. Framing European politics: A content analysis of press and television news. Journal of Communication 2000;50(2):93-109. [doi: 10.1111/j.1460-2466.2000.tb02843.x]
43. Matthes J, Kohring M. The content analysis of media frames: Toward improving reliability and validity. Journal of Communication 2008;58(2):258-279. [doi: 10.1111/j.1460-2466.2008.00384.x]
44. Zappavigna M. Searchable talk: the linguistic functions of hashtags. Social Semiotics 2015 Jan 09;25(3):274-291. [doi: 10.1080/10350330.2014.996948]
45. Trant J. Studying social tagging and folksonomy: A review and framework. Journal of Digital Information 2009;10(1):1 [FREE Full text]
46. Kehoe A, Gee M. Social Tagging: A new perspective on textual "aboutness". Studies in Variation, Contacts, and Change in English: Methodological and Historical Dimensions of Corpus Linguistics. 2011. URL: https://varieng.helsinki.fi/series/volumes/06/kehoe_gee/ [accessed 2021-08-27]
47. Silver A, Andrey J. Public attention to extreme weather as reflected by social media activity. J Contingencies and Crisis Management 2019 May 03;27(4):346-358. [doi: 10.1111/1468-5973.12265]
48. Ripberger J, Jenkins-Smith H, Silva C, Carlson D, Henderson M. Do tweets provide a valid indicator of public attention to severe weather risk communication? Weather, Climate, and Society 2014;6(4):520-530 [FREE Full text] [doi: 10.1175/WCAS-D-13-00028.1]
49. Szomszor M, Kostkova P, de Quincey E. #Swineflu: Twitter Predicts Swine Flu Outbreak in 2009. In: Szomszor M, Kostkova P, editors. Electronic Healthcare. eHealth 2010. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 69. Berlin, Heidelberg: Springer; 2010:18-26.

50. Lampos V, Cristianini N. Tracking the flu pandemic by monitoring the social web. 2010 Presented at: 2nd International Workshop on Cognitive Information Processing; June 14-16, 2010; Elba, Italy URL: https://ieeexplore.ieee.org/document/5604088 [doi: 10.1109/cip.2010.5604088]
51. DeMasi O, Mason D, Ma J. Understanding Communities via Hashtag Engagement: A Clustering Based Approach. Understanding communities via hashtag engagement; 2016 Presented at: Tenth International AAAI Conference on Web and Social Media; May 17–20, 2016; Cologne, Germany.
52. Ferrara E. # covid-19 on twitter: Bots, conspiracies, and social media activism. Cornell University. 2020. URL: https://arxiv.org/abs/2004.09531v1 [accessed 2021-08-27]
53. mkearney / tweetbotornot. GitHub. 2019. URL: https://github.com/mkearney/tweetbotornot [accessed 2021-08-27]
54. Davis CA, Varol O, Ferrara E, Flammini A, Menczer F. BotOrNot: A System to Evaluate Social Bots. 2016 Presented at: 25th International Conference Companion on World Wide Web; April 11-15, 2016; Montréal, Québec, Canada. [doi: 10.1145/2872518.2889302]
55. Lindekilde L. Discourse and frame analysis: in-depth analysis of qualitative data in social movement research. In: della Porta D, editor. Methodological Practices in Social Movement Research. Oxford, England: Oxford University Press; 2014:195-227.
56. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
57. Kirby J. The economics of the toilet paper panic—and why more stockpiling is inevitable. Maclean's. 2020 Mar 10. URL: https://www.macleans.ca/economy/economicanalysis/the-toilet-paper-panic-and-why-more-stockpiling-is-inevitable/ [accessed 2021-08-27]
58. Stephenson P. Community engagement through hashtags. Perry Stephenson. 2018. URL: https://perry.stephenson.id/2018/08/25/community-engagement-through-hashtags/ [accessed 2021-08-27]
59. Price V, Tewksbury D. News values and public opinion: A theoretical account of media priming and framing. In: Barnett GA, Boster FJ, editors. Progress in communication sciences. New York, NY: Ablex; 1997:173-212.
60. Chong D, Druckman JN. Framing Public Opinion in Competitive Democracies. Am Polit Sci Rev 2007 Nov 08;101(4):637-655. [doi: 10.1017/s0003055407070554]
61. van der Linden C, Savoie J. Does Collective Interest or Self-Interest Motivate Mask Usage as a Preventive Measure Against COVID-19? Can J Pol Sci 2020 May 14;53(2):391-397. [doi: 10.1017/s0008423920000475]
62. Valenzuela S, Piña M, Ramírez J. Behavioral Effects of Framing on Social Media Users: How Conflict, Economic, Human Interest, and Morality Frames Drive News Sharing. J Commun 2017 Aug 28;67(5):803-826 p 802. [doi: 10.1111/jcom.12325]
63. Cairney P, Wellstead A. COVID-19: effective policymaking depends on trust in experts, politicians, and the public. Policy Design and Practice 2020 Oct 26:1-14. [doi: 10.1080/25741292.2020.1837466]
64. Jaiswal A, Mathur A, Mattu S. Automatic Humour Detection in Tweets using Soft Computing Paradigms. 2019 Presented at: International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon); February 14-16, 2019; Faridabad, India. [doi: 10.1109/comitcon.2019.8862259]
65. Kopper A. The use of humour in diplomatic tweets: The affiliative potential of ridicule. Cooperation and Conflict 2020 Nov 26:001083672097545. [doi: 10.1177/0010836720975458]
66. Hyland-Wood B, Gardner J, Leask J, Ecker UKH. Toward effective government communication strategies in the era of COVID-19. Humanit Soc Sci Commun 2021 Jan 27;8(1):1-11. [doi: 10.1057/s41599-020-00701-w]
67. Oppenheim B, Gallivan M, Madhav NK, Brown N, Serhiyenko V, Wolfe ND, et al. Assessing global preparedness for the next pandemic: development and application of an Epidemic Preparedness Index. BMJ Glob Health 2019 Jan 29;4(1):e001157 [FREE Full text] [doi: 10.1136/bmjgh-2018-001157] [Medline: 30775006]
68. Naeem M. Do social media platforms develop consumer panic buying during the fear of Covid-19 pandemic. Journal of Retailing and Consumer Services 2021 Jan;58:102226. [doi: 10.1016/j.jretconser.2020.102226]
69. Jang SM, Hart PS. Polarized frames on “climate change” and “global warming” across countries and states: Evidence from Twitter big data. Global Environmental Change 2015 May;32:11-17 p 13. [doi: 10.1016/j.gloenvcha.2015.02.010]
70. Potnis D, Tahamtan I. Hashtags for gatekeeping of information on social media. J Assoc Inf Sci Technol 2021 Mar 11:1-13 [FREE Full text] [doi: 10.1002/asi.24467]
71. Bruns A, Moe H. Structural layers of communication on Twitter. In: Bruns A, Mahrt W, Weller K, Burgess J, Puschmann C, editors. Twitter and society [Digital Formations, Volume 89]. Bern, Switzerland: Peter Lang Publishing; 2014:15-28.
72. Siapera E, Boudourides M, Lenis S, Suiter J. Refugees and Network Publics on Twitter: Networked Framing, Affect, and Capture. Social Media + Society 2018 Mar 26;4(1):205630511876443. [doi: 10.1177/2056305118764437]
73. Shteynberg G. Shared Attention. Perspect Psychol Sci 2015 Sep 17;10(5):579-590. [doi: 10.1177/1745691615589104] [Medline: 26385997]
74. Reynolds B, W Seeger M. Crisis and emergency risk communication as an integrative model. J Health Commun 2005 Feb 23;10(1):43-55. [doi: 10.1080/10810730590904571] [Medline: 15764443]
75. Padidar S, Liao S, Magagula S, Mahlaba TAM, Nhlabatsi NM, Lukas S. Assessment of early COVID-19 compliance to and challenges with public health and social prevention measures in the Kingdom of Eswatini, using an online survey. PLoS One 2021 Jun 29;16(6):e0253954 p 4 [FREE Full text] [doi: 10.1371/journal.pone.0253954] [Medline: 34185804]

Abbreviations

API: application programming interface
WHO: World Health Organization

Edited by C Basch; submitted 29.05.21; peer-reviewed by J Yu, A Eisenstadt; comments to author 19.06.21; revised version received 13.07.21; accepted 22.07.21; published 10.09.21.

Please cite as:
Tahamtan I, Potnis D, Mohammadi E, Miller LE, Singh V
Framing of and Attention to COVID-19 on Twitter: Thematic Analysis of Hashtags
J Med Internet Res 2021;23(9):e30800
URL: https://www.jmir.org/2021/9/e30800
doi:10.2196/30800
PMID:34406961

Original Paper

Using Telehealth to Deliver Primary Care to Adolescents During and After the COVID-19 Pandemic: National Survey Study of US Primary Care Professionals

Melissa B Gilkey[1], PhD; Wei Yi Kong[1], MA; Qian Huang[1], PhD; Brigid K Grabert[1], PhD, JD; Peyton Thompson[2], MScR, MD; Noel T Brewer[1], PhD

[1]Department of Health Behavior, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[2]Department of Pediatrics, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Corresponding Author:
Melissa B Gilkey, PhD
Department of Health Behavior
University of North Carolina at Chapel Hill
CB 7440
Chapel Hill, NC, 27599
United States
Phone: 1 9199668650
Email: gilkey@email.unc.edu

Abstract

Background: The COVID-19 pandemic has led to unprecedented use of telehealth, including by primary care professionals (PCPs) who serve adolescents.

Objective: To inform future practice and policies, we sought to characterize PCPs' recent experience using adolescent telehealth as well as their support for it after the COVID-19 pandemic is over.

Methods: From February to March 2021, we conducted a web-based survey of 1047 PCPs in the United States. Our national sample included physicians (747/1047, 71%), advanced practice providers (177/1047, 17%), and nurses (123/1047, 12%) who provided primary care to adolescents aged 11-17 years.

Results: Most PCPs reported using telehealth for a low, moderate, or high proportion of their adolescent patients in the three months prior to the survey (424/1047, 40%, 286/1047, 27%, and 219/1047, 21%, respectively); only 11% (118/1047) reported no use. A majority of respondents agreed that adolescent telehealth increases access to care (720/1047, 69%) and enables them to provide high-quality care (560/1047, 53%). Few believed that adolescent telehealth takes too much time (142/1047, 14%) or encourages health care overuse (157/1047, 15%). Most supported giving families the option of adolescent telehealth for primary care after the pandemic is over (683/1047, 65%) and believed that health insurance plans should continue to reimburse for telehealth visits (863/1047, 82%). Approximately two-thirds (702/1047, 67%) wanted to offer adolescent telehealth visits after the pandemic, with intentions being higher among those with recent telehealth experience ($P<.001$).

Conclusions: PCPs in our national sample reported widespread use of and predominantly positive attitudes toward adolescent telehealth. Our findings also suggest broad support among PCPs for continuing to offer adolescent telehealth after the COVID-19 pandemic ends.

(J Med Internet Res 2021;23(9):e31240) doi:10.2196/31240

KEYWORDS

adolescent health; primary care; telemedicine; health communication; health services; telehealth; adolescent; young adult; teenager; COVID-19; survey; policy; access

Introduction

The COVID-19 pandemic has rapidly transformed health care delivery in the United States; for the first time, telehealth is playing a central role in the delivery of primary care for adolescents. Research conducted prior to the pandemic suggests that telehealth is a promising mode of delivering discrete types of health care for adolescents, including in the areas of mental

health, asthma and diabetes management, and gender-affirming care [1-4]. A small body of research also documents the adolescent medicine community's impressive efforts to rapidly scale up telehealth programs in the first months of the pandemic [5-7]. However, to our knowledge, no published national studies have explored the experience of primary care professionals (PCPs) in delivering care to adolescents in the ensuing period, during which telehealth has presumably become a standard offering for many. To address this gap, we surveyed a national sample of PCPs. Our aims were to characterize PCPs' recent adolescent telehealth use and attitudes as well as their support for continuing to offer adolescent telehealth after the COVID-19 pandemic is over.

Methods

Participants and Procedures

We conducted a web-based survey of PCPs from February to March 2021. We contracted with a survey research company to administer the survey, which we developed, to a standing national panel. The company maintained the panel using a combination of recruitment methods, including web-based registration, referrals, marketing emails, and digital advertisements. As part of the recruitment process, physicians provided licensure information used to verify their identity. For our survey, eligible panel members were US physicians, advanced practitioners (ie, nurse practitioners and physician assistants), and nurses who provided primary care, including vaccinations, to adolescents aged 11-17 years. In compliance with state policies governing PCPs' survey participation, our sample excluded residents of Vermont.

The survey company emailed invitations and up to two reminders to panel members. A total of 1055 panel members responded by accessing the survey. The response rate was 61% among physicians and 41% among advanced practitioners and nurses (American Association for Public Opinion Research response rate 4) [8]. Participants provided informed consent and received up to US $80 for their participation, depending on market rates in their area. Based on survey responses, we excluded 8 PCPs who indicated that they saw no adolescent patients in a typical week, resulting in a final sample size of 1047 participants. The University of North Carolina Institutional Review Board approved the study protocol.

Measures

Our survey began with an introductory statement that defined adolescent telehealth as visits by videoconference or telephone for patients aged 11-17 years. The survey next assessed the extent of PCPs' recent telehealth use with a closed-ended question on the proportion of adolescent patients they saw by telehealth in the 3 months prior to the survey; we recategorized responses as high (51%-75%, 76%-99%, and 100%), moderate (26%-50%), low (1%-25%), or no (0%) use. Among PCPs with any (>0%) use, the survey used 6 closed-ended items to assess telehealth practice. Of these items, 4 used pre-specified lists to assess the type of care provided, perceived advantages, perceived disadvantages, and technological barriers. One item assessed which vaccines PCPs always recommended during telehealth visits for adolescents who were due: seasonal influenza; human papillomavirus (HPV); tetanus, diphtheria, and acellular pertussis (Tdap); and meningococcal ACWY. One item assessed how often PCPs requested confidential time to speak with adolescents during telehealth visits.

Our survey assessed PCPs' perceptions of adolescent telehealth with 7 closed-ended questions that used 5-point response scales, ranging from "strongly disagree" (1) to "strongly agree" (5). Of these items, 4 assessed PCPs' attitudes on whether adolescent telehealth increases access to care, is a way they can provide high-quality care, takes too much time, or encourages health care overuse. A total of 2 items assessed PCPs' support for adolescent telehealth after the COVID-19 pandemic is over in terms of whether families should still have the option to use telehealth for primary care visits and whether health insurance plans should continue to reimburse for visits. One item assessed PCPs' intentions in terms of whether they wanted to offer adolescent telehealth visits once the pandemic is over. For all 7 items, we recategorized responses as disagree ("somewhat" or "strongly"), neither agree or disagree, or agree ("somewhat" or "strongly").

Our survey assessed PCPs' demographic and professional characteristics, including their training, gender, race, number of years in practice, and number of adolescent patients seen in a typical week (Table 1). The survey also assessed characteristics of the clinics in which the PCPs worked. These measures included clinic specialty (family medicine or pediatrics), practice type (solo/group vs other), whether the clinic was part of a health care system or network, the rurality of the area the clinic served, the US census region in which the clinic was located, and the extent to which the clinic experienced financial strain due to the COVID-19 pandemic.

Table 1. Sample characteristics (N=1047).

Characteristic	Value, n (%)
Primary care professional characteristics	
Training	
Physician	747 (71)
Advanced practice provider[a]	177 (17)
Nurse	123 (12)
Gender	
Woman	515 (49)
Man	492 (47)
Other[b]	40 (4)
Race	
White	717 (68)
Black	41 (4)
Asian	170 (16)
Other	119 (11)
Years in practice	
0-9	252 (24)
10-19	395 (38)
20 or more	400 (38)
Adolescent patients seen in a typical week	
1-9	283 (27)
10-24	431 (41)
25 or more	333 (32)
Proportion of adolescents seen by telehealth in prior 3 months	
None (0%)	118 (11)
Low (1%-24%)	424 (40)
Moderate (25%-50%)	286 (27)
High (51%-100%)	219 (21)
Clinic or practice characteristics	
Specialty	
Family medicine	748 (71)
Pediatrics	299 (29)
Practice type	
Solo or group	696 (66)
Other[c]	351 (34)
Part of a health care system	
No	457 (44)
Yes	590 (56)
Rurality	
Urban	363 (35)
Suburban	525 (50)
Rural	159 (15)

Characteristic	Value, n (%)
Region	
Northeast	265 (25)
Midwest	247 (24)
South	333 (32)
West	202 (19)
COVID-19–related financial strain	
None or a little	360 (34)
Moderate to high	687 (66)

[a]Includes nurse practitioners and physician assistants.

[b]Includes neither woman nor man, prefer to self-describe, and prefer not to say.

[c]Includes hospital- and university-affiliated clinics, Federally Qualified Health Centers, and community, public health, and nonprofit clinics.

Statistical Analysis

We used Pearson chi-square tests to compare the proportions of PCPs who indicated that they always recommended the seasonal influenza vaccine versus each of the other vaccines (HPV, Tdap, meningococcal) during adolescent telehealth visits. We compared the number of advantages and disadvantages that PCPs endorsed for adolescent telehealth using a Wilcoxon signed-rank test. We used bivariate logistic regression to identify correlates of PCPs' intentions to offer adolescent telehealth visits once the COVID-19 pandemic is over, modeling the outcome as yes ("agree") versus no ("neither agree or disagree" and "disagree"). We then simultaneously entered statistically significant correlates into a multivariable model. We conducted analyses using Stata, version 15.1 (StataCorp LLC). Statistical tests were two-tailed with a critical alpha of .05.

Results

Participant Characteristics

Our sample of 1047 respondents comprised physicians (n=747, 71%), advanced practitioners (n=177, 17%), and nurses (n=123, 12%; Table 1). Most had 10 or more years of experience in practice (795/1047, 76%) and saw 10 or more adolescent patients in a typical week (764/1047, 73%). PCPs worked in clinics focusing on family medicine (748/1047, 71%) or pediatrics (299/1047, 29%). The clinics were located in all four US regions; some were in clinical systems or networks (590/1047, 56%), and some served rural areas (159/1047, 15%). Two-thirds (687/1047, 66%) of PCPs reported that their clinics had experienced moderate to high financial strain due to the COVID-19 pandemic.

Telehealth Practice

Almost all PCPs reported using telehealth to see adolescent patients in the 3 months prior to the survey (Table 1). Approximately one-fifth of our sample (219/1047, 21%) indicated high adolescent telehealth use, while others reported more moderate (286/1047, 27%) or low (424/1047, 40%) use. Only 11% of PCPs (118/1047) reported no recent adolescent telehealth use.

The 929 PCPs with recent telehealth visits most often used telehealth for chronic disease management (n=599, 64%), acute care (n=571, 61%), mental and behavioral health (n=561, 60%), or vaccine consultations (n=406, 44%) (Table 2). Most of these 929 PCPs indicated that they always recommended seasonal influenza vaccination during telehealth visits if adolescents were due (n=798, 86%), but somewhat fewer said the same for HPV, Tdap, and meningococcal vaccines (n=715, 77%, n=709, 76%, and n=612, 66%, respectively; all $P<.001$). Approximately one-quarter of these PCPs (263/929, 28%) reported that they "always" or "often" requested time to speak to adolescents confidentially during telehealth visits.

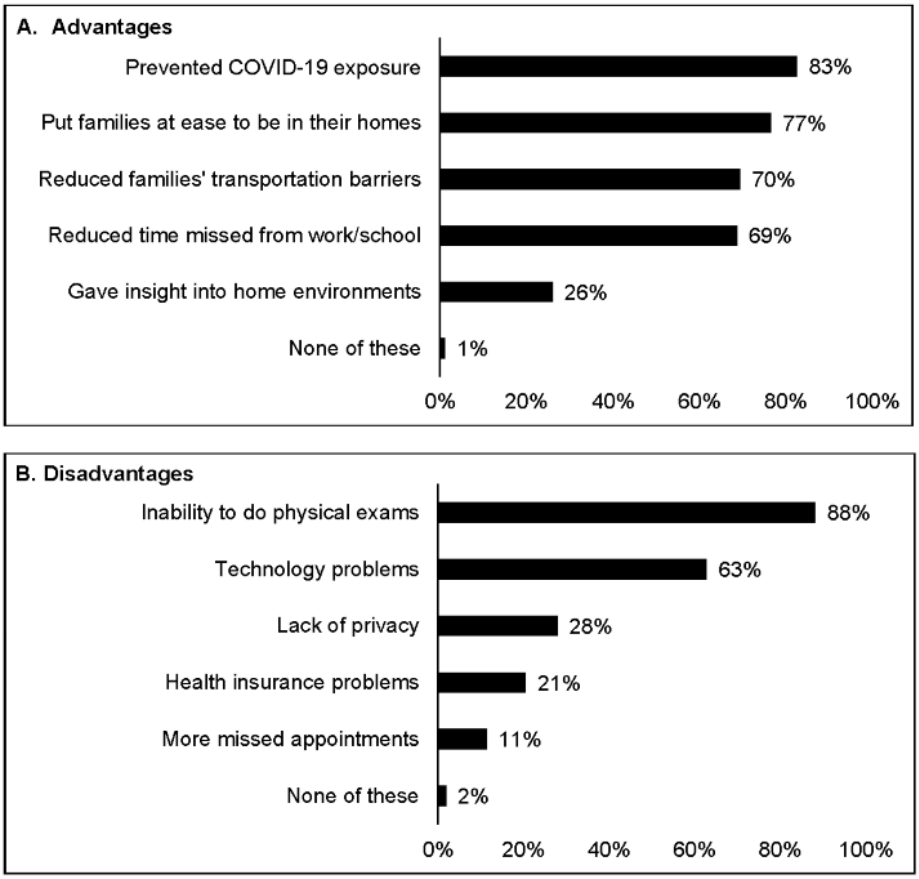
Table 2. Telehealth practice among primary care professionals with recent adolescent telehealth visits (n=929).

	Value, n (%)
Type of adolescent telehealth visits provided	
Chronic disease management	599 (64)
Acute care	571 (61)
Mental and behavioral health	561 (60)
Vaccine consultation	406 (44)
Other well-child care	387 (42)
Sexual health and contraceptive counseling	359 (39)
None of these	14 (2)
Vaccines always recommended if due	
Seasonal influenza	798 (86)
Human papillomavirus	715 (77)
Tetanus, diphtheria, and acellular pertussis	709 (76)
Meningococcal	612 (66)
None of these	48 (5)
Request to speak with adolescent confidentially	
Always or often	263 (28)
Sometimes	356 (38)
Rarely or never	310 (33)
Technology problems[a]	
Poor-quality internet connections	499 (85)
Families' lack of internet-enabled devices	353 (60)
Lack of training for providers and staff	195 (33)
Difficulty working with medical interpreters	128 (22)
None of these	21 (4)

[a]Among the subset of participants who reported technology problems as a disadvantage (n=584).

On average, PCPs reported more advantages than disadvantages of their recent adolescent telehealth use (mean=3.2 of 5 advantages, SD 1.3, vs 2.1 of 5 disadvantages, SD 0.9; $P<.001$). The most common advantages were preventing COVID-19 exposure (768/929, 83%), putting families at ease (712/929, 77%), and reducing families' transportation (646/929, 70%) or time (639/929, 69%) burdens (Figure 1). Only approximately one-quarter of PCPs reported that gaining insight into families' home environments was an advantage (243/929, 26%). The most common disadvantages were the inability to perform physical examinations (820/929, 88%) and technology problems (584/929, 63%). Only a minority of PCPs indicated that a lack of privacy (260/929, 28%), health insurance problems (191/929, 21%), or more missed appointments (106/929, 11%) were disadvantages. Of the 584 PCPs who indicated technology problems as a disadvantage, most reported that poor quality internet connections (n=499, 85%) and families' lack of internet-enabled devices (n=353, 60%) were common barriers (Table 2).

Figure 1. Perceived advantages (A) and disadvantages (B) of adolescent telehealth by primary care providers (n=929).

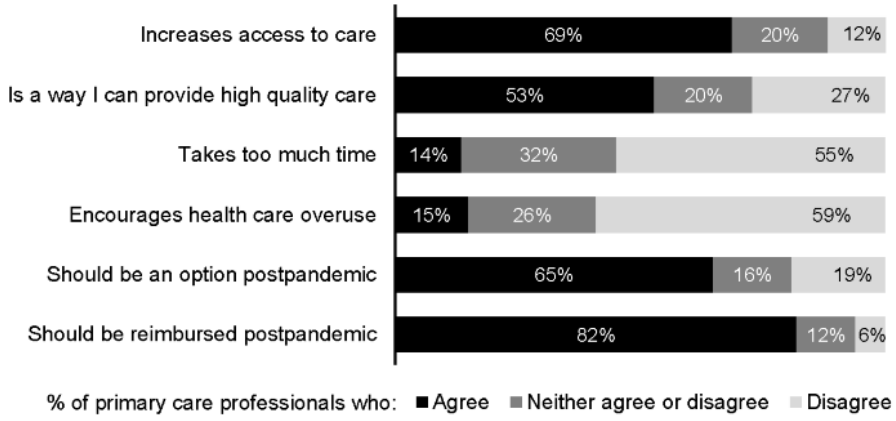


Telehealth Attitudes and Postpandemic Support

Most PCPs reported positive attitudes toward adolescent telehealth. A majority agreed that telehealth increased access to care for adolescents (720/1047, 69%) and was a way they could provide high-quality care (560/1047, 53%, Figure 2). Few agreed that telehealth took too much time (142/1047, 14%) or encouraged families to overuse health care for adolescents (157/1047, 15%).

Most PCPs indicated support for adolescent telehealth after the COVID-19 pandemic is over (Figure 2). Approximately two-thirds (683/1047, 65%) agreed that families should continue to have the option of telehealth for adolescent primary care visits. Approximately four-fifths (863/1047, 82%) agreed that health insurance plans should continue to reimburse for telehealth visits.

Figure 2. Primary care providers' attitudes toward and support for postpandemic adolescent telehealth (n=1047).



Telehealth Intentions

Approximately two-thirds of the 1047 PCPs agreed that they want to offer adolescent telehealth visits once the COVID-19 pandemic is over (n=702, 67%), while the remainder disagreed (n=188, 18%) or neither agreed nor disagreed (n=157, 15%). In multivariable analyses, wanting to offer telehealth was more common among PCPs with high, moderate, and low versus no experience seeing adolescents by telehealth in the 3 months prior to the survey (175/219, 80%, 211/286, 74%, and 280/424, 66%, vs 36/118, 31%, respectively; all $P<.001$) (Table 3). Wanting to offer telehealth was also more common among PCPs who worked in clinics that served urban versus rural areas (258/363, 71%, vs 94/159, 59%, $P=.02$) or that were located in the South or West regions versus the Northeast region (231/333, 69%, and 143/202, 71%, vs 161/265, 61%; $P=.003$ and .047, respectively). Wanting to offer telehealth was less common for PCPs with more years in practice (262/395, 66%, for 10-19 years and 246/400, 62%, for ≥20 years vs 194/252, 77%, for ≤9 years; both $P=.02$). In bivariate analyses, PCPs' adolescent telehealth intentions correlated with working outside of a solo/group practice or within a health care system, but these associations did not retain statistical significance in the multivariable model.

Table 3. Correlates of PCPs' intentions to offer adolescent telehealth after the COVID-19 pandemic (N=1047).

	PCPs[a] who want to offer telehealth, n (%)	Bivariate analysis		Multivariable analysis	
		OR[b] (95% CI)	*P* value	OR (95% CI)	*P* value
PCP characteristics					
Training					
Physician (n=747)	492 (66)	1	N/A[c]	—[d]	N/A
Advanced practice provider (n=177)	126 (71)	1.28 (0.89-1.83)	.18	—	N/A
Nurse (n=123)	84 (68)	1.12 (0.74-1.68)	.60	—	N/A
Years in practice					
0-9 (n=252)	194 (77)	1	N/A	1	N/A
10-19 (n=395)	262 (66)	0.59 (0.41-0.84)	.004	0.63 (0.43-0.92)	.02
20 or more (n=400)	246 (62)	0.48 (0.33-0.68)	<.001	0.64 (0.44-0.94)	.02
Adolescent patients seen in typical week					
1-9 (n=283)	190 (67)	1	N/A	—	N/A
10-24 (n=431)	289 (67)	1.00 (0.72-1.37)	.98	—	N/A
25 or more (n=333)	223 (67)	0.99 (0.71-1.39)	.96	—	N/A
Proportion of adolescents seen by telehealth in prior 3 months					
None (0%) (n=118)	36 (31)	1	N/A	1	N/A
Low (1%-24%) (n=424)	280 (66)	4.43 (2.85-6.88)	<.001	4.61 (2.92-7.27)	<.001
Moderate (25%-50%) (n=286)	211 (74)	6.41 (4.00-10.28)	<.001	6.53 (4.01-10.64)	<.001
High (51%-100%) (n=219)	175 (80)	9.06 (5.43-15.13)	<.001	8.99 (5.31-15.23)	<.001
Clinic or practice characteristics					
Specialty					
Family medicine (n=748)	510 (68)	1	N/A	—	N/A
Pediatrics (n=299)	192 (64)	0.84 (0.63-1.11)	.22	—	N/A
Practice type					
Solo or group (n=696)	450 (65)	1	N/A	1	N/A
Other (n=351)	252 (72)	1.39 (1.05-1.84)	.02	1.16 (0.83-1.61)	.39
Part of clinical system					
No (n=457)	280 (61)	1	N/A	1	N/A
Yes (n=590)	422 (72)	1.59 (1.22-2.06)	<.001	1.33 (0.98-1.79)	.07
Rurality					
Urban (n=363)	258 (71)	1	N/A	1	N/A
Suburban (n=525)	350 (67)	0.81 (0.61-1.09)	.17	0.87 (0.63-1.19)	.38
Rural (n=159)	94 (59)	0.59 (0.40-0.87)	.008	0.59 (0.39-0.90)	.02
Region					
Northeast (n=265)	161 (61)	1	N/A	1	N/A
Midwest (n=247)	167 (68)	1.35 (0.94-1.94)	.11	1.35 (0.92-1.99)	.13
South (n=333)	231 (69)	1.46 (1.04-2.05)	.03	1.73 (1.20-2.49)	.003
West (n=202)	143 (71)	1.57 (1.06-2.31)	.03	1.52 (1.01-2.31)	.047
COVID-19–related financial strain					
None or a little (n=360)	241 (67)	1	N/A	—	N/A
Moderate or more (n=687)	461 (67)	1.01 (0.77-1.32)	.96	—	N/A

[a]PCPs: primary care professionals.

[b]OR: odds ratio.

[c]N/A: not applicable.

[d]The variable was not included in the multivariable model because it was not statistically significant at the bivariate level.

Discussion

Principal Findings

The findings of our national study suggest that adolescent telehealth has achieved widespread adoption in the year since the COVID-19 pandemic began, with most PCPs in our national study reporting that they used adolescent telehealth and wanted to keep using it. The vast majority (89%) of PCPs reported using telehealth to see adolescents in the prior 3 months, including for chronic disease management, acute care, and mental health. This level of adoption is far higher than that reported in prepandemic studies, which found that very few pediatricians or family physicians were using telehealth to deliver care (13% and 15%, respectively) [9,10]. On average, the PCPs in our sample reported that adolescent telehealth offered more advantages than disadvantages, with advantages including increased access to care and reduced time and transportation burdens for families. The most commonly noted disadvantages were the inability to perform physical examinations and technology problems. Despite these limitations, PCPs indicated broad support for adolescent telehealth after the pandemic is over, with approximately two-thirds wanting to offer such visits themselves. This support suggests a pressing need to build on the strengths, address the challenges, and evaluate the quality of adolescent telehealth to ensure it remains a viable option for primary care delivery in the postpandemic era.

Our study provides novel data on two potential challenges for the delivery of adolescent telehealth: vaccine communication and privacy. With regard to vaccine communication, we found that the proportion of PCPs reporting that they "always" recommended vaccines during adolescent telehealth visits was high for seasonal influenza vaccine, perhaps in response to concerns about an influenza–COVID-19 "twindemic." Consistent recommendations were less common for HPV, Tdap, and meningococcal vaccines. In the case of the HPV vaccine, prepandemic studies that used similar measures in the context of traditional primary care documented levels of recommendation consistency that are comparable to or lower than what we observed [11-13]. Although such points of comparison are not available for other adolescent vaccines, our findings suggest that many PCPs are including vaccine counseling in adolescent telehealth visits, and these data can inform interventions to support and further strengthen this communication. For example, PCPs may benefit from electronic health record prompts to remind them to recommend adolescent vaccines during telehealth visits and to counsel families about how to schedule those visits. This care coordination could help to ensure that adolescent telehealth complements rather than competes with the in-person care that is integral to the delivery of vaccinations and other routine preventive health services for adolescents [14].

Consistent with prior studies [5,6], our findings suggest that privacy constitutes an important consideration but may not be a primary barrier to the delivery of adolescent telehealth. Among the PCPs in our sample with recent experience, only approximately one-quarter perceived a lack of privacy as a disadvantage of adolescent telehealth. At the same time, however, only approximately one-quarter routinely offered adolescents confidential time during telehealth visits. Despite being a recommended practice, confidential time is also inconsistently offered during in-person visits [15,16]; therefore, this low level of guideline adherence is perhaps unsurprising. Nevertheless, PCPs may have the opportunity to strengthen their telehealth practice by more consistently offering confidential time and by counseling adolescents to take steps that may help protect their privacy, such as using earphones [5,7,17].

Our study can inform future policy making by documenting PCPs' support for offering adolescent telehealth after the COVID-19 pandemic is over. Telehealth adoption during the pandemic has been possible due to expanded payer reimbursement, which prepandemic studies identified as the single largest barrier to bringing pediatric telehealth programs to scale [9,14,18]. Importantly, most PCPs in our sample believed that health insurance plans should continue to reimburse for adolescent telehealth after the pandemic is over. Furthermore, many PCPs believed that families should have the option of telehealth for adolescent primary care visits and wanted to offer telehealth visits themselves. Wanting to offer adolescent telehealth was more common among PCPs who had recently used it or who worked in the South or West regions of the United States, but it was less common among those serving rural areas. Future research can extend the present study by assessing factors, such as internet connectivity, that might explain these geographic differences.

Strengths and Limitations

This study is, to our knowledge, the first national evaluation of PCPs' experience using adolescent telehealth in the context of the COVID-19 pandemic. Study strengths include the use of data from a large, national sample of PCPs, including physicians, advanced practitioners, and nurses. Although evaluating adolescent telehealth from the perspective of those who deliver it is a study strength, our use of self-reported measures also constitutes a limitation. For example, PCPs may have overestimated the consistency with which they recommended vaccines or offered confidential time during adolescent telehealth visits. Another limitation is the modest response rate among advanced practice providers and nurses. Including nonphysician participants was important for enriching our data with diverse perspectives, but additional research with larger sample sizes may be needed to more fully understand how adolescent telehealth practice and attitudes vary across clinical roles. We acknowledge that other perspectives and data sources are important for understanding the impact of adolescent telehealth. Most notably, PCPs in our sample perceived telehealth as expanding access to care and reducing burden on

their patients, but future research is needed to understand the extent to which adolescents and their families experience telehealth as patient-centered and equitable. Finally, we note that our study examines telehealth broadly as including visits by video and telephone. PCPs' experience of adolescent telehealth may vary between these two delivery modes as well as by health system factors, such as the extent to which telehealth is integrated with electronic health record platforms and patient portals. Future research will be needed to understand the influence of these contextual factors.

Conclusion

One year into the COVID-19 pandemic, our national study finds that PCPs have widely adopted adolescent telehealth and endorse its continued use. In light of this endorsement, health care system leaders, payers, professional organizations, researchers, and other key stakeholders should redouble their efforts to support PCPs in adolescent telehealth delivery, including by further evaluating and fairly reimbursing such services. In this way, we can ensure that telehealth realizes its potential to increase health care access and to serve adolescents in a way that is effective, patient centered, and equitable.

Data Availability Statement

Deidentified individual participant data will not be made available.

Acknowledgments

This project was supported by a Gillings Innovation Laboratory award from the UNC Gillings School of Global Public Health and the Lineberger Comprehensive Cancer Center through a Tier 1 Innovation Award made possible by the University Cancer Research Fund. The authors' time was supported by grants from the National Institutes of Health (T32CA057726 for BKG and K08AI148607 for PT). The funders played no role in the study design; the collection, analysis, and interpretation of data; the writing of the report; or the decision to submit the manuscript for publication.

Conflicts of Interest

PT receives research funding from the National Institutes of Health and the American Society of Tropical Medicine and Hygiene/Burroughs-Wellcome fund, as well as research support from Abbott Laboratories. NTB has served as a paid advisor to the Centers for Disease Control and Prevention, the World Health Organization, and Merck & Co, Inc.

References

1. Myers KM, Valentine JM, Melzer SM. Child and adolescent telepsychiatry: utilization and satisfaction. Telemed J E Health 2008 Mar;14(2):131-137. [doi: 10.1089/tmj.2007.0035] [Medline: 18361702]
2. Sequeira GM, Kidd KM, Coulter RW, Miller E, Fortenberry D, Garofalo R, et al. Transgender youths' perspectives on telehealth for delivery of gender-affirming care. J Adolesc Health 2021 Jun;68(6):1207-1210 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.08.028] [Medline: 32980246]
3. Davies B, Kenia P, Nagakumar P, Gupta A. Paediatric and adolescent asthma: a narrative review of telemedicine and emerging technologies for the post-COVID-19 era. Clin Exp Allergy 2021 Mar 10;51(3):393-401 [FREE Full text] [doi: 10.1111/cea.13836] [Medline: 33527625]
4. Wood CL, Clements SA, McFann K, Slover R, Thomas JF, Wadwa RP. Use of telemedicine to improve adherence to American Diabetes Association Standards in pediatric type 1 diabetes. Diabetes Technol Ther 2016 Jan;18(1):7-14. [doi: 10.1089/dia.2015.0123] [Medline: 26295939]
5. Wood SM, White K, Peebles R, Pickel J, Alausa M, Mehringer J, et al. Outcomes of a rapid adolescent telehealth scale-up during the COVID-19 pandemic. J Adolesc Health 2020 Aug;67(2):172-178 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.05.025] [Medline: 32611509]
6. Barney A, Buckelew S, Mesheriakova V, Raymond-Flesch M. The COVID-19 pandemic and rapid implementation of adolescent and young adult telemedicine: challenges and opportunities for innovation. J Adolesc Health 2020 Aug;67(2):164-171 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.05.006] [Medline: 32410810]
7. Carlson JL, Goldstein R. Using the electronic health record to conduct adolescent telehealth visits in the time of COVID-19. J Adolesc Health 2020 Aug;67(2):157-158 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.05.022] [Medline: 32517972]
8. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. The American Association for Public Opinion Research. 2016. URL: https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf [accessed 2021-08-30]
9. Sisk B, Alexander J, Bodnar C, Curfman A, Garber K, McSwain SD, et al. Pediatrician attitudes toward and experiences with telehealth use: Results from a national survey. Acad Pediatr 2020 Jul;20(5):628-635 [FREE Full text] [doi: 10.1016/j.acap.2020.05.004] [Medline: 32437881]
10. Klink K, Coffman M, Moore M, Jetty A, Petterson S, Bazemore A. Family Physicians and Telehealth: Findings from a National Survey. Robert Graham Center. 2015 Oct 30. URL: https://www.graham-center.org/content/dam/rgc/documents/publications-reports/reports/RGC%202015%20Telehealth%20Report.pdf [accessed 2021-08-30]

11. Finney Rutten LJ, St Sauver JL, Beebe TJ, Wilson PM, Jacobson DJ, Fan C, et al. Association of both consistency and strength of self-reported clinician recommendation for HPV vaccination and HPV vaccine uptake among 11- to 12-year-old children. Vaccine 2017 Oct 27;35(45):6122-6128 [FREE Full text] [doi: 10.1016/j.vaccine.2017.09.056] [Medline: 28958810]
12. Vadaparampil ST, Malo TL, Sutton SK, Ali KN, Kahn JA, Casler A, et al. Missing the target for routine human papillomavirus vaccination: consistent and strong physician recommendations are lacking for 11- to 12-year-old males. Cancer Epidemiol Biomarkers Prev 2016 Aug 02;25(10):1435-1446. [doi: 10.1158/1055-9965.epi-15-1294]
13. Gilkey MB, Malo TL, Shah PD, Hall ME, Brewer NT. Quality of physician communication about human papillomavirus vaccine: findings from a national survey. Cancer Epidemiol Biomarkers Prev 2015 Nov;24(11):1673-1679 [FREE Full text] [doi: 10.1158/1055-9965.EPI-15-0326] [Medline: 26494764]
14. Fiks AG, Jenssen BP, Ray KN. A defining moment for pediatric primary care telehealth. JAMA Pediatr 2021 Jan 01;175(1):9-10. [doi: 10.1001/jamapediatrics.2020.1881] [Medline: 32658256]
15. Sieving RE, Mehus C, Catallozzi M, Grilo S, Steiner RJ, Brar P, et al. Understanding primary care providers' perceptions and practices in implementing confidential adolescent sexual and reproductive health services. J Adolesc Health 2020 Oct;67(4):569-575. [doi: 10.1016/j.jadohealth.2020.03.035] [Medline: 32389456]
16. Grilo SA, Catallozzi M, Santelli JS, Yan H, Song X, Heitel J, et al. Confidentiality discussions and private time with a health-care provider for youth, United States, 2016. J Adolesc Health 2019 Mar;64(3):311-318. [doi: 10.1016/j.jadohealth.2018.10.301] [Medline: 30638752]
17. McSwain SD, Bernard J, Burke BL, Cole SL, Dharmar M, Hall-Barrow J, et al. American Telemedicine Association operating procedures for pediatric telehealth. Telemed J E Health 2017 Sep;23(9):699-706. [doi: 10.1089/tmj.2017.0176] [Medline: 28829680]
18. Olson CA, McSwain SD, Curfman AL, Chuo J. The current pediatric telehealth landscape. Pediatrics 2018 Mar;141(3):e20172334 [FREE Full text] [doi: 10.1542/peds.2017-2334] [Medline: 29487164]

Abbreviations

HPV: human papillomavirus
PCP: primary care professional
Tdap: tetanus, diphtheria, and acellular pertussis

Edited by C Basch; submitted 14.06.21; peer-reviewed by A Barney, M Salimi; comments to author 01.07.21; revised version received 10.07.21; accepted 14.07.21; published 10.09.21.

Please cite as:
Gilkey MB, Kong WY, Huang Q, Grabert BK, Thompson P, Brewer NT
Using Telehealth to Deliver Primary Care to Adolescents During and After the COVID-19 Pandemic: National Survey Study of US Primary Care Professionals
J Med Internet Res 2021;23(9):e31240
URL: https://www.jmir.org/2021/9/e31240
doi:10.2196/31240
PMID:34406974

Original Paper

A 3D Hologram With Mixed Reality Techniques to Improve Understanding of Pulmonary Lesions Caused by COVID-19: Randomized Controlled Trial

Songxiang Liu[1,2*], MD; Mao Xie[1*], MD; Zhicai Zhang[1], MD; Xinghuo Wu[1], MD; Fei Gao[1], MD; Lin Lu[1], MD; Jiayao Zhang[1], MD; Yi Xie[1], MD; Fan Yang[3], MD; Zhewei Ye[1,2], MD

[1]Department of Orthopedics, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[2]Intelligent Medical Laboratory, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[3]Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[*]these authors contributed equally

Corresponding Author:
Zhewei Ye, MD
Department of Orthopedics
Union Hospital, Tongji Medical College
Huazhong University of Science and Technology
1277 Jiefang Avenue
Wuhan, 430019
China
Phone: 86 13971213880
Email: yezhewei@hust.edu.cn

Abstract

Background: The COVID-19 outbreak has now become a pandemic and has had a serious adverse impact on global public health. The effect of COVID-19 on the lungs can be determined through 2D computed tomography (CT) imaging, which requires a high level of spatial imagination on the part of the medical provider.

Objective: The purpose of this study is to determine whether viewing a 3D hologram with mixed reality techniques can improve medical professionals' understanding of the pulmonary lesions caused by COVID-19.

Methods: The study involved 60 participants, including 20 radiologists, 20 surgeons, and 20 medical students. Each of the three groups was randomly divided into two groups, either the 2D CT group (n=30; mean age 29 years [range 19-38 years]; males=20) or the 3D holographic group (n=30; mean age 30 years [range 20=38 years]; males=20). The two groups completed the same task, which involved identifying lung lesions caused by COVID-19 for 6 cases using a 2D CT or 3D hologram. Finally, an independent radiology professor rated the participants' performance (out of 100). All participants in two groups completed a Likert scale questionnaire regarding the educational utility and efficiency of 3D holograms. The National Aeronautics and Space Administration Task Load Index (NASA-TLX) was completed by all participants.

Results: The mean task score of the 3D hologram group (mean 91.98, SD 2.45) was significantly higher than that of the 2D CT group (mean 74.09, SD 7.59; P<.001). With the help of 3D holograms, surgeons and medical students achieved the same score as radiologists and made obvious progress in identifying pulmonary lesions caused by COVID-19. The Likert scale questionnaire results showed that the 3D hologram group had superior results compared to the 2D CT group (teaching: 2D CT group median 2, IQR 1-2 versus 3D group median 5, IQR 5-5; P<.001; understanding and communicating: 2D CT group median 1, IQR 1-1 versus 3D group median 5, IQR 5-5; P<.001; increasing interest: 2D CT group median 2, IQR 2-2 versus 3D group median 5, IQR 5-5; P<.001; lowering the learning curve: 2D CT group median 2, IQR 1-2 versus 3D group median 4, IQR 4-5; P<.001; spatial awareness: 2D CT group median 2, IQR 1-2 versus 3D group median 5, IQR 5-5; P<.001; learning: 2D CT group median 3, IQR 2-3 versus 3D group median 5, IQR 5-5; P<.001). The 3D group scored significantly lower than the 2D CT group for the "mental," "temporal," "performance," and "frustration" subscales on the NASA-TLX.

Conclusions: A 3D hologram with mixed reality techniques can be used to help medical professionals, especially medical students and newly hired doctors, better identify pulmonary lesions caused by COVID-19. It can be used in medical education to improve spatial awareness, increase interest, improve understandability, and lower the learning curve.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2100045845; http://www.chictr.org.cn/showprojen.aspx?proj=125761

(J Med Internet Res 2021;23(9):e24081) doi:10.2196/24081

KEYWORDS

COVID-19; mixed reality; hologram; pulmonary; lesion; diagnostic; imaging

Introduction

The COVID-19 outbreak has now become a pandemic [1]. It has had a serious adverse impact on global public health [2]. Many doctors are on the front line of fighting the epidemic, including orthopedic surgeons, general surgeons, and neurosurgeons. Two-dimensional computed tomography (CT) scanning, which requires high spatial imagination on the part of medical professionals, has been traditionally used to determine the status of a pulmonary infection [3,4]. Often, judging the specific condition of the pulmonary lesions caused by COVID-19 would require the expertise of experienced respiratory, radiology, infectious disease, and intensive care department experts. Doctors who are not specialized in respiratory-related cases, not to mention medical students and the public, may not find it easy to understand the clinical significance of a 2D CT scan.

Both doctors and the public have played a vital role in combating COVID-19. Doctors treat patients with COVID-19 and fight the virus directly. The public's self-protection and isolation efforts help to stop the spread of the virus. However, there are still many members of the public who are not aware of the necessary protective measures. In addition, it is very important for medical students to have a better understanding of COVID-19, as they are the successors in the fight against this virus. This dilemma may be experienced by other countries as well.

Mixed reality techniques can overlap virtual and real worlds. The user can view the real world while manipulating the digital content generated by a device [5]. Mixed reality has been used clinically to help doctors better understand anatomical structures [6-8]. However, the use of 3D holograms to image lungs affected by COVID-19 has not been reported in the literature. Therefore, we present a new method that applies mixed reality techniques to create a 3D hologram of lungs affected by COVID-19.

Methods

Trial Design

The study was a parallel-group randomized controlled trial. It was conducted in Wuhan, China, from March 2020 to September 2020. Participants were randomized into two groups in a 1:1 ratio: (1) a 3D holographic intervention group or (2) a 2D CT control group.

Participants and Recruitment

We randomly selected radiologists, surgeons, and medical students from our institution and asked them to participate in this study. The radiologists, surgeons, and medical students were eligible to participate in this trial if they had not previously seen 3D holograms of lungs affected by COVID-19. The study involved 60 participants, including 20 radiologists, 20 surgeons, and 20 medical students. Each of the three groups was randomly divided into two groups: (1) the 2D CT group and (2) the 3D holographic group (Figure S1 in Multimedia Appendix 1). Exclusion criteria were as follows: surgeons and medical students trained in COVID-19 lung CT imaging, and participants that had previously seen 3D holograms of COVID-19 lungs.

Intervention

For this study, 3D reconstructions of patients' lungs were made using data from CT scans using a StarCloud workstation (a 3D reconstruction software from Visual3D). Polygon mesh files were exported from the StarCloud workstation. These were uploaded into the StarCloud Mixed Reality system (Visual3D). After uploading the data, 3D holographic images were automatically converted as a case-specific computer graphic for mixed reality, referred to as a hologram. The data pertaining to the hologram can be downloaded onto a Microsoft HoloLens (Microsoft Corp).

The two groups completed the same task, which involved identifying lesion areas in the lungs of 6 cases of COVID-19 using a 2D CT or 3D hologram of the lungs. The 6 cases of COVID-19 used for the study involved moderate to severe COVID-19 (Table S1 in Multimedia Appendix 1).

Assessments

An independent radiology professor rated the participants' performance (out of 100). All participants in the two groups completed a Likert scale questionnaire regarding the educational utility and efficiency of 3D holograms. The National Aeronautics and Space Administration Task Load Index (NASA-TLX) was completed by all participants. The NASA-TLX is a multidimensional rating procedure that provides an overall workload score between 0 and 100 based on a weighted average of ratings on 6 subscales [9]: (1) mental demands ("How mentally demanding was the task?"), (2) physical demands ("How physically demanding was the task?"), (3) temporal demands ("How hurried or rushed was the pace of the task?"), (4) own performance ("How successful were you in performing the task?"), (5) effort ("How hard did you have to work to achieve your level of performance?"), and (6) frustration ("How insecure, discouraged, irritated, stressed, and annoyed were you?"). The ethics committee of Wuhan Union Hospital, Tongji Medical College, Huazhong University of Science and Technology approved this study, and all participants provided signed informed consent.

Statistical Analysis

Statistical analyses were performed using SPSS (version 19; IBM Corp). Continuous variables were expressed as means and standard deviations. Results of the NASA-TLX questionnaires were summarized in terms of means and standard deviations.

Data were processed using analysis of variance to determine possible relationships between individual characteristics and workload.

Results

Sample Characteristics

Participant baseline characteristics are described in Table 1. The study involved 60 participants, including 20 radiologists, 20 surgeons, and 20 medical students. Each of the three groups was randomly divided into two groups: (1) the 2D CT group (n=30; mean age 29 years [range 19-38 years]; males=20) and (2) the 3D holographic group (n=30; mean age 30 years [range 20-38 years]; males=20; Figure S1 in Multimedia Appendix 1).

Table 1. Demographic information.

Variables	3D holographic group (n=30)	2D computed tomography group (n=30)
Gender		
Male	20	20
Female	10	10
Age (year), mean (range)	30 (20-38)	29 (19-38)
Participants		
Radiologists, total (male)	10 (5)	10 (5)
Surgeons, total (male)	10 (10)	10 (10)
Medical students, total (male)	10 (5)	10 (5)

Outcomes

Clear 3D visual holographic renderings were obtained for the lungs of patients with COVID-19 (Figure 1). In the 3D holographic (mixed reality) group, the task scores for all participants (mean 91.98, SD 2.45), radiologists (mean 93.60, SD 2.25), surgeons (mean 91.50, SD 2.31), and medical students (mean 91.25, SD 2.18) were significantly higher than that of the conventional (2D) group (all participants: mean 74.09, SD 7.59; radiologists: mean 82.63, SD 2.28; surgeons: mean 74.55, SD 2.52; medical students: mean 65.10, SD 65.10; P<.001). With the help of 3D holograms, surgeons and medical students achieved the same scores as radiologists and made obvious progress in identifying pulmonary lesions caused by COVID-19 (Figure 2, Table 2). The Likert scale questionnaire revealed that the 3D hologram group had superior results compared to the 2D CT group (teaching: 2D CT group median score 2, IQR 1-2 versus 3D group median 5, IQR 5-5; P<.001; understanding and communicating: 2D CT group median score 1, IQR 1-1 versus 3D group median 5, IQR 5-5; P<.001; increasing interest: 2D CT group median score 2, IQR 2-2 versus 3D group median 5, IQR 5-5; P<.001; lowering the learning curve: 2D CT group median score 2, IQR 1-2 versus 3D group median 4, IQR 4-5; P<.001; spatial awareness: 2D CT group median score 2, IQR 1-2 versus 3D group median 5, IQR 5-5; P<.001; learning: 2D CT group median score 3, IQR 2-3 versus 3D group median 5, IQR 5-5; P<.001; Figure 3, Table 2).

The 3D hologram group scored significantly lower than the 2D CT group on the “mental,” “temporal,” “performance,” and “frustration” subscales of the NASA-TLX. Compared with the traditional 2D group, the 3D hologram group reported higher physical demands. The main reason is that participants need to wear a mixed reality headset, but the weight of the mixed reality glasses was acceptable (Figure 4, Table 2).

Figure 1. A 3D hologram of the pulmonary lesions caused by COVID-19. (A) A computed tomography scan of a patient with COVID-19 revealed patchy pure ground-glass opacities (red arrow). (B) A hologram can clearly show the lesion site and lesion range (green color) of the patient's lung infection, allowing stereoscopic viewing from 360 degrees in a physical world. (C) Simultaneous analysis of 2D and 3D images.

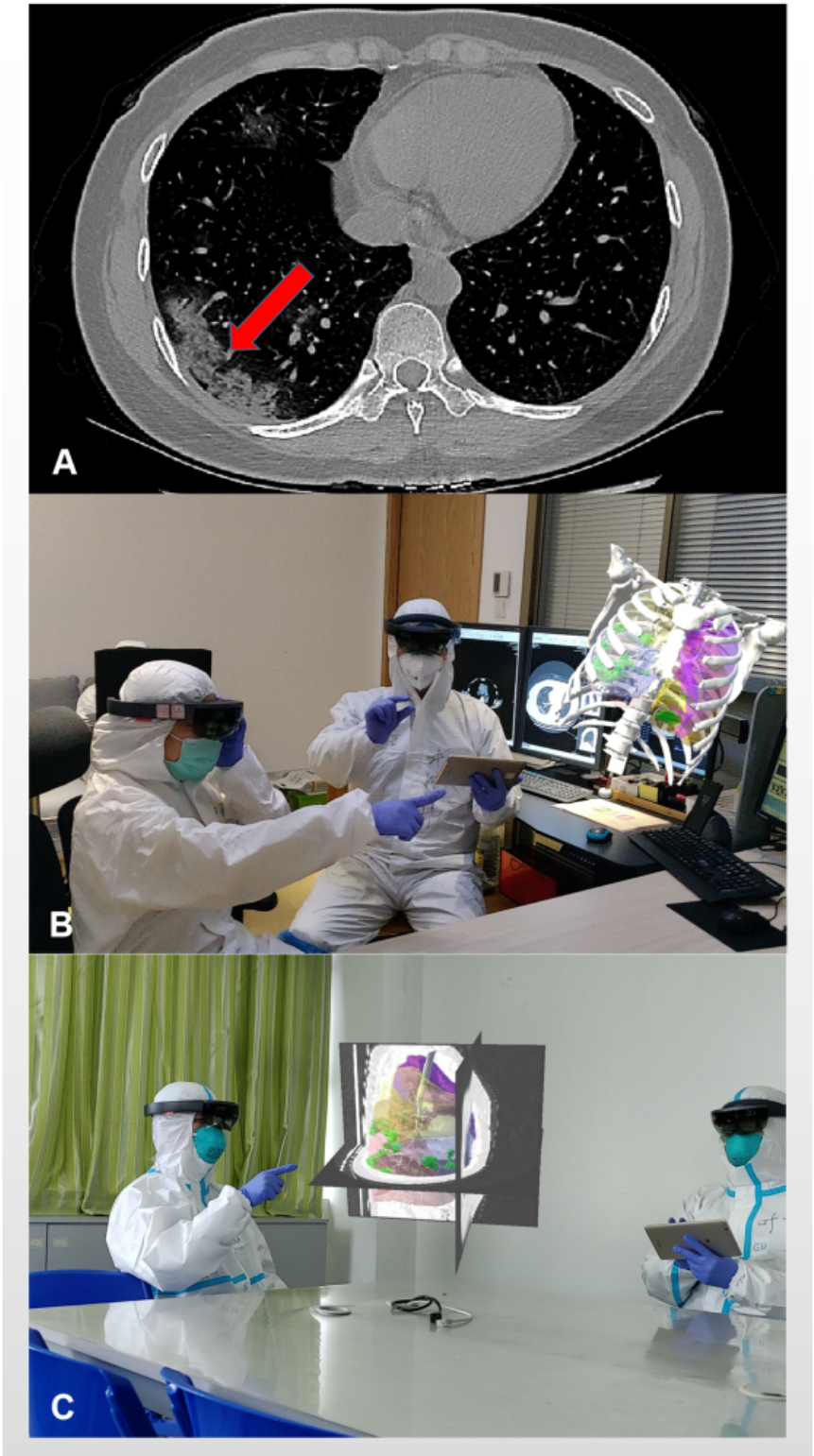


Figure 2. Task scores for identifying COVID-19 lesion areas in lungs. In the mixed reality group, the mean task score was significantly higher than that of the conventional group, with *P*<.001. With the help of 3D holograms, surgeons and medical students achieved the same score as radiologists and made obvious progress in identifying the pulmonary lesions caused by COVID-19.

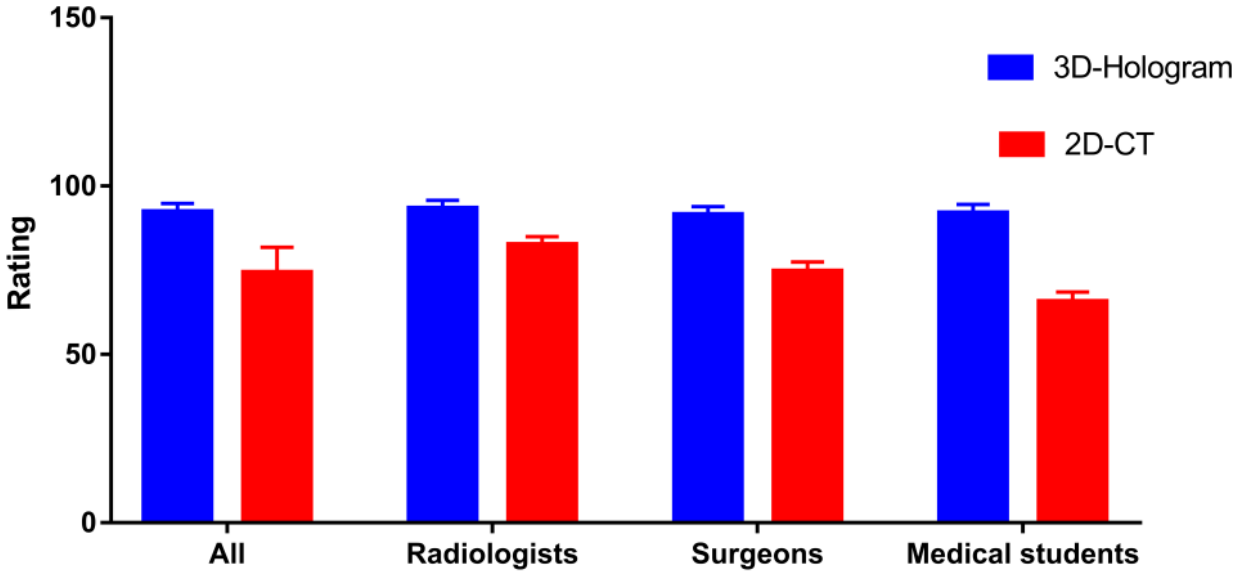


Table 2. Task scores, NASA Task Load Index scores, and Likert scale questionnaire scores.

Category	2D CT[a] group, mean (SD)	2D CT group, median (IQR)	3D holographic group, mean (SD)	3D holographic group, median (IQR)	*P* value
Task score					
All participants	74.09 (7.59)	75 (68-80)	91.98 (2.45)	90 (90-95)	<.001
Radiologists	82.63 (2.28)	83 (80-85)	93.60 (2.25)	95 (90-95)	<.001
Surgeons	74.55 (2.52)	75 (75-75)	91.50 (2.31)	90 (90-95)	<.001
Medical students	65.10 (2.61)	65 (65-68)	91.25 (2.18)	90 (90-94)	<.001
NASA Task Load Index scores					
Mental	47.03 (5.71)	50 (40-50)	20.06 (2.85)	20 (20-20)	<.001
Physical	19.75 (1.09)	20 (20-20)	25.08 (0.64)	25 (25-25)	<.001
Temporal	63.11 (6.45)	60 (60-70)	40.31 (3.04)	40 (40-40)	<.001
Performance	26.17 (6.03)	25 (20-30)	10.08 (0.99)	10 (10-10)	<.001
Effort	69.92 (0.99)	70 (70-70)	69.92 (0.99)	70 (70-70)	.99
Frustration	29.22 (5.95)	30 (25-35)	14.58 (1.39)	15 (15-15)	<.001
Likert-scale questionnaire scores					
Effectiveness as teaching tool	1.73 (0.51)	2 (1-2)	4.8 (0.40)	5 (5-5)	<.001
Better understanding and communication	1.1 (0.30)	1 (1-1)	4.83 (0.37)	5 (5-5)	<.001
Increasing interest	1.8 (0.4)	2 (2-2)	4.9 (0.3)	5 (5-5)	<.001
Lowering the learning curve	1.73 (0.51)	2 (1-2)	4.4 (0.49)	4 (4-5)	<.001
Better spatial awareness	1.6 (0.49)	2 (1-2)	4.93 (0.25)	5 (5-5)	<.001
Learning is easier	2.63 (0.55)	3 (2-3)	4.83 (0.37)	5 (5-5)	<.001

[a]CT: computed tomography.

Figure 3. Responses to a Likert scale questionnaire regarding the educational utility and efficiency of 3D holograms. The 3D hologram group indicated higher educational utility and efficiency than the 2D CT group. CT: computed tomography.

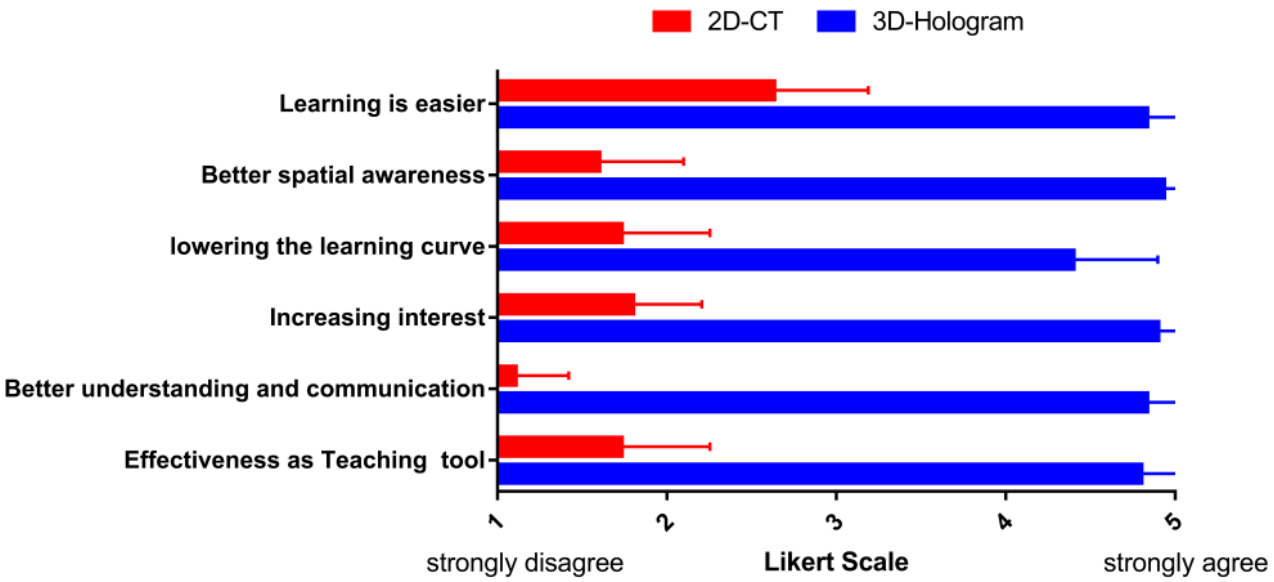
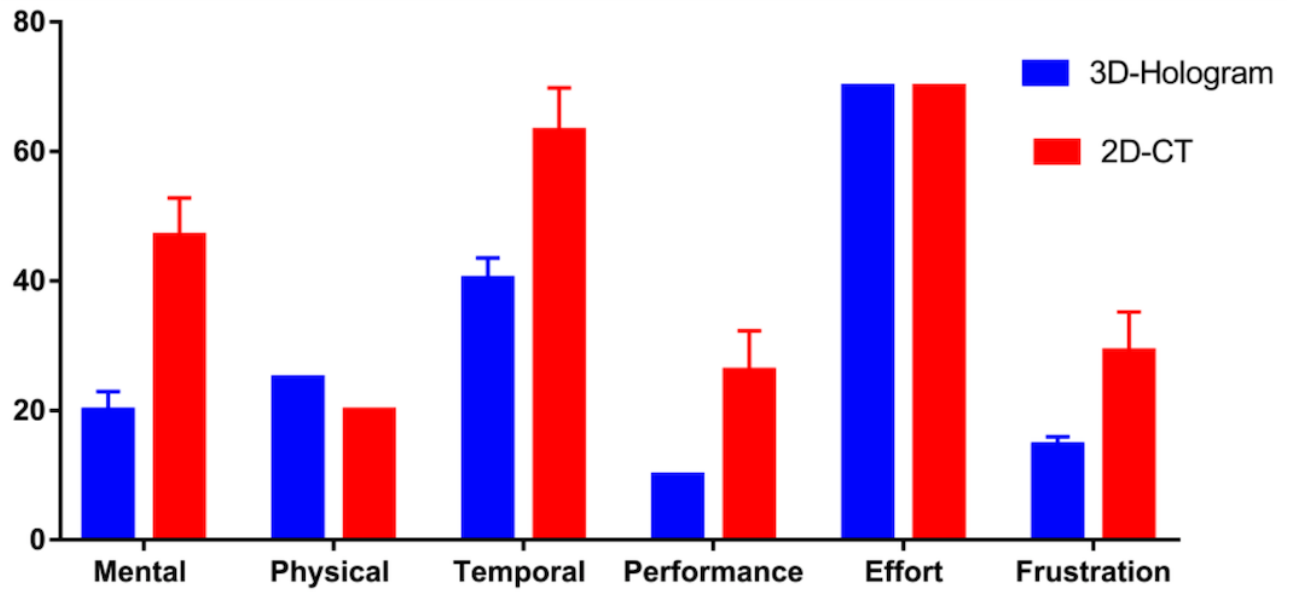


Figure 4. Results from the National Aeronautics and Space Administration (NASA) Task Load Index assessment. The mean rating for each subscale is given. Those that used holograms scored significantly lower than those that used 2D CT scans for task load on the "mental," "temporal," "performance," and "frustration" subscales. CT: computed tomography.



Discussion

A 3D hologram with mixed reality techniques can be used to help medical professionals, especially medical students and newly hired doctors, better identify pulmonary lesions caused by COVID-19. It can be used in medical education to increase interest, improve understandability, improve spatial awareness, and lower the learning curve.

A CT scan of the lungs of patients with COVID-19 revealed patchy pure ground-glass opacities (Figure 1A). We visualized the patients' lungs using mixed-reality technology, which can transform the 2D CT image of the lung into a 3D hologram (Figure 1B) to clearly show the lesion site and lesion range (green color) of the patient's lung infection, allowing stereoscopic viewing from 360 degrees in a physical world. Conversely, a standard 2D CT usually only allows one to see the local area of a certain layer of the lung. The 3D hologram can also analyze 2D and 3D images simultaneously (Figure 1C). As the 3D hologram can clearly show the spatial anatomical neighborhood, it does not require the human brain to mentally transform the complex 2D structure into a complete 3D structure. The results of this study showed that in the 3D hologram group, compared with the traditional 2D CT group, task loads for the "mental," "temporal," and "frustration" subscales were significantly reduced, and better performance was obtained. Thus, this study can help doctors—whether they have just started to practice, are in nonrespiratory specialties, or are in respiratory specialties—better identify and understand the pulmonary lesions caused by COVID-19. In addition, this

can improve medical students' understanding of COVID-19, which is important as they are the successors in the fight against this virus.

This technique of image reconstruction and presentation can also be applied to other anatomical systems throughout the body (Figure 5). It can contribute to a better understanding of normal and abnormal body structure for both medical and nonmedical individuals and could be especially useful for medical students when used in future medical education. In addition, unlike traditional 3D anatomical drawings, mixed reality technology allows multiple people to view the same 3D hologram from 360 degrees during the teaching process.

Figure 5. Images show a 3D holographic mixed reality technology being used to present and teach the digestive system and classification of pelvic fractures.

The COVID-19 outbreak has also created challenges for anatomy education [10]. Cadaver specimens were the standard learning method of anatomy classes in the past [11]. However, substantial financial, ethical, and supervisory constraints on their use and the shortage of cadaver specimens are ongoing problems faced by teaching colleges and universities [12]; in addition, during the COVID-19 epidemic, the use of cadaver specimens has the potential risk of virus transmission [10,13]. Digital anatomical imaging is a feasible alternative solution to the use of cadaver specimens and is more visual, accessible, clean, fun, and inexpensive.

Although this study shows promising results, further studies in more institutions, populations, and locations are needed in the future.

In conclusion, a 3D hologram with mixed reality techniques can provide a better understanding of the pulmonary lesions caused by COVID-19 and will play an important role in future medical education.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (grant no. 81974355), Major Technical Innovation Project of Hubei Province (grant no. 2016-176), and Huazhong University of Science and Technology COVID-19 Emergency Science and Technology Project (grant no. 2020kfyXGYJ095).

Authors' Contributions

Authors ZY and FY contributed equally as corresponding authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Information about computed tomography images of COVID-19 lungs.
[DOC File , 47 KB - jmir_v23i9e24081_app1.doc]

Multimedia Appendix 2
CONSORT-eHEALTH checklist V 1.6.1.
[PDF File (Adobe PDF File), 692 KB - jmir_v23i9e24081_app2.pdf]

References

1. Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. Acta Biomed 2020 Mar 19;91(1):157-160 [FREE Full text] [doi: 10.23750/abm.v91i1.9397] [Medline: 32191675]
2. Steeb T, Follmann M, Hagen RM, Berking C, Heppt MV. Implications of the COVID-19 Pandemic for the Development and Update of Clinical Practice Guidelines: Viewpoint. J Med Internet Res 2020 Dec 29;22(12):e20064 [FREE Full text] [doi: 10.2196/20064] [Medline: 33347419]
3. Li Y, Xia L. Coronavirus Disease 2019 (COVID-19): Role of Chest CT in Diagnosis and Management. AJR Am J Roentgenol 2020 Jun;214(6):1280-1286. [doi: 10.2214/AJR.20.22954] [Medline: 32130038]
4. Aljondi R, Alghamdi S. Diagnostic Value of Imaging Modalities for COVID-19: Scoping Review. J Med Internet Res 2020 Aug 19;22(8):e19673 [FREE Full text] [doi: 10.2196/19673] [Medline: 32716893]
5. Verhey JT, Haglin JM, Verhey EM, Hartigan DE. Virtual, augmented, and mixed reality applications in orthopedic surgery. Int J Med Robot 2020 Apr;16(2):e2067. [doi: 10.1002/rcs.2067] [Medline: 31867864]
6. Hu H, Feng X, Shao Z, Xie M, Xu S, Wu X, et al. Application and Prospect of Mixed Reality Technology in Medical Field. Curr Med Sci 2019 Feb 13;39(1):1-6. [doi: 10.1007/s11596-019-1992-8] [Medline: 30868484]
7. Wainman B, Pukas G, Wolak L, Mohanraj S, Lamb J, Norman GR. The Critical Role of Stereopsis in Virtual and Mixed Reality Learning Environments. Anat Sci Educ 2020 May 20;13(3):401-412. [doi: 10.1002/ase.1928] [Medline: 31665563]
8. Hilt AD, Mamaqi Kapllani K, Hierck BP, Kemp AC, Albayrak A, Melles M, et al. Perspectives of Patients and Professionals on Information and Education After Myocardial Infarction With Insight for Mixed Reality Implementation: Cross-Sectional Interview Study. JMIR Hum Factors 2020 Jun 23;7(2):e17147 [FREE Full text] [doi: 10.2196/17147] [Medline: 32573464]
9. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Hancock PA, Meshkati N, editors. Human Mental Workload. Amsterdam: Elsevier; 1988:139-183.
10. Singal A, Bansal A, Chaudhary P. Cadaverless anatomy: Darkness in the times of pandemic Covid-19. Morphologie 2020 Sep;104(346):147-150 [FREE Full text] [doi: 10.1016/j.morpho.2020.05.003] [Medline: 32518047]
11. Turney B. Anatomy in a Modern Medical Curriculum. Ann R Coll Surg Engl 2007 Mar;89(2):104-107. [doi: 10.1308/003588407x168244]
12. Ravi KS. Dead Body Management in Times of Covid-19 and its Potential Impact on the Availability of Cadavers for Medical Education in India. Anat Sci Educ 2020 May 28;13(3):316-317 [FREE Full text] [doi: 10.1002/ase.1962] [Medline: 32279462]
13. Pather N, Blyth P, Chapman JA, Dayal MR, Flack NA, Fogg QA, et al. Forced Disruption of Anatomy Education in Australia and New Zealand: An Acute Response to the Covid-19 Pandemic. Anat Sci Educ 2020 May 10;13(3):284-300 [FREE Full text] [doi: 10.1002/ase.1968] [Medline: 32306555]

Abbreviations

CT: computed tomography
NASA-TLX: National Aeronautics and Space Administration Task Load Index

Edited by C Basch; submitted 03.09.20; peer-reviewed by Z Ni, W Ji; comments to author 03.11.20; revised version received 07.12.20; accepted 26.05.21; published 10.09.21.

Please cite as:
Liu S, Xie M, Zhang Z, Wu X, Gao F, Lu L, Zhang J, Xie Y, Yang F, Ye Z
A 3D Hologram With Mixed Reality Techniques to Improve Understanding of Pulmonary Lesions Caused by COVID-19: Randomized Controlled Trial
J Med Internet Res 2021;23(9):e24081
URL: https://www.jmir.org/2021/9/e24081
doi:10.2196/24081
PMID:34061760

Original Paper

Revealing Public Opinion Towards COVID-19 Vaccines With Twitter Data in the United States: Spatiotemporal Perspective

Tao Hu[1,2*], PhD; Siqin Wang[3*], PhD; Wei Luo[4*], PhD; Mengxi Zhang[5], PhD; Xiao Huang[6], PhD; Yingwei Yan[4], PhD; Regina Liu[7], BSc; Kelly Ly[8], BSc; Viraj Kacker[9], BSc; Bing She[10], PhD; Zhenlong Li[11], PhD

[1]Department of Geography, Oklahoma State University, Stillwater, OK, United States
[2]Center for Geographic Analysis, Harvard University, Cambridge, MA, United States
[3]School of Earth and Environmental Sciences, University of Queensland, Brisbane, Australia
[4]Department of Geography, National University of Singapore, Singapore, Singapore
[5]Department of Nutrition and Health Science, Ball State University, Muncie, IN, United States
[6]Department of Geosciences, University of Arkansas, Fayetteville, AR, United States
[7]Department of Biology, Mercer University, Macon, GA, United States
[8]Department of Computer Science, University of Massachusetts Lowell, Lowell, MA, United States
[9]College of Computing, Georgia Institute of Technology, Atlanta, GA, United States
[10]Institute for Social Research, University of Michigan, Ann Arbor, MI, United States
[11]Geoinformation and Big Data Research Laboratory, Department of Geography, University of South Carolina, Columbia, SC, United States
[*]these authors contributed equally

Corresponding Author:
Wei Luo, PhD
Department of Geography
National University of Singapore
1 Arts Link, #04-32 Block AS2
Singapore, 117570
Singapore
Phone: 65 65163851
Email: geowl@nus.edu.sg

Abstract

Background: The COVID-19 pandemic has imposed a large, initially uncontrollable, public health crisis both in the United States and across the world, with experts looking to vaccines as the ultimate mechanism of defense. The development and deployment of COVID-19 vaccines have been rapidly advancing via global efforts. Hence, it is crucial for governments, public health officials, and policy makers to understand public attitudes and opinions towards vaccines, such that effective interventions and educational campaigns can be designed to promote vaccine acceptance.

Objective: The aim of this study was to investigate public opinion and perception on COVID-19 vaccines in the United States. We investigated the spatiotemporal trends of public sentiment and emotion towards COVID-19 vaccines and analyzed how such trends relate to popular topics found on Twitter.

Methods: We collected over 300,000 geotagged tweets in the United States from March 1, 2020 to February 28, 2021. We examined the spatiotemporal patterns of public sentiment and emotion over time at both national and state scales and identified 3 phases along the pandemic timeline with sharp changes in public sentiment and emotion. Using sentiment analysis, emotion analysis (with cloud mapping of keywords), and topic modeling, we further identified 11 key events and major topics as the potential drivers to such changes.

Results: An increasing trend in positive sentiment in conjunction with a decrease in negative sentiment were generally observed in most states, reflecting the rising confidence and anticipation of the public towards vaccines. The overall tendency of the 8 types of emotion implies that the public trusts and anticipates the vaccine. This is accompanied by a mixture of fear, sadness, and anger. Critical social or international events or announcements by political leaders and authorities may have potential impacts on public opinion towards vaccines. These factors help identify underlying themes and validate insights from the analysis.

Conclusions: The analyses of near real-time social media big data benefit public health authorities by enabling them to monitor public attitudes and opinions towards vaccine-related information in a geo-aware manner, address the concerns of vaccine skeptics, and promote the confidence that individuals within a certain region or community have towards vaccines.

(J Med Internet Res 2021;23(9):e30854) doi:10.2196/30854

KEYWORDS

Twitter; public opinion; COVID-19 vaccines; sentiment analysis; emotion analysis; topic modeling; COVID-19

Introduction

As of May 21, 2021, the COVID-19 pandemic had led to more than 160 million confirmed cases and more than 3 million deaths worldwide [1]. COVID-19 has continued to spread worldwide due to its highly contagious nature, diverse variants, and the mass public's inconsistent adherence to effective public health measures, such as wearing masks and maintaining social distance [2]. Meanwhile, the emergence of asymptomatic cases (which are difficult to detect) has become more frequent, potentially leading to a substantial accumulation in the number of infections over time [3]. As such, it is important to keep COVID-19 vaccines widely available and accessible [4].

Since January 2020, scientists and medical experts around the world have been developing and testing COVID-19 vaccines; 16 vaccines have been approved for emergency use around the world so far, but the progress of vaccination has been subject to hesitancy, distrust, and debate. Vaccine hesitancy was identified by the World Health Organization as one of the top 10 global health threats in 2019 [5]. In many countries, such hesitancy, along with vaccine misinformation, have presented substantial obstacles towards vaccinating a sufficient amount of the population in order to establish herd immunity [6,7].

Therefore, it is crucial for governments, public health officials, and policy makers to understand the potential drivers that affect public opinion towards COVID-19 vaccines [8]. A number of campaigns against antivaccination activists have been made through multiple channels since January 2020. Notably, the accelerated pace of vaccine development has further heightened public anxieties and could compromise the public's acceptance of the vaccine [9]. However, this acceptance varies across geographic contexts and the pandemic timeline. As governments put more effort into developing strategies for promoting vaccine acceptance and uptake, the key questions regarding the willingness to be vaccinated persist — what are the public's opinions and perceptions towards COVID-19 vaccines and what are the potential drivers that affect such opinions?

The internet and social media have provided rich user-generated data sources, in the form of infodemiology studies [10], in real time for performing public health surveillance [11]. Social media, especially Twitter, have been considered as major channels for the distribution of health information and opinion exchange, helping people to make intelligent decisions [12,13]. The analysis of big data derived from Twitter has been an emerging trend in recent COVID-19 vaccine–related studies. Geotagged tweets (hereinafter termed as geotweets) provide a rich volume of cost-effective content, including news, events, public comments, and the locational information of Twitter users. Through sentiment analysis and topic modeling methods that have been widely used in existing studies, qualitative tweet contents can be retrieved to reflect public opinions and attitudes towards COVID-19 vaccines. Additionally, users' location information enables researchers to investigate the spatiotemporal patterns of the public's opinions and attitudes. In general, existing studies have investigated people's reactions towards COVID-19 vaccines, with a geographical emphasis on the United States [14-19]. Some papers have also studied other countries in the world, including China [20], South Africa [21], Australia [22], the United Kingdom [14,23], Canada [24], and Africa [25], and to a global scale [26]. However, the study period of these works is relatively limited to or predominantly focused on the early stage of the pandemic or up to the end of 2020. None of these studies cover early 2021, the period of implementing mass systemic vaccine distribution. Furthermore, although sentiment analysis and topic modeling have been broadly applied, what remains less explored are the potential drivers that induce a change in public sentiment and opinion on vaccines, such as important events and announcements by political leaders (eg, the propaganda of vaccine success or vaccine conspiracy theories). There is a pressing need to investigate public opinion towards COVID-19 vaccines across a longer timeline and to explore the potential drivers that influence the change in such opinion over time.

To address these knowledge gaps, this study aimed to analyze the spatiotemporal patterns of public sentiment and emotion and explore the keywords and major topics of tweets regarding COVID-19 vaccines that were tweeted by Twitter users. Drawing on more than 300,000 geotweets from March 1, 2020 to February 28, 2021 in the United States, we employed sentiment and emotion analysis at both the national and state levels. We identified 3 phases along the pandemic timeline that display sharp changes in public sentiment and emotion. Using cloud mapping of keywords and topic modeling, we identified 11 key events and major topics as the potential drivers that induced such changes. Findings from this study can help governments, policymakers, and public health officials understand factors that motivate and cause hesitance in the public towards vaccination. With this understanding, these entities can better design potential interventions during their vaccination campaigns.

Methods

Data

Using the Twitter streaming Application Programming Interface (API), the Harvard Center for Geographic Analysis collected geotweets from March 1, 2020 to February 28, 2021. Geotweets provide the location information of user-defined places. If users

activate the GPS function in Twitter, their longitude and latitude are provided. We used the keyword "vaccin*" to query vaccine-related tweets, generating a total of 308,755 geotweets. In the results, 1.43% (44,118/308,755) of geotweets' geographic locations are at a state level (ie, Massachusetts, United States), and others are geocoded at a city level (ie, Cambridge, MA) or at a finer geographical level (ie, Uptown Coffee, Oxford, MS). We then conducted a series of data preprocessing of the geotweets' contents. First, we generalized the variations of COVID-related terms to "COVID-19," including "corona," "covid," "covid19," and "coronavirus"; second, we removed unrelated website links from the search results, including links starting with the fragment of "https"; third, we removed punctuation (eg, period, question mark, comma, colon, and ellipsis) and other key symbols (eg, bracket, single and double quotes) and converted capital letters into lower-case letters; fourth, we removed inflectional endings (eg, "ly") and reverted words to their root or dictionary form (eg, "peopl" from people, "dai" from daily, and "viru" from virus), by employing the *word lemmatization* function provided in the Python package Natural Language Toolkit 3.6.2 [27].

Methodology

To explore the spatiotemporal patterns of public sentiment and emotion towards COVID-19 vaccines, we conducted 4 sets of analyses, including sentiment analysis, emotion analysis, topic modeling, and word cloud mapping. For the sentiment analysis, we applied Valence Aware Dictionary for Sentiment Reasoning (VADER), a well-known rule-based model, to estimate sentiment compound scores [28]. The sentiment compound score is computed by summing the score of each word in the lexicon, adjusted according to the rules. The rules embody grammatical and syntactical conventions for expressing and emphasizing sentiment intensity. Then, the score is normalized to be between –1 (most extreme negative) and +1 (most extreme positive). To reclassify sentences as positive, neutral, or negative sentiment, threshold values are set as follows: A tweet with a compound score larger than 0.05 is classified as positive sentiment; a tweet with a compound score smaller than –0.05 is classified as negative sentiment; otherwise, it is classified as neutral sentiment [28]. We then cross-tabulated the 3 types of sentiment on daily and weekly bases with the number of geotweets. We generated line graphs at the national level and in the top 10 states with the largest number of geotweets.

Different from sentiment analysis, which detects positive, neutral, or negative feelings from tweet contents, emotion analysis aims to recognize the types of feelings more specifically through the content expression, such as anger, fear, and happiness. The emotion analysis of this study was performed based on the National Research Council Canada Lexicon (NRCLex) [29]. NRCLex examines 4 pairs of primary bipolar emotions: joy (feeling happy) versus sadness (feeling sad); anger (feeling angry) versus fear (feeling of being afraid); trust (stronger admiration and weaker acceptance) versus disgust (feeling something is wrong or nasty); and surprise (being unprepared for something) versus anticipation (looking forward positively to something). We then examined the temporal patterns of these 8 types of emotion at both national and state levels.

In order to investigate the potential drivers of such changes, we applied the Latent Dirichlet Allocation (LDA) model [30] to detect popular topics based on a certain number of key dates as the turning points of sentiment scores or with a sharp change in the number of geotweets. The LDA model generates automatic summaries of topics in terms of a discrete probability distribution over words for each topic and further infers per-document discrete distributions over topics [31]. Each topic is treated as a cluster, and each document is assigned to a cluster that represents its dominant topic. LDA is an unsupervised algorithm [32], meaning that, prior to running the model, users need to predefine the number of topics. To estimate the optimal number of topics, we used the Python package [33] and pyLDAvis [34] to compare the results with topic numbers from 3 to 10 and found that the smallest overlap among topics occurs when the topic number is 3. We further visualized the topic modeling results in bar graphs with the Y-axis, which indicates the top 10 keywords associated with that topic, and the X-axis, which shows the weight of each keyword (to reveal the extent to which a certain keyword contributes to that topic). Based on the top 10 most relevant keywords to each topic, we generalized and presented the name of each topic at the bottom of each graph.

We then categorized the study period into 3 phases based on 2 iconic events: the results of Phase 1 clinical trials by Moderna that were published in *The New England Journal of Medicine* on July 14, 2020 [35] and the first COVID-19 vaccine shots that were given in the United States on December 14, 2020 [36]. Phase 1, dating from March 1, 2020 to July 13, 2020, is the stage in which the public was waiting for official announcements regarding the effectiveness of COVID-19 vaccines; Phase 2, ranging from July 14, 2020 to December 13, 2020, is when the positive news of COVID-19 vaccine development began to arrive; and Phase 3 starts from December 14, 2020, when the first vaccine shots were given in the United States. We then aggregated sentiment scores at the state level and analyzed the changes in sentiment over the 3 phases in the top 10 states. Finally, we produced word cloud maps over the 3 predefined phases based on the frequency of keywords appearing in Tweet contents, with the size of a keyword reflecting its frequency and popularity.

Results

Sentiment Analysis and Topic Modeling

Figure 1 shows the overall trends in the weekly sentiment scores, unveiling the increased positive attitude towards COVID-19 vaccines within the study period. We identified 11 key dates as turning points in sentiment scores or in the number of geotweets. Correspondingly, a total of 33 topics on these 11 key dates are summarized and presented in Figures 2-4. In Phase 1, changes in the sentiment score were relatively stable, except for a sharp drop on June 21, 2020. This drop could have resulted from the misinformation and conspiracy theories related to Bill Gates. Vaccine-adverse conspiracy related to Gates claimed that the pandemic is a cover for his plan to implant trackable microchips made by Microsoft [37]. Topic modeling suggests that Gates

was referred to as "satan," "terrorist," and "evil" on that day (Figure 2).

In Phase 2, the first stimulus was observed on July 14, 2020, when the results of Phase 1 clinical trials by Moderna were published [35]. However, we did not observe a dramatic change in sentiment score until July 15, 2020, when Donald Trump tweeted "Great News on Vaccines!" [35]. Topic modeling suggests that keywords related to "good," "trial," "promis," and "test" were widely discussed on July 15, 2020 (Figure 2). Speculation suggests that, compared to key events in the development of COVID-19 vaccines, comments from public figures on vaccination could trigger bigger changes in public sentiment.

Figure 1. Sentiment scores and the number of geotweets over the entire study timeline at the national level.

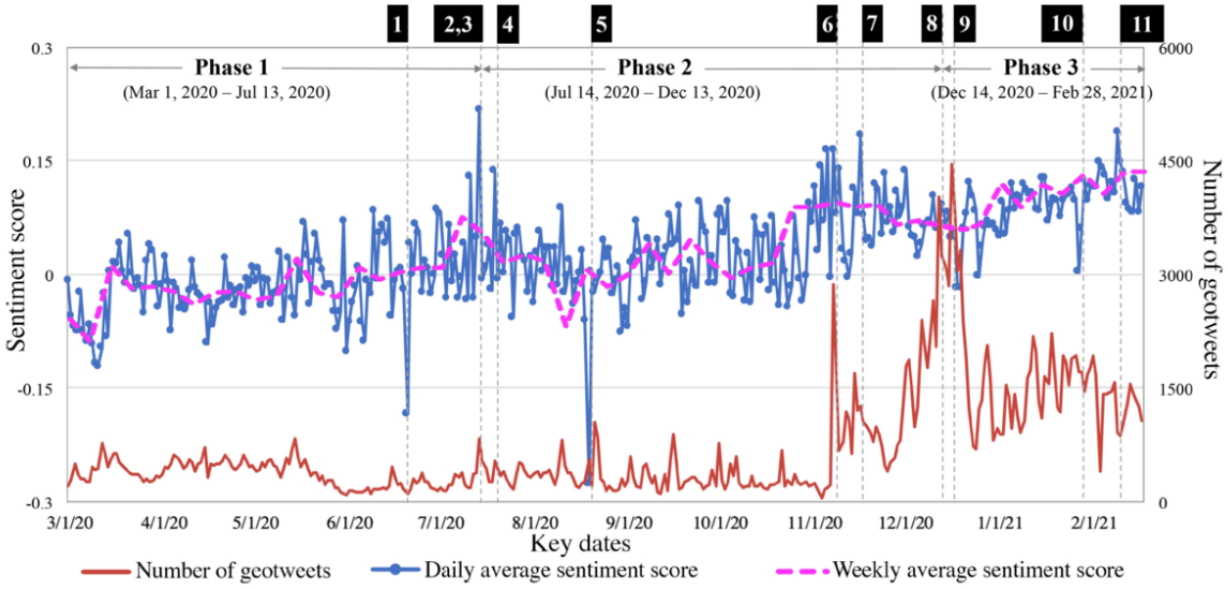
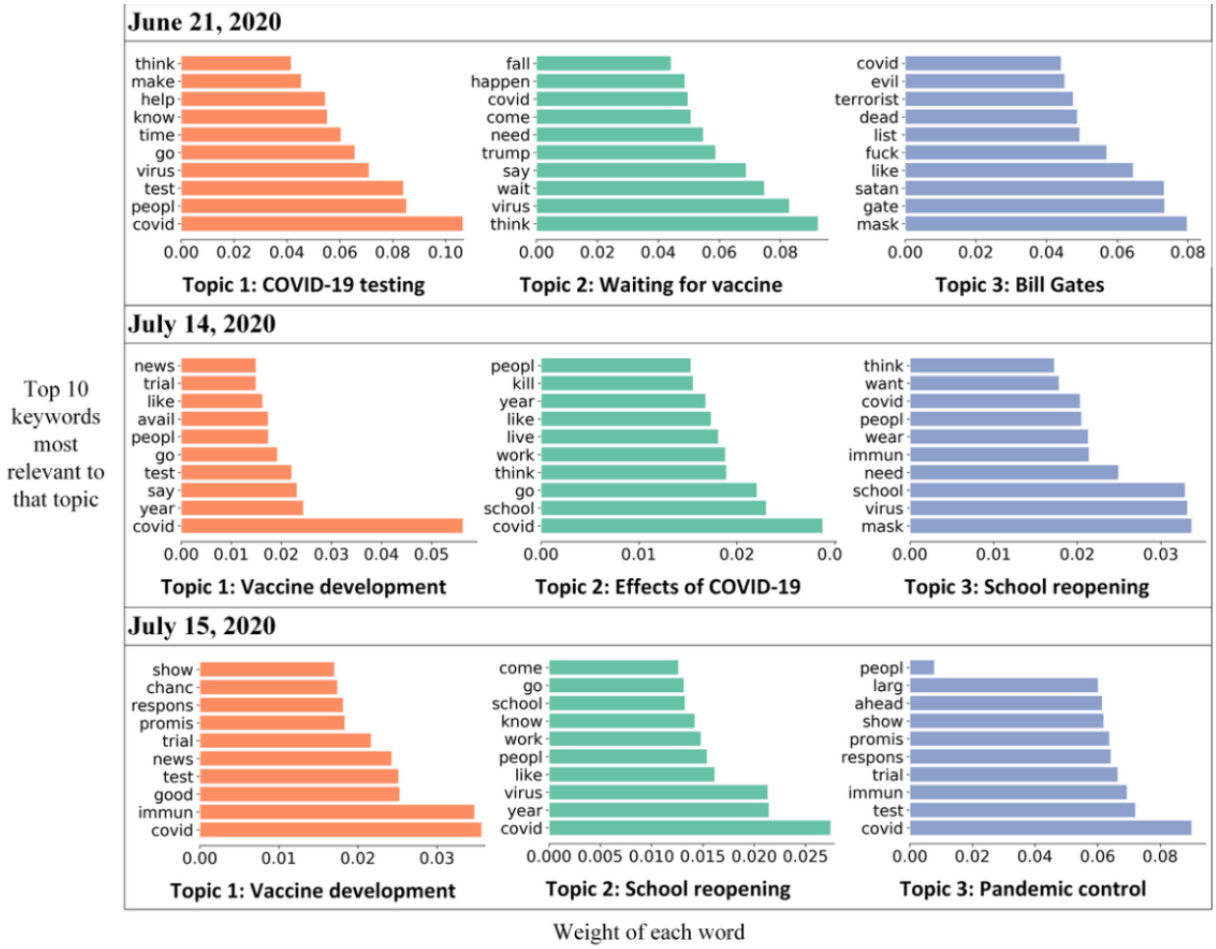
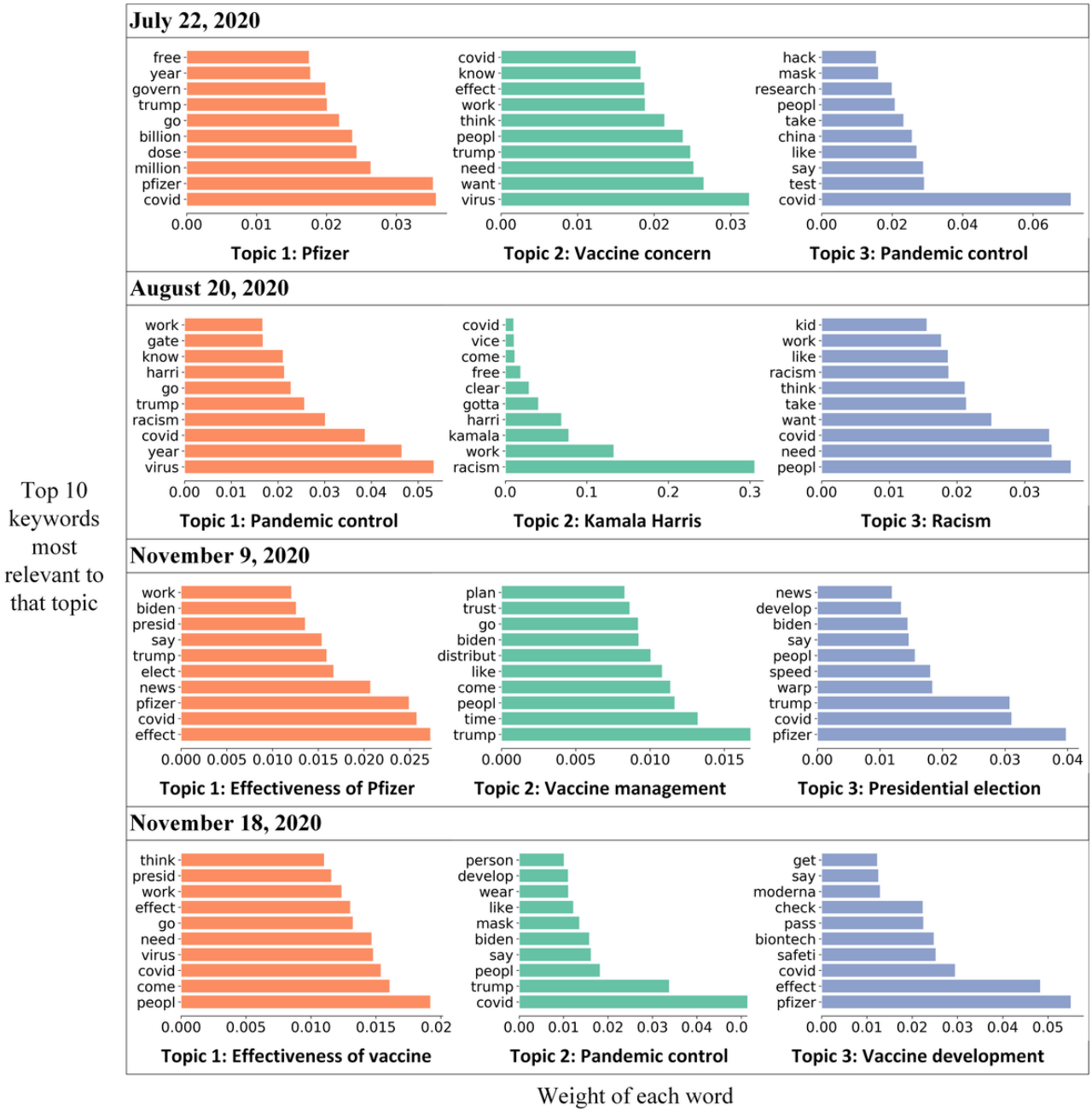


Figure 2. Three topics discussed on each of 3 key dates: June 21, 2020; July 14, 2020; and July 15, 2020.



Another sharp increase in sentiment score was observed on July 22, 2020, when the partnership between Pfizer and the US government accelerated the production and delivery of 100 million doses of COVID-19 vaccines [38]. The keywords "pfizer," "govern," and "million" were widely discussed and identified through topic modeling (Figure 3). On August 20, 2020, the sentiment score dropped dramatically after Kamala Harris formally accepted the Democrats' vice-presidential nomination at the 2020 Democratic National Convention. Harris advocated, "There is no vaccine for racism," mentioning the context of the racism protests for George Floyd and Breonna Taylor [39]. Of the keywords, "racism" and "kamala" were observed through topic modeling. Another increase in sentiment score appeared on November 9, 2020, when Pfizer announced that its vaccine is 90% effective (Figure 3) [40]. On the same day, Trump tweeted "STOCK MARKET UP BIG, VACCINE COMING SOON. REPORT 90% EFFECTIVE. SUCH GREAT NEWS!" Amid positive news from Pfizer, people questioned whether Pfizer purposefully released study results after Election Day, though Pfizer's CEO claimed that the release timing had nothing to do with politics [41]. On that day, widely discussed keywords included "trump," "pfizer," and "elect" (Figure 3).

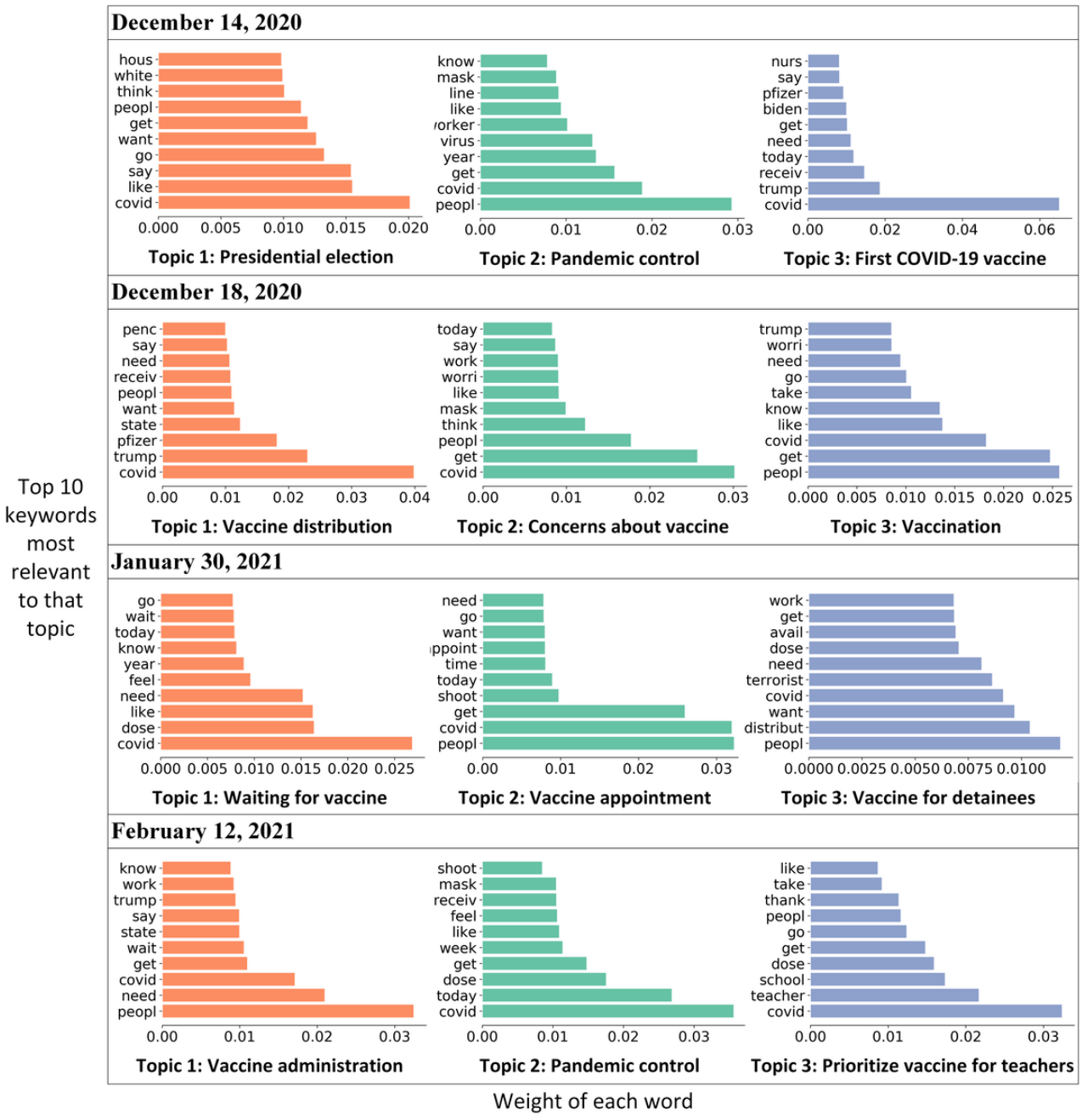
Figure 3. Three topics discussed on each of 4 key dates: July 22, 2020; August 20, 2020; November 9, 2020; and November 18, 2020.



In Phase 3 on December 14, 2020, an increased sentiment score was observed when an intensive care unit nurse received the first COVID-19 vaccine in New York. On the same day, the Electoral College voted to cement Biden's victory over Trump. Discussion regarding COVID-19 vaccines ("pfizer," "nurs," "receive") quickly increased on Twitter, while other related discussions regarding mask wearing ("wear" and "mask") and the presidential election ("house," "trump," "biden") remained popular (Figure 4). By December 18, 2020, the sentiment score remained high as both Pfizer and Moderna were authorized for

emergency use by the US Food and Drug Administration [42]. Trump tweeted "Moderna vaccine overwhelmingly approved. Distribution to start immediately." Additionally, the fact that former Vice President Pence and second lady Karen Pence received a COVID-19 vaccine [43] was widely discussed ("penc" and "receiv"). Expectations for the COVID-19 vaccines were also discussed ("need" and "want"; Figure 4). On January 30, 2021, the Department of Defense paused a plan to give COVID-19 vaccines to detainees in the Guantanamo Bay prison camp [44], which raised queries of COVID-19 vaccine delivery, leading to a moderate decrease in the sentiment score. Keywords were observed, including "terrorist" and "distribut" through topic modeling (Figure 4). On February 12, 2021, an increased sentiment score was observed after the Biden administration announced the purchase of 200 million COVID-19 vaccine doses from Pfizer and Moderna [45]. Discussion surrounding the administration of COVID-19 vaccines was extensive ("wait," "get," "need"; Figure 4). Topic modeling also suggests that complaints were pervasive ("teacher," "school," and "get"; Figure 4) because teachers were not prioritized for vaccination in states despite the Center for Disease Control and Prevention's recommendation.

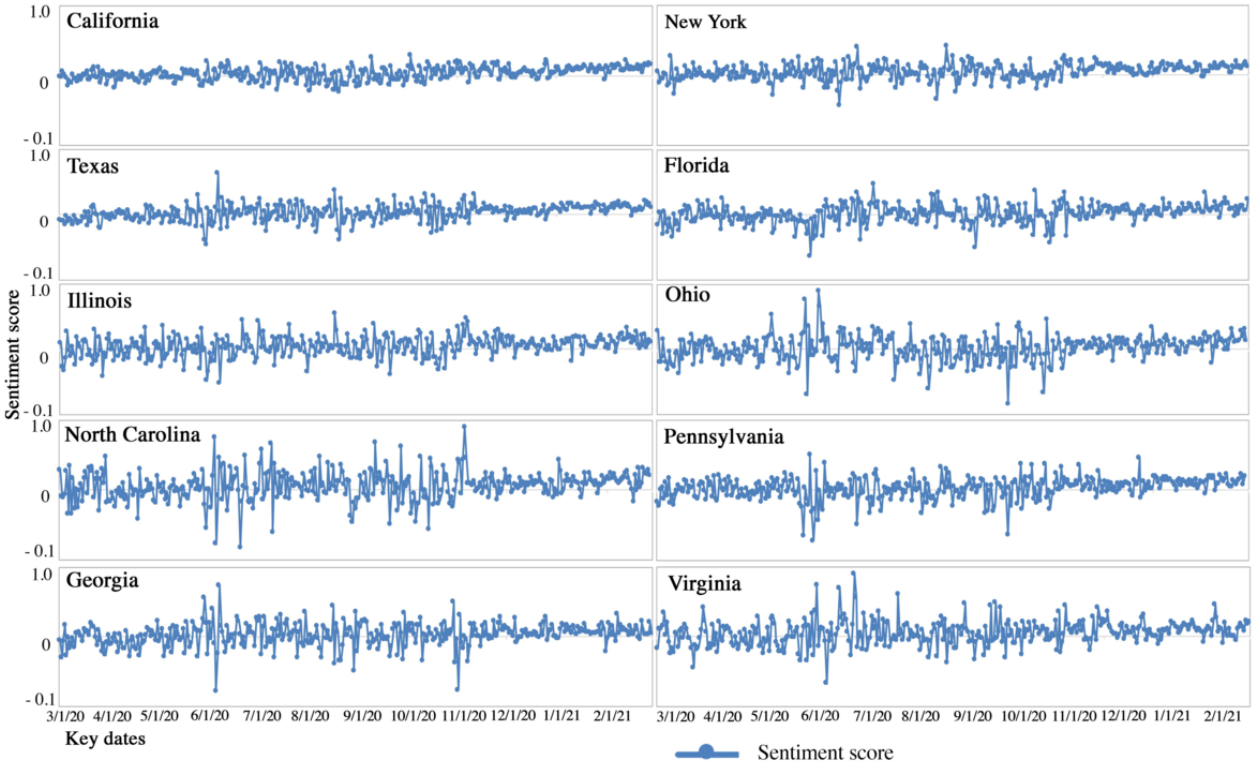
Figure 4. Three topics discussed on each of 4 key dates: December 14, 2020; December 18, 2020; January 30, 2021; and February 12, 2021.



We then broke down the sentiment scores by state in tandem, along with the pandemic timeline. We present the results in the top 10 states with the largest number of geotweets (Figure 5), including California, New York, Texas, Florida, Illinois, Ohio, North Carolina, Pennsylvania, Georgia, and Virginia. The temporal patterns in sentiment scores vary across states, with more obvious fluctuations before November 2020 in Illinois, Ohio, North Carolina, Georgia, Pennsylvania, and Virginia. A

number of sharp decreases in sentiment scores was observed in June 2020 in Illinois, North Carolina, Ohio, Pennsylvania, Georgia, and Virginia, in line with the tendency of sentiment drops at the national level. The states with relatively larger numbers of geotweets (ie, California, New York, Texas, and Florida) had more stable temporal trends and sentiment scores compared with the states with relatively smaller numbers of geotweets (eg, Ohio, North Carolina, Pennsylvania, Georgia, and Virginia).

Figure 5. Sentiment scores in 10 selected states.



We further examined the absolute values of the average positive and negative sentiment scores by states in Figure 6. In the majority of the states, the absolute positive sentiment score was larger than that of the negative sentiment score. The difference between the positive and negative sentiment scores was relatively more obvious in the mainland states of Alabama, Utah, Nebraska, Minnesota, and West Virginia (highlighted in dark grey in Figure 6), as well as in Hawaii and Alaska; the potential drivers triggering such differences across states may either relate to information or news spreading locally or be subject to the variations caused by the different sampling size in each state.

The changes in positive and negative sentiment scores over 2 periods of time (Phase 1 to Phase 2; Phase 2 to Phase 3) were compared and are presented in Figure 7. From Phase 1 to Phase 2, an increase in positive sentiment scores (orange bars) appeared in most states, most obviously in South Dakota, followed by North Dakota and Arkansas; meanwhile, a decrease in negative sentiment scores (dark blue bars) was also observed in the majority of states, most obviously in South Dakota and Rhode Island, followed by Montana, North Dakota, and Arkansas. From Phase 2 to Phase 3, the decrease in negative sentiment scores (light blue bars) appeared in most states, most obviously in Idaho and Rhode Island, followed by North Dakota, Vermont, and New Hampshire. However, the change in positive sentiment scores (red bars) from Phase 2 to Phase 3 varied across states, with a slight increase that is more obviously observed in Idaho, North Dakota, and New Mexico, while a slight decrease is more obviously observed in South Dakota, Rhode Island, and Connecticut. In addition, the magnitude of both positive and negative sentiment scores from Phase 1 to Phase 2 (the height of dark blue and orange bars) was more obvious in most states than that of Phase 2 to Phase 3 (the height of light blue and red bars). This indicates that the fluctuation in people's opinions towards vaccines became less obvious with the gradual development of vaccines and more encouraging news.

Figure 6. Absolute values of negative and positive sentiment scores at the state level.

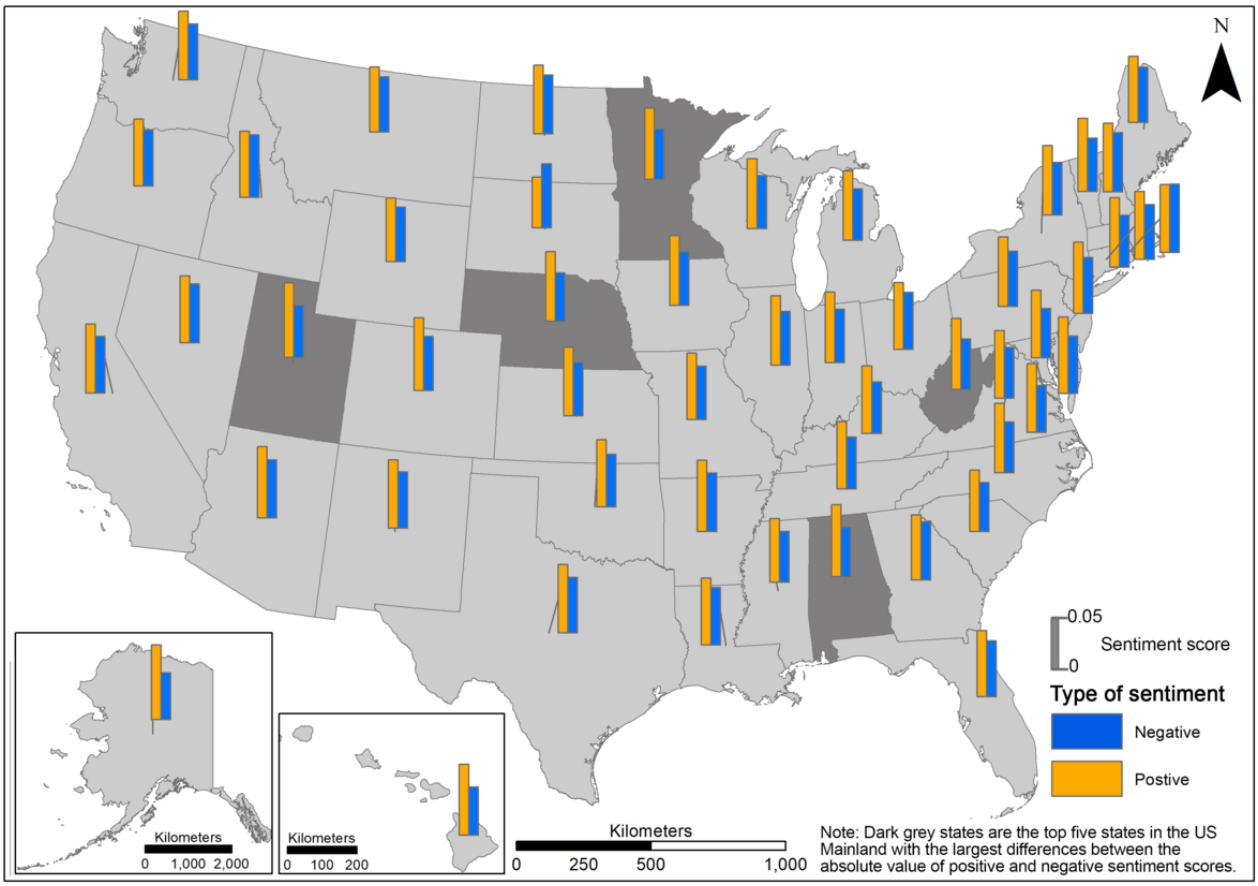
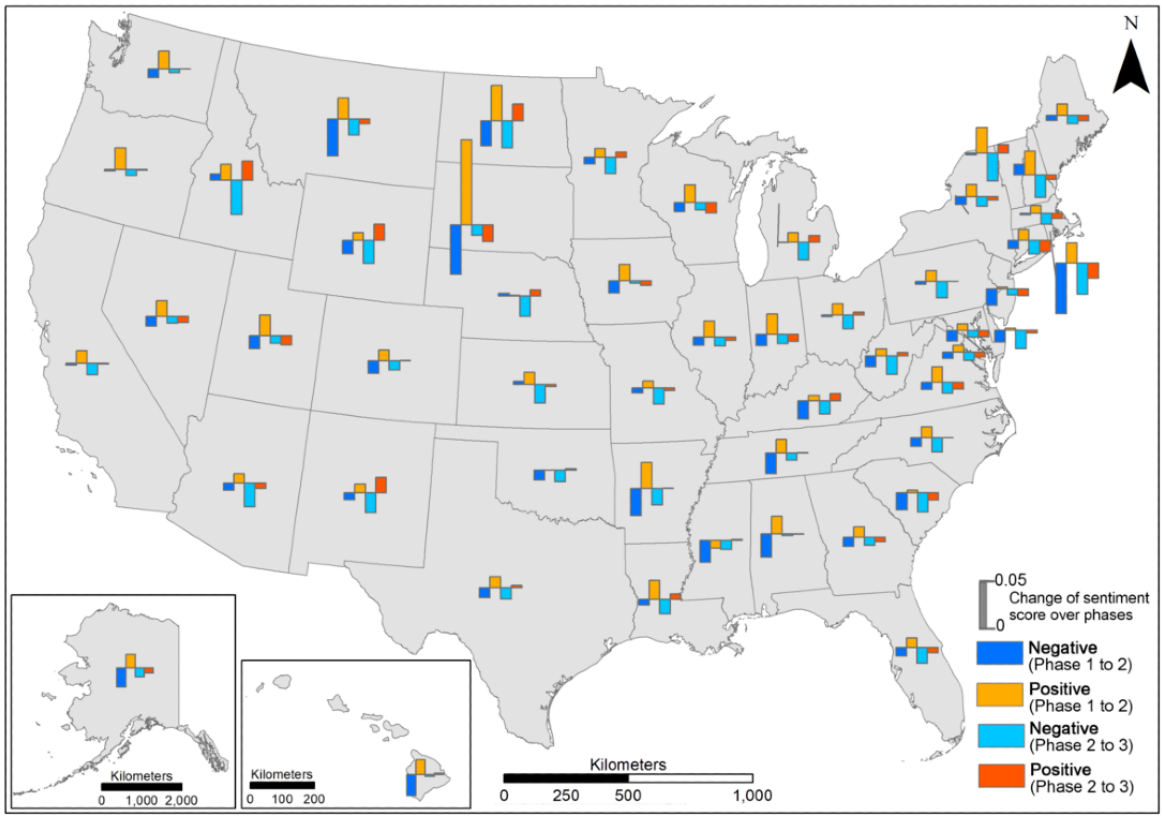


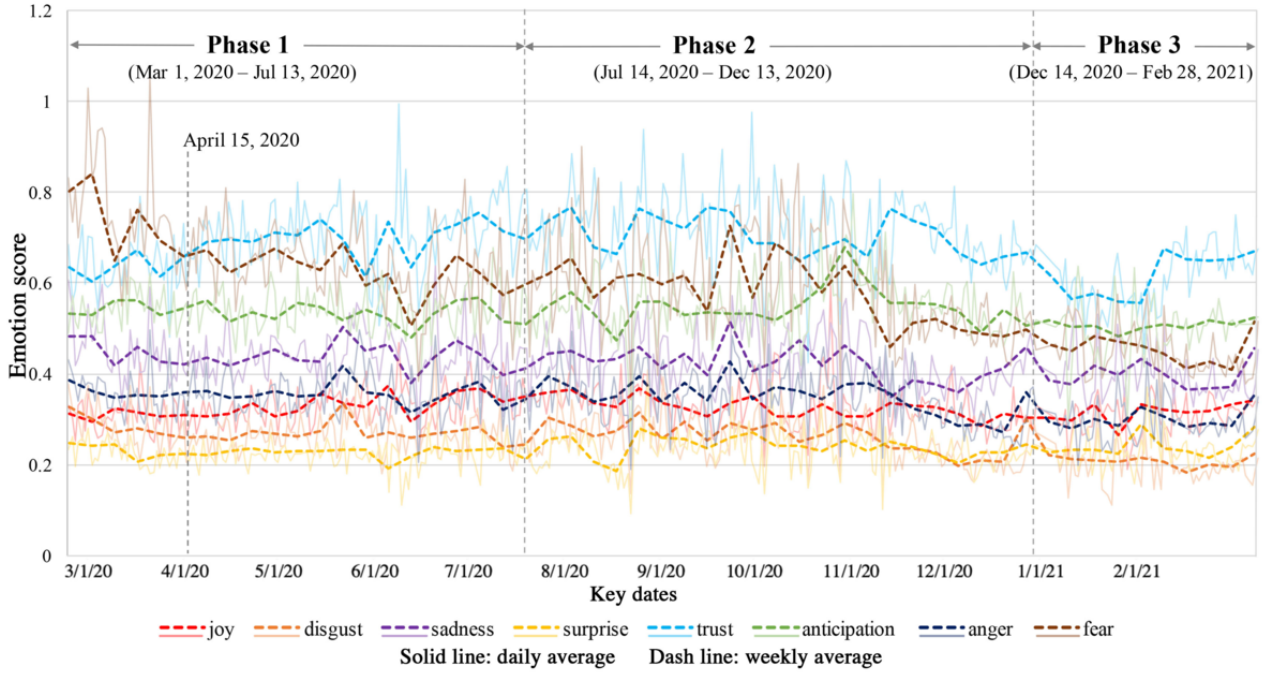
Figure 7. Changes in sentiment scores over the 3 phases at the state level.



Emotion Analysis

Figure 8 shows the temporal patterns in the 8 types of emotion, including joy, trust, anticipation, trust, surprise, disgust, sadness, and fear. Through the vertical comparison of the weekly average trend lines (dashed lines), we found that the emotion with the highest weekly average scores along the majority of the timeline was trust (blue dashed line), followed by fear, anticipation, sadness, anger, joy, disgust, and surprise. It is worth noting that the weekly average emotion score of fear was higher than that of trust before mid April 2020, possibly due to rapid COVID-19 infection and ineffective control of viral spread at the early stage of the pandemic. These events may have caused fear, uncertainty, or even feelings of panic [46]. Although fluctuations in emotion scores (eg, local peaks and valleys) can be found within each of the 8 emotions, the general trend implies that the public's trust in and anticipation towards vaccination were accompanied by a mixture of fear, sadness, and anger.

Figure 8. Average daily and weekly emotion scores over the entire study timeline at the national level.



We further investigated the relative distributions of 8 emotions in each state, as indicated by the percentage of emotion scores for each type with different colors (Figure 9). The overall patterns of the 8 emotions are consistent across most states. Throughout the entire timeline and in each of the 3 phases of the pandemic, trust was the dominant emotion towards vaccination over the full timeline of the pandemic. It was followed by anticipation, fear, sadness, anger, joy, disgust, and surprise. The state-level patterns largely align with the national pattern as depicted in Figure 9, although there are some exceptions, such as fear outweighing anticipation, joy, and trust (eg, Washington) and with fear, anger, and sadness outweighing other emotions (eg, Maine). As shown in Figures 10 and 11, the emotion of trust stayed consistent over time, while the changes in trends for other types of emotion were distinct across phases and by state.

We further compared the change in the percentage of emotions over 2 periods of time (Phase 1 to Phase 2; Phase 2 to Phase 3). From Phase 1 to Phase 2 (Figure 10), a decrease in fear (dark blue bars) was observed in most states, though its magnitude varied across states. This decrease was most obvious in South Dakota, followed by North Dakota, Arkansas, Mississippi, North Carolina, and South Carolina. The changes in anger, sadness, and disgust varied across states, with a general decrease in most states but sporadic increases in others (eg, Idaho, New Mexico, and New Hampshire). Furthermore, the combination of a decrease in fear and an increase in joy, trust, and anticipation was observed in most states except South Dakota. Throughout the period from Phase 2 to Phase 3 (Figure 11), it is difficult to generalize the pattern of emotion change across states in terms of type and magnitude. An increase in joy, trust, anticipation, and surprise along with a decrease in fear, anger, sadness, and disgust were the most notable (high bars) in Idaho and Rhode Island, followed by Missouri, Vermont, and New Hampshire. On the contrary, some states encountered a decrease in trust and anticipation in tandem with an increase in anger and sadness, including South Dakota, North Dakota, Montana, Kansas, Indiana, Maine, and Delaware. The complexity of emotion changes from Phase 2 to Phase 3 varied across states, reflecting the diversity in people's opinions and psychological reactions to vaccination, which should be subject to an in-depth investigation of causality.

Figure 9. Percentage of 8 emotions expressed at the state level.

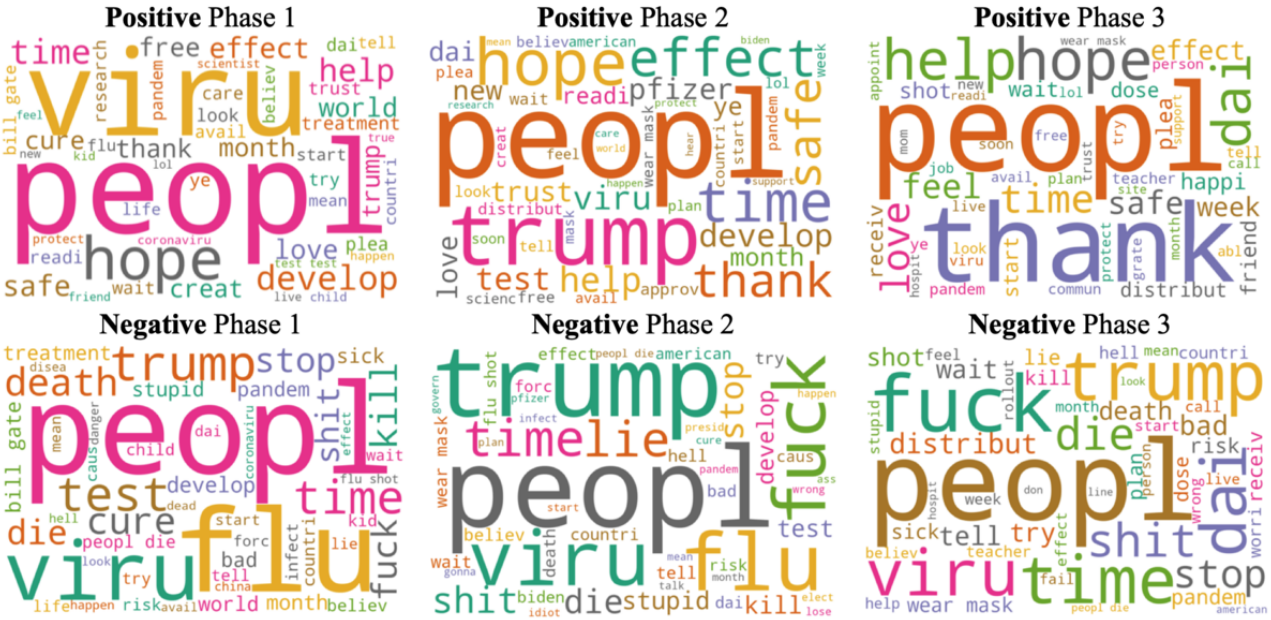
Figure 10. Change in emotions from Phase 1 to Phase 2 at the state level.

Figure 11. Changes in emotion from Phase 2 to Phase 3 at the state level.

Word Cloud Visualization

We produced word cloud mappings of 50 popular words associated with positive and negative sentiments over the 3 phases (Figure 12). The size of a word represents its popularity and the frequency with which it appears in tweets. Among the words associated with positive sentiment, the popular ones were “hope,” “help,” “thank,” “love,” “safe,” “cure,” and “free,” although the word “peopl,” with a more neutral nature, appears to be the most popular. Throughout the 3 phases, “hope,” “safe,” and “thank” grew larger from Phase 1 to Phase 3; in particular, “thank” became the most popular word in Phase 3. On the contrary, “flu,” “death,” “trump,” “fuck,” “lie,” “die,” “kill,” “shit,” and “stupid” were popular words associated with negative sentiment. Over the 3 phases, “flu” became smaller from Phase 1 to Phase 3 whereas “die,” “fuck,” “shit,” and “trump” evolved to be larger from Phase 1 to Phase 3; in particular, “trump” became predominant in Phase 2 possibly due to Trump’s increasing popularity caused by the 2020 Presidential Election. More specifically, while people were waiting for the news of COVID-19 vaccine development during Phase 1, their uncertainties on potential vaccines were reflected in the included keywords, which were related to the coronavirus and public’s frustration of the pandemic (eg, “viru,” “death,” “cure,” and “test”). Some keywords related to the COVID-19 vaccine were also observed, including “hope” and “develop.” Positive news about the development of COVID-19 vaccines appeared in Phase 2, which brought hope as well as misinformation regarding the vaccines to the public. At this stage, more specific information about COVID-19 vaccines was discussed (eg, “Pfizer,” “effect,” “risk,” “develop,” and “approve”), as compared to Phase 1. With Pfizer and Moderna vaccines approved during Phase 3, the public’s attention moved from vaccine development towards vaccine distribution (“distribution,” “wait,” and “free”), effectiveness (“safe” and “risk”), and priority (“teacher”). In all 3 phases, public figures (eg, “Trump,” “Biden,” and “Bill Gates”) contributed to hot topics with impacts on both positive and negative sentiments.

Figure 12. Popular keywords associated with positive and negative sentiments over 3 phases.



Discussion

Principal Findings

Drawing on geotweets from March 1, 2020 to February 28, 2021, this study examined public opinion on COVID-19 vaccines in the United States, by unveiling the spatiotemporal patterns of public sentiment and emotion over time, modeling the popular keywords and topics of Twitter contents, and analyzing the potential drivers of public opinion on vaccines. Our findings indicate that critical social or international events or announcements by political leaders and authorities may have potential impacts on public opinion towards COVID-19 vaccines. Such examples include the vaccine-adverse conspiracy related to Bill Gates on June 21, 2020, the tweet by Donald Trump of "Great News on Vaccines!" on July 14, 2020, Kamala Harris's advocacy of "There is no vaccine for racism" on August 20, 2020, Biden's victory of the presidential election over Trump on December 14, 2020, and the authorized emergent usage of Pfizer and Moderna on December 18, 2020. In the proposed 3 phases over the study timeframe, changes in public opinions on vaccines varied across space and time. More specifically, the fluctuation in people's sentimental response to the vaccine during the earlier stage of the pandemic was more obvious compared to that in the later stage of the pandemic. However, an increase in positive sentiment in parallel with a decrease in negative sentiment were generally observed in most states, reflecting the rising confidence and anticipation of the public towards COVID-19 vaccines. Furthermore, the public's 8 types of emotion towards the COVID-19 vaccine displayed a general trend of a combination of trust and anticipation with a mixture of fear, sadness, and anger. Moreover, the word cloud mapping showed that positive keywords including "hope," "safe," and "thank" grew larger from Phase 1 to Phase 3; in particular, "thank" became the most popular word in Phase 3, indicating the public's increasingly positive response towards vaccination. In all 3 phases, public figures (eg, "Trump," "Biden," and "Bill Gates") contributed to the most popular topics, impacting both positive and negative sentiments. The aforementioned findings reveal the diversity and complexity of people's perception on and their psychological reaction towards COVID-19 vaccines, which indicates a further need to be cautious in the interpretation of analytical outcomes and to initiate an additional in-depth investigation of the causality.

Our findings are partially supported by the current literature. Hussain et al [14] observed a marked increase in positive sentiment toward COVID-19 vaccines in the United States from March 1, 2020 to November 22, 2020. Guntuku et al [18] and Roy and Ghosh [47] found that Republican legislators became more engaged in public discussion on vaccine progress, which may have implications for COVID-19 vaccine uptake among their followers. Germani and Biller-Andorno [17] revealed that antivaccination supporters have been heavily engaged in discussions and dissemination of misinformation and conspiracy theories. Considering the limitations (ie, random sample) inherent in Twitter data, it is important to propose alternative data that provide a complementary understanding of public opinions towards the COVID-19 vaccine to promote vaccination in the United States.

Implications and Recommendations

The emergence of the internet and social media has provided new platforms for persuasion and the rapid spread of (mis)information, which leads to new opportunities for and challenges to the communication of vaccine information [48]. There are over 4.3 billion people using the internet nowadays, with 3.8 billion of these individuals as social media users [49]. The popularity of social media platforms coupled with the advent of digital detection strategies benefit public health authorities by enabling the monitoring of public sentiment towards vaccine-relevant information in a geo-aware, (near) real-time manner. This can inform more effective policymaking and promote participatory dialogue to establish confidence towards vaccines, in order to maximize vaccine uptake. Some of our findings add new value to the current scholarship and

also provide new insights and suggestions for policy implications with regard to safeguarding societal and economic health.

First, our findings indicate that public figures, especially politicians, play a crucial role in impacting the public's opinions on vaccination. Negative opinions expressed by public figures about a vaccine could impact a large population of people, especially those who do not hold an unswayable opinion [48]. People tend to believe public figures' opinions, as they are elected officials who can influence health care systems and are perceived to have more information about a vaccine [50,51]. Thus, public figures have a responsibility to disseminate accurate health information and should be cautious in expressing their opinions in public. This also highlights the necessity of considering the impact that public figures within vaccine campaigns have on upholding the public's confidence towards the concept of vaccination.

Second, our study reveals that vaccine-adverse conspiracy theories led to a sharp decline in sentiment scores. We need to be aware of the fact that social media platforms with a massive number of users, to some degree, "disrupted" traditional vaccine information communication [52], allowing antivaccination advocates to disseminate misleading messages to a certain audience, whose views on vaccination could be susceptible to change. However, it also means that governmental officials should consider using these platforms to communicate with individuals directly about vaccination via geotailored messages to address concerns specific to a certain region.

Third, different states demonstrated various trends in sentimental and emotional scores. Our geospatial analysis and map visualization [53] better portray more aspects of users' attitudes towards COVID-19 vaccines. This helps identify the areas with high negative sentimental and emotional scores that require further research to understand the public's underlying fears and concerns about COVID-19 vaccines. We also recommend government and public health agencies conduct COVID-19 vaccine campaigns in these areas to address people's fears and concerns about COVID-19 vaccines and provide guidance to access available vaccines.

Limitations and Future Work

Our study has several limitations that can be improved in future studies. First, the demographics of Twitter users is typically characterized by younger users who are avid users of mobile phone apps and the internet, and such users may not be able to reflect the opinion and perception of the general public with varying demographics and socioeconomic statuses [54,55]. In addition, the representativeness of Twitter users is not stationary but geographically varying [56,57]. Like other studies that rely on digital devices, the "digital divide" [58] issue needs to be acknowledged. This study only accounts for the reactions from Twitter users to vaccines, which, to some degree, neglect the underprivileged members of society (especially the poor and elderly), inhabitants of rural areas (who do not have access to digital devices), and those who are not willing to share their thoughts on social media platforms. Additionally, the Twitter API that we used allows access to approximately only 1% of the total records [59]. As Padilla et al [60] demonstrated, tweet sentiment can be impacted based on attraction visits throughout the course of a day. Hence, future work needs to increase the sample size to reduce the uncertainties and fluctuations of sentiment scores and emotions. Efforts are also needed to distinguish between local residents and visitors and also conduct investigations under finer temporal scales. In early 2021, Twitter released a new Twitter API (academic research product track) that grants free access to a full-archive search with enhanced features and functionality for researchers to obtain more precise, complete, and unbiased data for analyzing the public conversation [61]. Further efforts can be made to explore the potential of this new API in mining public opinions towards COVID-19 vaccines at a more granular scale. Since emotion is a complex and integrated product of human feelings [62], future research efforts can be put into exploring more diverse dimensions of emotion, on top of the 8 primary types of emotion. Moreover, disaster and crisis management includes 4 phases, namely prevention (capacity building), preparation (early warning), response (search, rescue, and emergency relief), and recovery (rehabilitation) [63]. Management of the COVID-19 pandemic is still in the response phase. For policy and decision-making endeavors that are pertinent to COVID-19 crisis management, it will be highly beneficial if researchers and practitioners continuously monitor emotional and perspective variations throughout the response and also extend the study timeline to the recovery phase or massive vaccination phase in the post-pandemic years. More importantly, to understand the impact of vaccination on countries, the workflow and methodology used in this study can be applied in multiple languages to global-scale geotweets.

Acknowledgments

This research was partially funded by the National University of Singapore Start-up Grant under WBS R-109-000-270-133 awarded to WL and NSF under Grant 1841403, 2027540, and 2028791. This research has also been partially supported by Ball State University Digital Fellowship Funding and the Faculty of Arts & Social Sciences Staff Research Support Scheme FY2021 of National University of Singapore (WBS: C-109-000-222-091).

Conflicts of Interest

None declared.

References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases 2020 May 19;20(5):533-534. [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
2. Brown TS, Walensky RP. Serosurveillance and the COVID-19 Epidemic in the US: Undetected, Uncertain, and Out of Control. JAMA 2020 Aug 25;324(8):749-751. [doi: 10.1001/jama.2020.14017] [Medline: 32692350]
3. Nishiura H, Kobayashi T, Miyama T, Suzuki A, Jung S, Hayashi K, et al. Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). Int J Infect Dis 2020 May;94:154-155 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.020] [Medline: 32179137]
4. Haynes BF. A New Vaccine to Battle Covid-19. N Engl J Med 2021 Feb 04;384(5):470-471 [FREE Full text] [doi: 10.1056/NEJMe2035557] [Medline: 33378607]
5. Ten threats to global health in 2019. World Health Organization. URL: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 [accessed 2021-05-30]
6. Lazarus JV, Ratzan SC, Palayew A, Gostin LO, Larson HJ, Rabin K, et al. A global survey of potential acceptance of a COVID-19 vaccine. Nat Med 2021 Feb;27(2):225-228 [FREE Full text] [doi: 10.1038/s41591-020-1124-9] [Medline: 33082575]
7. Lane S, MacDonald NE, Marti M, Dumolard L. Vaccine hesitancy around the globe: Analysis of three years of WHO/UNICEF Joint Reporting Form data-2015-2017. Vaccine 2018 Jun 18;36(26):3861-3867 [FREE Full text] [doi: 10.1016/j.vaccine.2018.03.063] [Medline: 29605516]
8. Cornwall W. Officials gird for a war on vaccine misinformation. Science 2020 Jul 03;369(6499):14-15. [doi: 10.1126/science.369.6499.14] [Medline: 32631873]
9. Fadda M, Albanese E, Suggs L. When a COVID-19 vaccine is ready, will we all be ready for it? Int J Public Health 2020 Jul;65(6):711-712 [FREE Full text] [doi: 10.1007/s00038-020-01404-4] [Medline: 32529534]
10. Zeraatkar K, Ahmadi M. Trends of infodemiology studies: a scoping review. Health Info Libr J 2018 Jun 04;35(2):91-120 [FREE Full text] [doi: 10.1111/hir.12216] [Medline: 29729073]
11. Luo W, MacEachren AM. Geo-social visual analytics. JOSIS 2014 Jun 20;8:27-66. [doi: 10.5311/JOSIS.2014.8.139]
12. Kwak H, Lee C, Park H, Moon S. What is Twitter, a social network or a news media? 2010 Presented at: 19th international conference on World wide web; April 26-30, 2010; Raleigh, NC. [doi: 10.1145/1772690.1772751]
13. Paul M, Dredze M. You Are What You Tweet: Analyzing Twitter for Public Health. 2011 Presented at: Fifth International AAAI Conference on Weblogs and Social Media; July 17–21, 2011; Barcelona, Catalonia, Spain.
14. Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, et al. Artificial Intelligence-Enabled Analysis of Public Attitudes on Facebook and Twitter Toward COVID-19 Vaccines in the United Kingdom and the United States: Observational Study. J Med Internet Res 2021 Apr 05;23(4):e26627 [FREE Full text] [doi: 10.2196/26627] [Medline: 33724919]
15. Jang H, Rempel E, Roth D, Carenini G, Janjua NZ. Tracking COVID-19 Discourse on Twitter in North America: Infodemiology Study Using Topic Modeling and Aspect-Based Sentiment Analysis. J Med Internet Res 2021 Feb 10;23(2):e25431 [FREE Full text] [doi: 10.2196/25431] [Medline: 33497352]
16. Engel-Rebitzer E, Stokes DC, Buttenheim A, Purtle J, Meisel ZF. Changes in legislator vaccine-engagement on Twitter before and after the arrival of the COVID-19 pandemic. Hum Vaccin Immunother 2021 May 10:1-5. [doi: 10.1080/21645515.2021.1911216] [Medline: 33970786]
17. Germani F, Biller-Andorno N. The anti-vaccination infodemic on social media: A behavioral analysis. PLoS One 2021;16(3):e0247642 [FREE Full text] [doi: 10.1371/journal.pone.0247642] [Medline: 33657152]
18. Guntuku SC, Purtle J, Meisel ZF, Merchant RM, Agarwal A. Partisan Differences in Twitter Language Among US Legislators During the COVID-19 Pandemic: Cross-sectional Study. J Med Internet Res 2021 Jun 03;23(6):e27300 [FREE Full text] [doi: 10.2196/27300] [Medline: 33939620]
19. Roy KC, Hasan S. Modeling the dynamics of hurricane evacuation decisions from twitter data: An input output hidden markov modeling approach. Transportation Research Part C: Emerging Technologies 2021 Feb;123:102976. [doi: 10.1016/j.trc.2021.102976]
20. Wang J, Zhou Y, Zhang W, Evans R, Zhu C. Concerns Expressed by Chinese Social Media Users During the COVID-19 Pandemic: Content Analysis of Sina Weibo Microblogging Data. J Med Internet Res 2020 Nov 26;22(11):e22152 [FREE Full text] [doi: 10.2196/22152] [Medline: 33151894]
21. Mutanga MB, Abayomi A. Tweeting on COVID-19 pandemic in South Africa: LDA-based topic modelling approach. African Journal of Science, Technology, Innovation and Development 2020 Oct 08:1-10. [doi: 10.1080/20421338.2020.1817262]
22. Kwok SWH, Vadde SK, Wang G. Tweet Topics and Sentiments Relating to COVID-19 Vaccination Among Australian Twitter Users: Machine Learning Analysis. J Med Internet Res 2021 May 19;23(5):e26953 [FREE Full text] [doi: 10.2196/26953] [Medline: 33886492]
23. Cotfas L, Delcea C, Roxin I, Ioanas C, Gherai DS, Tajariol F. The Longest Month: Analyzing COVID-19 Vaccination Opinions Dynamics From Tweets in the Month Following the First Vaccine Announcement. IEEE Access 2021;9:33203-33223. [doi: 10.1109/access.2021.3059821]

24. Griffith J, Marani H, Monkman H. COVID-19 Vaccine Hesitancy in Canada: Content Analysis of Tweets Using the Theoretical Domains Framework. J Med Internet Res 2021 Apr 13;23(4):e26874 [FREE Full text] [doi: 10.2196/26874] [Medline: 33769946]
25. Gbashi S, Adebo OA, Doorsamy W, Njobeh PB. Systematic Delineation of Media Polarity on COVID-19 Vaccines in Africa: Computational Linguistic Modeling Study. JMIR Med Inform 2021 Mar 16;9(3):e22916 [FREE Full text] [doi: 10.2196/22916] [Medline: 33667172]
26. Chang C, Monselise M, Yang CC. What Are People Concerned About During the Pandemic? Detecting Evolving Topics about COVID-19 from Twitter. J Healthc Inform Res 2021 Jan 17:1-28 [FREE Full text] [doi: 10.1007/s41666-020-00083-3] [Medline: 33490856]
27. Bird S, Klein E, Loper E. Natural language processing with Python: analyzing text with the natural language toolkit. 2019. URL: https://www.nltk.org/book/ [accessed 2021-08-14]
28. Hutto CJ, Gilbert E. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. 2014 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 1–4, 2014; Ann Arbor, MI.
29. Mohammad S, Turney P. NRC emotion lexicon. Saif Mohammad. 2013 Nov 15. URL: http://www.saifmohammad.com/WebDocs/NRCemotionlexicon.pdf [accessed 2021-08-14]
30. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. The Journal of Machine Learning Research 2003 Mar 01;3:993-1022. [doi: 10.5555/944919.944937]
31. Hu T, She B, Duan L, Yue H, Clunis J. A Systematic Spatial and Temporal Sentiment Analysis on Geo-Tweets. IEEE Access 2020;8:8658-8667. [doi: 10.1109/access.2019.2961100]
32. Yan Y, Feng C, Huang W, Fan H, Wang Y, Zipf A. Volunteered geographic information research in the first decade: a narrative review of selected journal articles in GIScience. International Journal of Geographical Information Science 2020 Feb 26;34(9):1765-1791. [doi: 10.1080/13658816.2020.1730848]
33. Řehůřek R. GENSIM: topic modelling for humans. URL: https://radimrehurek.com/gensim/ [accessed 2021-05-30]
34. bmabey / pyLDAvis. GitHub. 2014 Mar 24. URL: https://github.com/bmabey/pyLDAvis [accessed 2021-05-30]
35. Jackson LA, Anderson EJ, Rouphael NG, Roberts PC, Makhene M, Coler RN, et al. An mRNA Vaccine against SARS-CoV-2 — Preliminary Report. N Engl J Med 2020 Nov 12;383(20):1920-1931. [doi: 10.1056/nejmoa2022483]
36. Guarino B, Cha AE, Wood J, Witte G. 'The weapon that will end the war': First coronavirus vaccine shots given outside trials in U.S. The Washington Post. URL: https://www.washingtonpost.com/nation/2020/12/14/first-covid-vaccines-new-york/ [accessed 2021-05-30]
37. Goodman J, Carmichael F. Coronavirus: Bill Gates 'microchip' conspiracy theory and other vaccine claims fact-checked. BBC News. 2020 May 30. URL: https://www.bbc.com/news/52847648 [accessed 2021-08-14]
38. U.S. Government Engages Pfizer to Produce Millions of Doses of COVID-19 Vaccine. U.S. Department of Health & Human Services. 2020 Jul 22. URL: https://www.hhs.gov/about/news/2020/07/22/us-government-engages-pfizer-produce-millions-doses-covid-19-vaccine.html [accessed 2021-05-30]
39. Nagourney A, Glueck K. Kamala Harris Takes the Spotlight, a Moment for Her and History. New York Times. 2020 Aug 19. URL: https://www.nytimes.com/2020/08/19/us/politics/kamala-harris-dnc.html [accessed 2021-05-30]
40. Callaway E. What Pfizer's landmark COVID vaccine results mean for the pandemic. Nature. 2020 Nov 9. URL: https://www.nature.com/articles/d41586-020-03166-8 [accessed 2021-05-30]
41. Benveniste A. Pfizer CEO: Our vaccine timing had nothing to do with politics. CNN. 2020 Nov 09. URL: https://www.cnn.com/2020/11/09/business/pfizer-covid-vaccine/index.html [accessed 2021-05-30]
42. Pfizer-BioNTech COVID-19 Vaccine. U.S. Food & Drug Administration. 2020 Dec 11. URL: https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/pfizer-biontech-covid-19-vaccine [accessed 2021-05-30]
43. Shabad R. Pence receives Covid vaccine in televised appearance, hails "medical miracle". NBC News. 2020 Dec 18. URL: https://www.nbcnews.com/politics/white-house/pence-set-receive-covid-vaccine-televi3ed-appearance-n1251655 [accessed 2021-05-30]
44. Herridge C, McDonald C. Department of Defense pauses plan to give COVID-19 vaccine to Guantanamo detainees. CBS News. 2021 Jan 30. URL: https://www.cbsnews.com/news/dod-pauses-plan-to-give-covid-vaccine-to-guantanamo-detainees/ [accessed 2021-05-30]
45. Biden Administration purchases additional doses of COVID-19 vaccines from Pfizer and Moderna. U.S. Department of Health & Human Services. 2021 Feb 11. URL: https://www.hhs.gov/about/news/2021/02/11/biden-administration-purchases-additional-doses-covid-19-vaccines-from-pfizer-and-moderna.html [accessed 2021-05-30]
46. Wang S, Liu Y, Hu T. Examining the Change of Human Mobility Adherent to Social Restriction Policies and Its Effect on COVID-19 Cases in Australia. Int J Environ Res Public Health 2020 Oct 29;17(21):1 [FREE Full text] [doi: 10.3390/ijerph17217930] [Medline: 33137958]
47. Roy S, Ghosh P. A Comparative Study on Distancing, Mask and Vaccine Adoption Rates from Global Twitter Trends. Healthcare (Basel) 2021 Apr 21;9(5):1 [FREE Full text] [doi: 10.3390/healthcare9050488] [Medline: 33919097]

48. Kang GJ, Ewing-Nelson SR, Mackey L, Schlitt JT, Marathe A, Abbas KM, et al. Semantic network analysis of vaccine sentiment in online social media. Vaccine 2017 Jun 22;35(29):3621-3638 [FREE Full text] [doi: 10.1016/j.vaccine.2017.05.052] [Medline: 28554500]
49. Digital in 2020. We Are Social. URL: https://wearesocial.com/digital-2020 [accessed 2021-05-30]
50. Zhang EJ, Chughtai AA, Heywood A, MacIntyre CR. Influence of political and medical leaders on parental perception of vaccination: a cross-sectional survey in Australia. BMJ Open 2019 Mar 26;9(3):e025866 [FREE Full text] [doi: 10.1136/bmjopen-2018-025866] [Medline: 30914408]
51. Abu-Akel A, Spitz A, West R. The effect of spokesperson attribution on public health message sharing during the COVID-19 pandemic. PLoS One 2021 Feb 3;16(2):e0245100 [FREE Full text] [doi: 10.1371/journal.pone.0245100] [Medline: 33534800]
52. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Hum Vaccin Immunother 2020 Nov 01;16(11):2586-2593 [FREE Full text] [doi: 10.1080/21645515.2020.1780846] [Medline: 32693678]
53. Mocnik F, Raposo P, Feringa W, Kraak M, Köbben B. Epidemics and pandemics in maps – the case of COVID-19. Journal of Maps 2020 Jun 18;16(1):144-152. [doi: 10.1080/17445647.2020.1776646]
54. Wilson K, Atkinson K, Deeks S. Opportunities for utilizing new technologies to increase vaccine confidence. Expert Rev Vaccines 2014 Aug;13(8):969-977. [doi: 10.1586/14760584.2014.928208] [Medline: 24931799]
55. Karami A, Kadari RR, Panati L, Nooli SP, Bheemreddy H, Bozorgi P. Analysis of Geotagging Behavior: Do Geotagged Users Represent the Twitter Population? IJGI 2021 Jun 02;10(6):373. [doi: 10.3390/ijgi10060373]
56. Gore RJ, Diallo S, Padilla J. You Are What You Tweet: Connecting the Geographic Variation in America's Obesity Rate to Twitter Content. PLoS One 2015 Sep 2;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
57. Jiang Y, Li Z, Ye X. Understanding demographic and socioeconomic biases of geotagged Twitter users at the county level. Cartography and Geographic Information Science 2018 Feb 09;46(3):228-242. [doi: 10.1080/15230406.2018.1434834]
58. Huang X, Li Z, Jiang Y, Li X, Porter D. Twitter reveals human mobility dynamics during the COVID-19 pandemic. PLoS One 2020 Nov 10;15(11):e0241957 [FREE Full text] [doi: 10.1371/journal.pone.0241957] [Medline: 33170889]
59. Yan Y, Chen J, Wang Z. Mining public sentiments and perspectives from geotagged social media data for appraising the post-earthquake recovery of tourism destinations. Applied Geography 2020 Oct;123:102306. [doi: 10.1016/j.apgeog.2020.102306]
60. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018 Jun 14;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]
61. Yan Y, Kuo C, Feng C, Huang W, Fan H, Zipf A. Coupling maximum entropy modeling with geotagged social media data to determine the geographic distribution of tourists. International Journal of Geographical Information Science 2018 Apr 20;32(9):1699-1736. [doi: 10.1080/13658816.2018.1458989]
62. Twitter API: Academic Research product track. Twitter Developer. URL: https://developer.twitter.com/en/products/twitter-api/academic-research [accessed 2021-05-30]
63. Plutchik R. The Nature of Emotions. Am. Sci 2001;89(4):344. [doi: 10.1511/2001.4.344]

Abbreviations

API: application programming interface
LDA: Latent Dirichlet Allocation
NRCLex: National Research Council Canada Lexicon
VADER: Valence Aware Dictionary for Sentiment Reasoning

Edited by C Basch; submitted 01.06.21; peer-reviewed by FB Mocnik, P Luz, ECY Su, R Zhang, C Lynch; comments to author 20.06.21; revised version received 12.07.21; accepted 26.07.21; published 10.09.21.

Please cite as:
Hu T, Wang S, Luo W, Zhang M, Huang X, Yan Y, Liu R, Ly K, Kacker V, She B, Li Z
Revealing Public Opinion Towards COVID-19 Vaccines With Twitter Data in the United States: Spatiotemporal Perspective
J Med Internet Res 2021;23(9):e30854
URL: https://www.jmir.org/2021/9/e30854
doi:10.2196/30854
PMID:34346888

Original Paper

Factors Associated With Longitudinal Psychological and Physiological Stress in Health Care Workers During the COVID-19 Pandemic: Observational Study Using Apple Watch Data

Robert P Hirten[1,2], MD; Matteo Danieletto[2], PhD; Lewis Tomalin[3], PhD; Katie Hyewon Choi[3], MS; Micol Zweig[2], MPH; Eddye Golden[2], MPH; Sparshdeep Kaur[2], BBA; Drew Helmus[1], MPH; Anthony Biello[1], BA; Renata Pyzik[4], MS; Claudia Calcagno[4], MD, PhD; Robert Freeman[5], MSN; Bruce E Sands[1], MD; Dennis Charney[6,7], MD; Erwin P Bottinger[2], MD; James W Murrough[7,8], MD; Laurie Keefer[1], PhD; Mayte Suarez-Farinas[3], PhD; Girish N Nadkarni[2,9], MD; Zahi A Fayad[4], PhD

[1]The Dr Henry D Janowitz Division of Gastroenterology, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[2]The Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[3]Department of Population Health Science and Policy, Center for Biostatistics, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[4]Department of Diagnostic, Molecular and Interventional Radiology, The BioMedical Engineering and Imaging Institute, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[5]Institute for Healthcare Delivery Science, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[6]Office of the Dean, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[7]Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[8]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[9]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States

Corresponding Author:
Robert P Hirten, MD
The Dr Henry D Janowitz Division of Gastroenterology
Icahn School of Medicine at Mount Sinai
1 Gustave L Levy Place
New York, NY, 10029
United States
Phone: 1 212 241 0150
Email: robert.hirten@mountsinai.org

Abstract

Background: The COVID-19 pandemic has resulted in a high degree of psychological distress among health care workers (HCWs). There is a need to characterize which HCWs are at an increased risk of developing psychological effects from the pandemic. Given the differences in the response of individuals to stress, an analysis of both the perceived and physiological consequences of stressors can provide a comprehensive evaluation of its impact.

Objective: This study aimed to determine characteristics associated with longitudinal perceived stress in HCWs and to assess whether changes in heart rate variability (HRV), a marker of autonomic nervous system function, are associated with features protective against longitudinal stress.

Methods: HCWs across 7 hospitals in New York City, NY, were prospectively followed in an ongoing observational digital study using the custom Warrior Watch Study app. Participants wore an Apple Watch for the duration of the study to measure HRV throughout the follow-up period. Surveys measuring perceived stress, resilience, emotional support, quality of life, and optimism were collected at baseline and longitudinally.

Results: A total of 361 participants (mean age 36.8, SD 10.1 years; female: n=246, 69.3%) were enrolled. Multivariate analysis found New York City's COVID-19 case count to be associated with increased longitudinal stress (P=.008). Baseline emotional support, quality of life, and resilience were associated with decreased longitudinal stress (P<.001). A significant reduction in stress during the 4-week period after COVID-19 diagnosis was observed in the highest tertial of emotional support (P=.03) and resilience (P=.006). Participants in the highest tertial of baseline emotional support and resilience had a significantly different circadian pattern of longitudinally collected HRV compared to subjects in the low or medium tertial.

Conclusions: High resilience, emotional support, and quality of life place HCWs at reduced risk of longitudinal perceived stress and have a distinct physiological stress profile. Our findings support the use of these characteristics to identify HCWs at risk of the psychological and physiological stress effects of the pandemic.

(J Med Internet Res 2021;23(9):e31295) doi:10.2196/31295

KEYWORDS

wearable device; COVID-19; stress; heart rate variability; psychological; psychology; physiology; mental health; health care worker; observational; app; heart rate; nervous system; resilience; emotion; support; quality of life

Introduction

Increasing rates of SARS-CoV-2 infections and hospitalizations, growing workloads, and concern regarding personal protective equipment have resulted in a large psychological burden on health care workers (HCWs) [1]. Prior pandemics have had psychological effects on HCWs, increasing posttraumatic stress, depression, and anxiety [1-3]. However, the scale and duration of the COVID-19 pandemic have further amplified the risk of these adverse outcomes. Cross-sectional studies have demonstrated that frontline HCWs are at a high risk of depression, anxiety, insomnia, and distress compared to the general population [4-6]. HCWs working in wards serving patients with COVID-19 reported higher levels of stress, exhaustion, depressive mood, and burnout [7,8]. However, there is limited longitudinal data on the pandemic's psychological impact on this group and across health care occupations, no means to identify which HCWs are at risk of developing psychological sequela over time, and no objective evaluation of the stress response among HCWs. Identification of at-risk HCWs will allow for the appropriate allocation of mental health resources.

Advances in digital technology provide a means to address these limitations. Smartphone apps can be used to administer surveys and integrate wearable devices, such as the Apple Watch, to monitor the autonomic nervous system (ANS), a primary component of the stress response. ANS function can be ascertained through measurement of heart rate variability (HRV), a measure of the parasympathetic and sympathetic nervous systems' impact on cardiac contractility through calculation of changes in the beat-to-beat intervals [9]. Systems such as the ANS promote adaption to stressors. However, the cumulative burden of stress, or allostatic load, can alter and impair the response of these systems and result in deleterious physical effects on processes ranging from immune function to cardiovascular health. Given the large differences in the response of individuals to stress, an evaluation of both the perceived and physiological consequences of the stressor could provide a more comprehensive overview of its impact and an understanding of who is at risk from the deleterious physical effects of chronic stress [10]. In response to the COVID-19 pandemic, we launched the Warrior Watch Study, comprised of our custom iOS app, which integrates survey metrics with physiological signatures acquired via the Apple Watch. The aim of the study was to understand the longitudinal perceived and physiological stress response among HCWs throughout the course of the pandemic.

Methods

Study Design

This was an observational cohort study. The primary objective of the study was to identify characteristics associated with longitudinal stress in HCWs. The secondary aim was to determine whether changes in HRV were associated with features protective against longitudinal stress development. HCWs across 7 hospitals in New York City (The Mount Sinai Hospital, Morningside Hospital, Mount Sinai West, Mount Sinai Beth Israel, Mount Sinai Queens, New York Eye and Ear Infirmary, and Mount Sinai Brooklyn) were eligible. Participants had to be current employees of one of the participating hospitals, ≥18 years of age, have an iPhone Series 6 or higher, and be willing to wear an Apple Watch Series 4 or higher. An underlying autoimmune disease or the use of medications that interfere with ANS function were exclusionary. This study was approved by the institutional review board at Mount Sinai. Informed consent was obtained from all participants.

Study Procedures

Participants downloaded the custom Warrior Watch Study app to their iPhones and completed eligibility questions prior to signing electronic consent. Through the Warrior Watch Study app, surveys were completed at enrollment and then longitudinally throughout the course of the study. At enrollment, demographic information and whether subjects had a diagnosis of anxiety or depression were collected. Psychological well-being was assessed through structured surveys evaluating perceived stress (Perceived Stress Scale-4 [PSS-4]) [11], resilience (Connor-Davidson Resilience Scale-2 [CD-RISC-2]) [12], emotional support (2-item Patient-Reported Outcomes Measurement Information System [PROMIS] questionnaire) [13], quality of life (2-item Global Health and Quality of Life [QoL]) [14], and optimism (Life Orientation Test) [15] (Table S1, Multimedia Appendix 1). Diagnosis of COVID-19 was defined as a positive SARS-CoV-2 nasal polymerase chain reaction (PCR) swab reported by the study subject.

To longitudinally evaluate subjects, daily surveys were administrated via the study app to collect COVID-19–related symptoms and severity, degree of COVID-19 exposure at work, types of patient care at work, whether participants left their home each day, if public transportation was used, the number of people that participants interacted with each day, the results of any COVID-19 nasal PCR or antibody tests, whether the subject was quarantined, if childcare needs were required, and if the subject was hospitalized. To assess trends in psychological well-being, subjects were prompted to complete the PSS-4 and

2-item General Health and QoL survey weekly. Participants were instructed to wear the Apple Watch for at least 8 hours per day throughout the study period.

Outcome Measures, Instrumentation, and the Wearable Device

HRV, a physiological marker of stress, was collected via the Apple Watch, while subjective outcome measures were assessed through standardized surveys. The Apple Watch Series 4 or 5 was worn by subjects on the wrist to capture HRV and was connected via Bluetooth to the participants' iPhone. A photoplethysmogram sensor on the Apple Watch pairs a green LED (light-emitting diode) light with a light-sensitive photodiode to generate time-series peaks [16]. The Apple Watch and Apple Health app calculates HRV using the standard deviation of NN intervals (SDNN) from the time differences between heart beats, categorized as the interbeat interval. SDNN is a time domain index reflecting sympathetic and parasympathetic nervous system activity [9]. This is recorded by the Apple Watch during approximately 60-second recording periods (ultrashort period). Ultrashort term analyses of HRV can be reliably performed to monitor mental health [17]. In accordance with current HRV guidelines, we did not compare SDNN values obtained from recordings of different duration, and we used one measuring device (Apple Watch) to assure standardization of the equipment [18].

The PSS-4 measures perceived stress. It is a 4-question survey that is scored from 0 to 16. Higher scores correlate with elevated perceived stress. The scale evaluates an individual's confidence in handling problems, whether individuals feel difficulties are piling up, whether things are felt to be going in the way of the individual, and one's sense of control [11].

The CD-RISC-2 measures resilience. It is a 2-question survey, with higher scores correlating to higher resilience. It includes questions about coping, bouncing back from adversity, whether an individual gets discouraged, and adaptation to change [12].

The 2-item emotional support PROMIS questionnaire is a 2-question survey inquiring about whether individuals have someone to listen to them when they need to talk and someone they trust to discuss their feelings with. It is scored on a scale from 2 to 10 points [13].

The 2-item Global Health and QoL questionnaire asks individuals to grade how their health and quality of life are in general, with higher scores correlating with lower health and quality of life [14].

The Life Orientation Test assesses subject optimism. It comprises 6 questions that ask individuals whether he or she expects the best in uncertain times, whether he or she expects things to go wrong, and whether more good is expected to happen compared to bad [15].

Statistical Analysis

Survey Analyses

To account for gaps created by unanswered weekly surveys and allow comparison for each patient, we created a chronological variable called a "period." To account for participants having different time windows between each weekly survey, a period was assigned to each weekly survey according to participants' starting and ending dates. When a participant's survey was completed less than 7 days from their previous survey date, the day after the previous survey date was regarded as the starting window date for the next period. When a participant's survey was done 7 days or more apart from the previous survey date, the starting window date was set to 6 days prior to the current survey date. To integrate weekly psychological metrics and daily risk/health metrics, the results of the daily surveys were summarized by the periods defined by the weekly surveys. Daily survey data were summarized for each period (eg, mean number of risk days per period, mean number of days left home per period). To examine the associations between the COVID-19 case count of New York City and perceived stress, raw case count data were obtained for the city and summarized as a mean case count per period [19].

Occupation Classification

The occupation of each participant was collected at enrollment. Due to the pandemic, the roles and responsibilities of these occupations may have changed compared to nonpandemic job descriptions. We therefore created a new occupation metric to identify which participants were seeing patients during the study. Occupation was calculated as follows:

1. Daily clinical occupation was calculated from the daily survey where participants classified the type of patient or nonpatient care responsibilities that day. Those who reported either (a) exposure to patient areas but without patients diagnosed with or being evaluated for COVID-19, or (b) exposure to areas with patients confirmed to have COVID-19 or being investigated for COVID-19 infection, were assigned as "clinical" for that day. Those who responded they were at work but not caring for patients or those who were working remotely were classified as "nonclinical" for that day.
2. If a participant had one or more clinical days in a period, that participant was assigned as clinical for that period.
3. If a participant had one or more clinical periods over the entire study, then they were deemed as either clinical nontrainee or clinical trainee. To be classified as a clinical trainee, a participant had to be either a resident or fellow. All other occupations were classified as staff.

Statistical Modeling

To model longitudinal changes in stress, we used linear mixed-effect models. Fixed effects included time-invariant covariates (gender, age, occupation, baseline resilience, optimism, and quality of life) and time-variant covariates (COVID-19 diagnosis, SARS-CoV-2/COVID-19 antibody positive test, mobility variables). A continuous first-order autoregressive correlation structure (over period) was found to be suitable for our data, significantly increasing the likelihood function (likelihood ratio test, $P<.001$) and leading to a minimal Akaike information criterion or Bayesian information criterion. Model coefficients were estimated using the restricted maximum likelihood approach via R's *nlme* packages (R Foundation for Statistical Computing). Hypotheses of interest were tested using contrasts through the capabilities of the *emmeans* package.

Univariate models tested the association of each variable with longitudinal stress and identified associated factors. Variables with $P<.10$ in the marginal ANOVA (analysis of variance) test were considered significant and included in the multivariate analysis. Although in univariate models random effects include only the intercept, in multivariate models, a random effect for the case burden in New York City was found to be significant (likelihood ratio test, $P<.001$), indicating heterogeneity in the association of this variable with stress across subjects.

Heart Rate Variability Modeling

HRV captured from the Apple Watch demonstrated a sparse nonuniform sampling and circadian pattern, making it amenable to analysis via a cosinor model. This approach models the daily HRV circadian rhythm over a period of 24 hours, which can be described using the circadian parameters: (1) midline statistic of rhythm (MESOR), a rhythm-adjusted mean; (2) amplitude, a measure of half the extent of variation within a day; and (3) acrophase, a measure of the time of overall high values recurring in each day. This allows testing of the effect that model covariates have on HRV. A cosinor model used the nonlinear function $Y(t)=M + Acos(2\pi t/\tau + \varphi + e_i(t)$, where τ is the period (τ=24 hours), M is the MESOR, *A* is the amplitude, and φ is the acrophase. This can be transformed into the linear model $x=\sin(2\pi t/\tau)$, $z=\sin(2\pi t/\tau)$, with HRV written as $Y(t)=M + \beta x_t + \gamma z_t + e_i(t)$. We identified a subject-specific daily pattern measuring departures from this pattern as a function of emotional support, resilience, and other covariates of interest. Utilizing a mixed-effect cosinor model HRV, the introduction of random effects intrinsically models the correlation due to the longitudinal sampling. Covariates, C, were introduced as fixed effects using the equation $HRV_{it} = M + a_oCi + (\beta + a_2C_i)x_{it} + (\gamma + a_3C_i)z_{it} + W_{it}\theta_i + e_i(t)$. As we have described previously to test if the cosinor curve differs between two populations of interest, we performed the bootstrapping procedure where for each iteration, we (1) fit a linear mixed-effect model using reweighted least squares, (2) estimated the marginal means for each group defined by a covariate, (3) estimated marginal means for each group using the inverse relationship, and (4) defined the bootstrapping statistics as a pairwise difference between groups [20]. Cosinor models were used to estimate HRV MESOR, amplitude, and acrophase for participants based on emotional support and resilience tertials (low, medium, high). Cosinor model covariates included time, gender, age, BMI, baseline emotional support, baseline resilience, optimism, and stress, with the participant serving as a random intercept.

Results

Overview

A total of 361 HCWs were enrolled in this ongoing observational study between April 29 and September 29, 2020, when data were censored for analysis (Table 1). Occupations were classified as (1) clinical nontrainees: participants who reported caring for a patient on any daily survey and was not a resident or fellow, (2) clinical trainee: a resident or fellow, and (3) staff: participants who did not report caring for patients on a daily survey. Participants had a mean age of 37 years, were 69.3% (n=246) female, and were followed for a mean of 60 days (IQR 21-98 days). Clinical trainees had higher baseline resilience compared to clinical nontrainees (P=.03) and staff (P=.01), higher optimism (P=.04) and emotional support (P=.01) compared to staff, and higher emotional support compared to clinical nontrainees (P=.01) (Table S2, Multimedia Appendix 1).

Table 1. Baseline demographic characteristics of the total cohort and by occupation category.

Characteristic	Total cohort (N=361)	Staff (n=65)	Clinical nontrainee (n=217)	Clinical trainee (n=40)
Age (years), mean (SD)	36.8 (10.1)	36.5 (11.0)	37.8 (10.4)	31.1 (3.6)
BMI, mean (SD)	25.7 (5.8)	—[a]	—	—
Female gender, n (%)	246 (69.3)	43 (66.2)	158 (73.8)	20 (51.3)
Race, n (%)				
Asian	90 (24.9)	14 (21.5)	49 (22.6)	14 (35.0)
Black	33 (9.1)	3 (4.6)	23 (10.6)	4 (10.0)
White	132 (36.6)	26 (40.0)	80 (36.9)	15 (37.5)
Other	47 (13.0)	7 (10.8)	31 (14.3)	6 (15.0)
Ethnicity, n (%)				
Hispanic	59 (16.3)	15 (23.1)	34 (15.7)	1 (2.5)
Baseline positive SARS-CoV-2 nasal PCR,[b] n (%)	22 (6.1)	2 (3.1)	16 (7.4)	2 (5.0)
Baseline positive SARS-CoV-2 serum antibody, n (%)	35 (9.7)	6 (9.2)	22 (10.1)	2 (5.0)
Baseline smoking status, n (%)				
Current/past smoker	48 (13.5)	10 (15.4)	31 (14.5)	0 (0.0)
Never/rarely smoked	307 (86.5)	55 (84.6)	183 (85.5)	39 (100.0)
Baseline immune-suppressing medication, n (%)	4 (1.4)	0 (0.0)	4 (1.9)	0 (0.0)
Anxiety or depression, n (%)	73 (20.6)	16 (24.6)	43 (20.1)	7 (17.9)
Baseline survey metrics, mean (SD)				
Perceived Stress Scale-4	5.3 (3.1)	5.3 (3.1)	5.5 (2.9)	5.4 (3.1)
Connor-Davidson Resilience Scale-2	5.7 (1.4)	5.4 (1.5)	5.7 (1.4)	6.2 (1.3)
Optimism	19.1 (4.2)	18.4 (4.3)	18.8 (4.2)	20.1 (3.7)
Emotional support	6.8 (1.5)	6.7 (1.7)	6.8 (1.5)	7.6 (0.9)
Quality of life	7.8 (1.5)	7.5 (1.4)	7.8 (1.4)	8.0 (1.5)
Baseline medical conditions, n (%)				
Asthma	41 (11.4)	13 (20)	19 (8.8)	5 (12.5)
Chronic lung disease	1 (0.3)	0 (0.0)	0 (0.0)	1 (2.5)
Heart disease	1 (0.3)	1 (1.5)	0 (0.0)	0 (0.0)
Cancer	2 (0.6)	0 (0.0)	2 (0.9)	0 (0.0)
Diabetes mellitus	6 (1.7)	2 (3.1)	4 (1.8)	0 (0.0)
Hypertension	20 (5.5)	5 (7.7)	11 (5.1)	0 (0.0)
Pneumonia	7 (1.9)	1 (1.5)	5 (2.3)	1 (2.5)

[a]Not applicable.

[b]PCR: polymerase chain reaction.

Factors Associated With Longitudinal Stress

The primary aim of the study was to assess the factors associated with longitudinal perceived stress. Univariate analysis evaluated the relationship between baseline demographics and prospectively collected survey metrics with longitudinal perceived stress (Table 2). Baseline factors including resilience, optimism, emotional support, quality of life, male gender, and age were significantly associated with lower longitudinal stress. Baseline anxiety or depression, BMI, weight, and asthma were significantly associated with increased longitudinal stress. Longitudinal quality of life (P<.001) was associated with reduced longitudinal stress, while the mean number of COVID-19 cases in New York City (P=.004) was positively associated with increased longitudinal stress. Occupation classification (staff vs clinical nontrainee, P=.81; staff vs clinical trainee, P=.15; clinical nontrainee vs clinical trainee, P=.17), mean number of days caring for patients (P=.88), and treatment of patients with COVID-19 (P=.73) were not associated with longitudinal stress. We observed a significant reduction in stress

during the 4-week period following diagnosis (*P*=.01) and over the follow-up period (*P*=.04). Multivariable analysis found only the New York City COVID-19 case count to be significantly associated with increased longitudinal stress (*P*=.008). The drop in stress during the 4-week period following COVID-19 diagnosis was not significant (*P*=.23); however, we noted a borderline significant increase in stress following the 4-week period after a COVID-19 diagnosis (*P*=.05). Baseline emotional support, baseline quality of life, and baseline resilience were associated with decreased longitudinal stress (*P*<.001) (Figure 1).

The COVID-19 case count of New York City and the 4-week period after a COVID-19 diagnosis via nasal PCR were further explored in the context of emotional support and resilience. Participants were stratified into emotional support tertials (low, medium, high). A significant reduction in stress during the 4-week period after COVID-19 diagnosis occurred only in participants in the highest tertial of emotional support (effect estimate –0.97, *P*=.03) but not in the medium (effect estimate –0.62, *P*=.48) and low tertials (effect estimate 0.08, *P*=.93) (Figure S1A, Multimedia Appendix 2). A significant trend between New York City's COVID-19 case count and longitudinal stress was observed only in the high tertial emotional support group (estimate 1.22, *P*=.005), not in the low (estimate –1.45, *P*=.26) or medium (estimate 0.98, *P*=.16) tertials (Figure S1B, Multimedia Appendix 2). Stratification of the cohort into tertials for resilience demonstrated a significant reduction in stress during the 4-week period after COVID-19 diagnosis via nasal PCR in the high (estimate –1.78, *P*=.006) but not medium (estimate 0.33, *P*=.64) and low tertials (estimate –0.60, *P*=.25) (Figure S1C, Multimedia Appendix 2). The impact of New York City's COVID-19 case count demonstrated a borderline significant relationship with stress in the medium (estimate 1.29, *P*=.098) and high (estimate 1.14, *P*=.09) tertials but not in the low resilience group (estimate 0.72, *P*=.21) (Figure S1D, Multimedia Appendix 2).

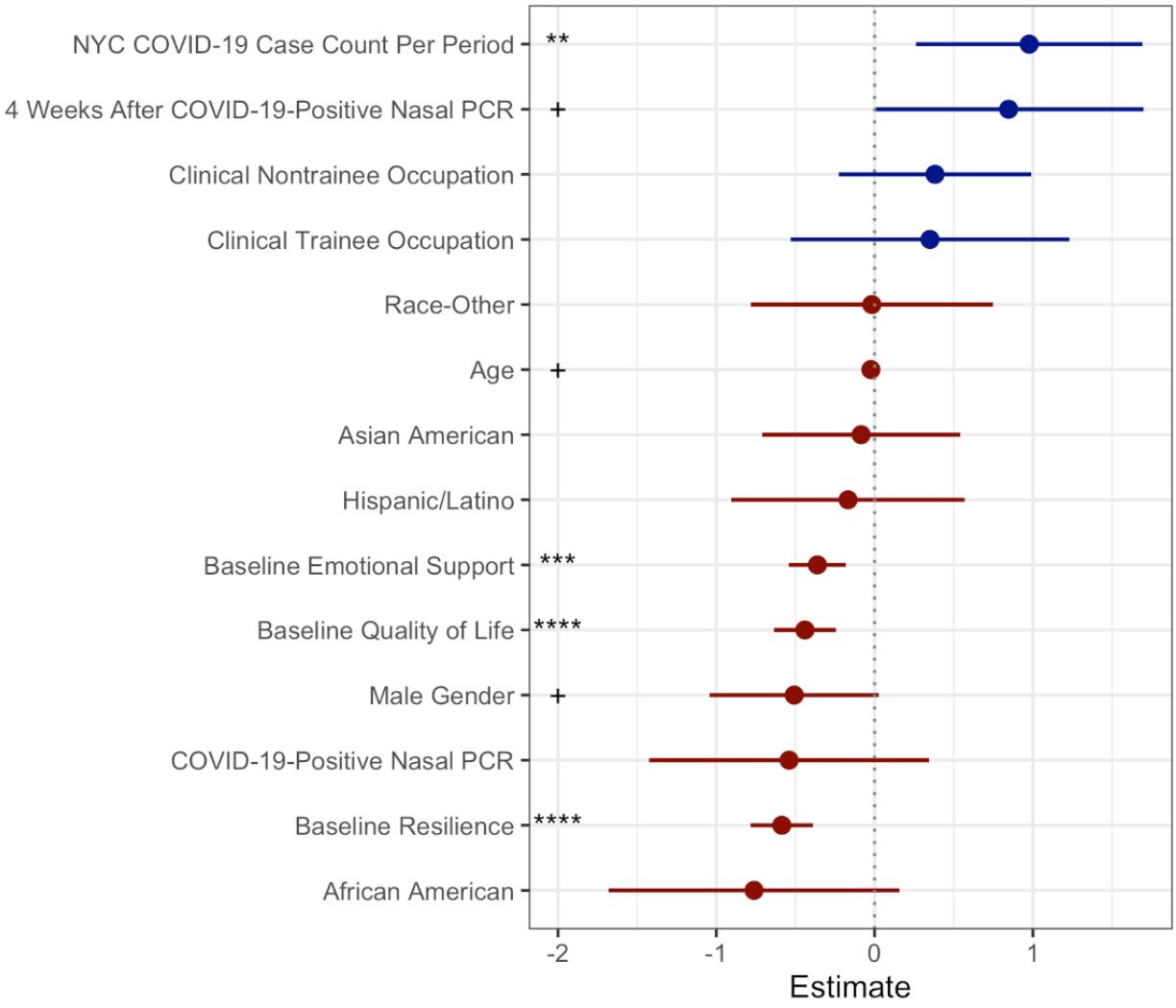
Table 2. Univariate analysis of factors associated with longitudinal perceived stress.

Factor	Effect estimate	*P* value
Baseline resilience	–0.84	<.001
Baseline optimism	–0.34	<.001
Baseline emotional support	–0.62	<.001
Baseline quality of life	–0.71	<.001
Longitudinal quality of life	–0.80	<.001
Baseline anxiety or depression	1.27	<.001
Baseline BMI	0.07	.001
Male gender	–0.94	.002
Mean New York City case count per period	0.82	.004
No positive COVID-19 nasal PCR[a] at baseline	0.91	.11
2 weeks post positive COVID-19 nasal PCR test	–0.18	.62
4 weeks post positive COVID-19 nasal PCR test	–0.82	.01
2 weeks post positive COVID-19 antibody test	–0.36	.19
4 weeks post positive COVID-19 antibody test	–0.10	.71
Any period post positive COVID-19 nasal PCR test	–0.85	.04
2 weeks post positive COVID-19 PCR or antibody test	–0.25	.28
Any period post positive COVID-19 antibody test	–0.26	.36
Weight	0.02	.04
Age	–0.03	.047
Baseline asthma	0.89	.045
Baseline heart disease	3.37	.19
Baseline hypertension	–0.34	.58
Baseline diabetes	–0.20	.85
Mean symptomatic days per period	0.32	.07
Staff vs clinical nontrainee	–0.03	.81
Staff vs clinical trainee	0.43	.15
Clinical vs clinical trainee	0.60	.17
Any period post positive COVID-19 PCR or antibody test	–0.47	.07
Height	–0.03	.08
Mean days traveled per period	–0.09	.58
Days left home per period	–0.05	.08
No immune-suppressing medication at baseline	–1.80	.17
No childcare needs at baseline	–0.39	.23
Smoking at baseline	0.53	.20
Mean days hospitalized per period	–1.66	.27
Mean days treating COVID-19–positive patients per period	0.22	.28
Number of days left home	–0.02	.30
Asian vs Black	0.02	.97
Asian vs other	–0.34	.47
Black vs other	–0.36	.55
White vs Asian	0.01	.99
White vs Black	0.03	.96

Factor	Effect estimate	*P* value
White vs other	–0.34	.45
No positive COVID-19 antibody test at baseline	0.03	.69
Days caring for patients with COVID-19	0.01	.73
Interacted with 1-3 people outside the home per day	0.01	.97
Interacted with 4-9 people outside the home per day	0.01	.96
Interacted with ≥10 people outside the home per day	–0.11	.64
Mean days quarantined per period	–0.20	.80
Total symptomatic days per period	0.01	.85
Days working weighted based on patient exposure	0.03	.69
Days hospitalized per period	0.07	.88
Days quarantined per period	–0.03	.90
Mean number of days the participant left the house per period	0.03	.90
Mean working days during this period	0.02	.88
Sum of the severity of COVID-19 symptoms per period	–0.001	.96
Mean severity of COVID-19 symptoms this period	0.08	.14

[a]PCR: polymerase chain reaction.

Figure 1. Multivariate analysis of factors associated with longitudinal stress. The scatter plot shows estimated coefficients (CIs) for variables used in the multivariate analysis. Stars indicate that the variable has a significant (*P*<.05) association with longitudinal stress while crosses indicate a borderline significant relationship (*P*<.10). Positive association is indicated in blue and negative association in red. NYC: New York City; PCR: polymerase chain reaction.



Physiological Stress Response

The secondary aim of the study was to evaluate whether features that buffer against perceived stress result in physiological differences in the stress response of HCWs. We fit a cosinor model evaluating differences in HRV (SDNN) (Table S3, Multimedia Appendix 1). A significant reduction in the amplitude of the circadian pattern of SDNN was observed between participants with high compared to medium (*P*<.001) and low (*P*=.008) emotional support (Figure 2A and 2B). There was a significant reduction in the acrophase of the circadian pattern of SDNN in participants with high emotional support compared to those with medium (*P*<.001) and low (*P*=.004) emotional support. Significant changes in the circadian pattern of SDNN were also observed when the cohort was stratified based on baseline resilience (Figure 2C and 2D). The amplitude of the circadian pattern of SDNN was significantly lower in subjects with high resilience compared to those with low (*P*<.001) and medium (*P*<.001) resilience. Similarly, the acrophase of participants with high resilience was significantly reduced compared to those with medium (*P*<.001) and low (*P*=.048) resilience (Table 3).

Figure 2. Exploring the relationship between heart rate variability (HRV), emotional support, and resilience. Plots A and C show mean (95% CIs) HRV midline statistic of rhythm (MESOR), amplitude, and acrophase for participants with low, medium, and high emotional support (A) or resilience (C). Stars indicate significant differences between groups. Plots B and D show average daily circadian HRV rhythm for participants with low, medium, and high emotional support (B) or resilience (D). +*P*<.10, **P*<.05, ***P*<.01, ****P*<.001. SDNN: standard deviation of NN intervals.

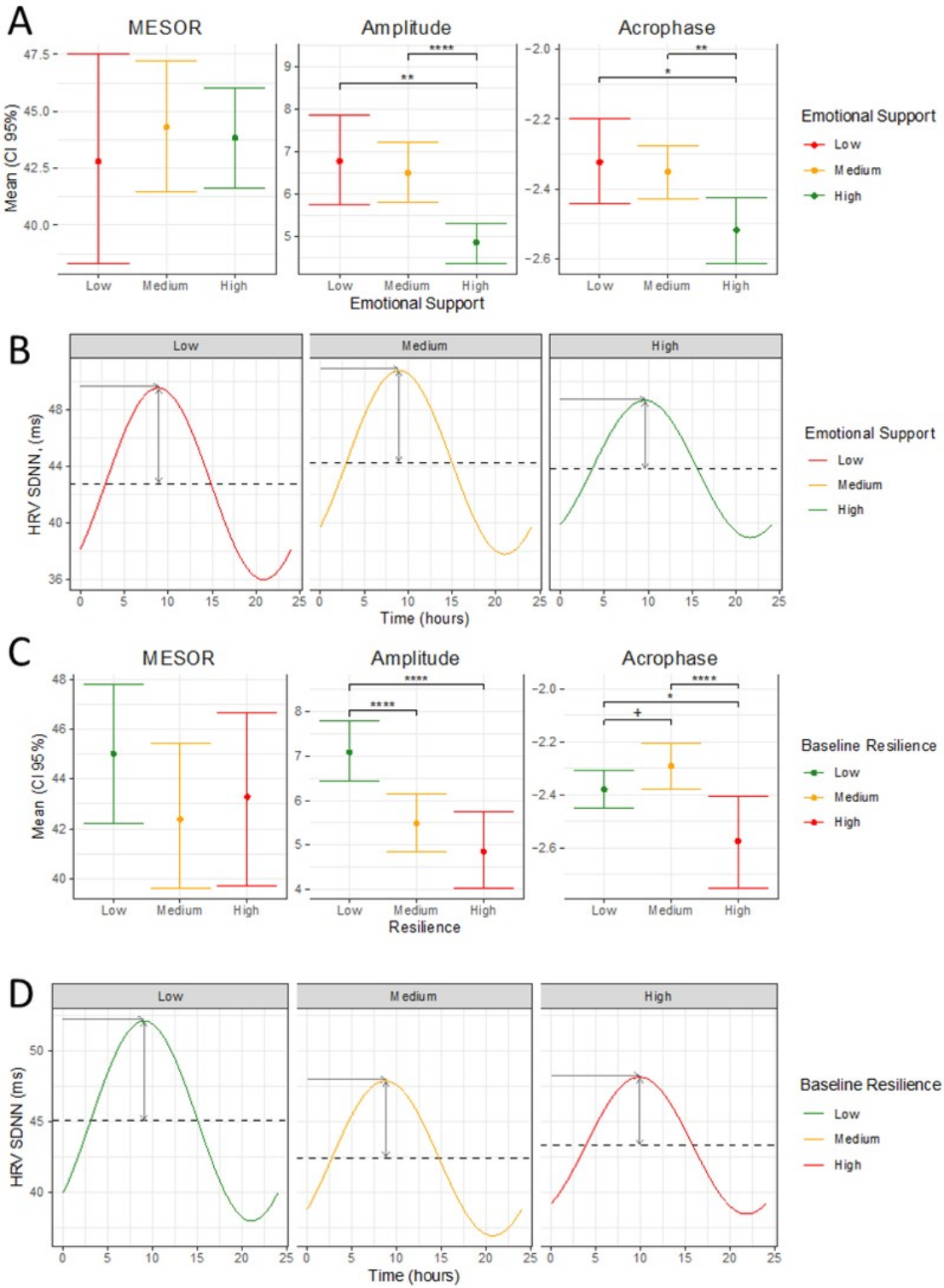


Table 3. Comparison of mean heart rate variability parameters stratified based on emotional support and resilience tertials.

Parameter and tertial comparisons	Emotional support, *P* value	Resilience, *P* value
MESOR[a]		
Low vs medium	.60	.10
Low vs high	.67	.46
Medium vs high	.78	.71
Amplitude		
Low vs medium	.68	<.001
Low vs high	.01	<.001
Medium vs high	<.001	.24
Acrophase		
Low vs medium	.70	.09
Low vs high	.004	.048
Medium vs high	<.001	<.001

[a]MESOR: midline statistic of rhythm.

Discussion

Principal Results and Comparison With Prior Work

In summary, to the best of our knowledge, we conducted the first study identifying HCW characteristics that correlate with longitudinal stress during the COVID-19 pandemic and identified employees at risk of psychological sequela. We found worsening longitudinal stress to be associated with the number of COVID-19 cases in the community, highlighting the effect of this environmental stressor. Baseline emotional support, resilience, and quality of life, rather than occupation class, defined which HCWs were prone to perceived longitudinal stress and characterized a unique ANS stress profile.

In line with our findings, prior work shows that emotional support and resilience buffer against stress [21,22]. Resilience, defined as a reduced vulnerability to environmental stressors and the ability to overcome difficulty, is crucial to establishing social relationships and is tied to social support, which also acts as an environmental protective factor against adversity [23-25]. In addition to demonstrating their stress protective effect in the multivariate analysis, when we further evaluated the COVID-19 case count of New York City, a factor associated with longitudinal stress over time, we again found that those with lower emotional support or resilience were vulnerable to a dynamic stress response uncoupled from the environmental COVID-19 stressor. Similarly, the transient reduction in stress that occurs after a COVID-19 diagnosis only occurs in those with high emotional support and resilience. Importantly, these latter findings highlight how high resilience or high emotional support can minimize the impact of factors associated with longitudinal stress in HCWs.

A strength of our study is the objective assessment of this observation through longitudinal HRV measurements. HRV is a marker of physiological stress on the ANS. Repeated stressors that are felt to be unmanageable and overwhelming may lead to an impaired stress response characterized by altered autonomic and hypothalamic-pituitary-adrenal axis function [26,27]. Buffers to stress, such as resilience, have also been shown to impact ANS function and are associated with an activation of the sympathetic nervous system that is sufficient to respond to a stressor but not to a degree that results in the development of negative psychological effects [19]. Resilience has been shown to moderate the relationship between perceived stress and the hypothalamic pituitary adrenal axis, which acts as a primary mediator of the physiological stress response [28]. Studies evaluating resilience have demonstrated an associated characteristic ANS profile, with higher vagal indices and with a baseline sympathovagal balance shifted toward parasympathetic predominance [27,29]. Resilience has been associated with the modulation of sympathetic nervous system activation, allowing for an adequate response to a stressor but with moderation of the degree of activation and a rapid return to baseline. This allows for a stable emotional response and reduced chronic anxiety [27,30,31]. Our findings extend these observations into HCWs during the COVID-19 pandemic. We found that buffers of stress, such as high resilience or high emotional support, were characterized by a physiologically distinct ANS profile confirming their impact on how individuals respond to stressors. These findings substantiate the effect these features have on longitudinal stress in HCWs in multiple dimensions, reaffirming their importance not only in the perception of stress but also in how HCWs are physiologically affected by the COVID-19 pandemic stressor.

Our findings have implications for how HCWs can be assessed for the pandemic's detrimental psychological effects, through screening for an individual's degree of resilience and social support. Evaluation of these characteristics will assist health care institutions in allocating often limited psychological support services to at-risk individuals. Importantly, one of these features, resilience, is modifiable through targeted interventions. This provides an opportunity to build resilience in HCWs who are found to have low resilience. While further studies are needed on the impact of such interventions in HCWs, they may mitigate the physiological impact of longitudinal stress. Several resilience

building interventions have demonstrated to be effective in HCWs [32,33]. However, our findings linking HRV alterations with degree of resilience, makes HRV focused resilience building exercises an attractive option [34].

Strengths and Limitations

Strengths of the study are its multicenter, longitudinal study design. Furthermore, the number and type of longitudinal variables we captured allowed for a robust multivariate analysis. Lastly, the incorporation of ANS parameters provided an objective assessment of the stress response. However, there are several limitations to our study. The Apple Watch provides HRV data in one-time dimension (SDNN), limiting evaluation of other metrics with outcomes of interests. The Apple Watch also provides HRV sampling sporadically throughout the day. While our modeling accounts for this, a denser sampling would allow expanded analyses. Additionally, HRV can be impacted by many environmental factors that cannot be fully accounted for in such studies. While we tried to control for relevant covariates, there is the potential for unmeasured factors to impact our results.

Conclusions

We identified features associated with longitudinal perceived stress in HCWs during the COVID-19 pandemic. The COVID-19 case count of New York City was significantly associated with increased longitudinal perceive stress, and baseline emotional support, quality of life, and resilience were associated with decreased longitudinal perceived stress. Furthermore, high resilience and high social support impacted the physiological stress response and were associated with a unique autonomic nervous system profile. This demonstrates the importance of resilience and social support on both perception of stress and its physiological impact.

Acknowledgments

Support for this study was provided by the Ehrenkranz Lab for Human Resilience, the BioMedical Engineering and Imaging Institute, the Hasso Plattner Institute for Digital Health at Mount Sinai, the Mount Sinai Clinical Intelligence Center, and the Dr Henry D Janowitz Division of Gastroenterology.

Authors' Contributions

RPH, MD, GNN, and ZAF developed the study concept. RPH assisted with the drafting of the manuscript. RPH, MD, LT, KHC, MZ, EG, SK, DH, AB, RP, CC, RF, BES, DC, EPB, JWM, LK, MS-F, GNN, and ZAF critically revised the manuscript for important intellectual content. RPH, MD, LT, KHC, MZ, EG, SK, DH, AB, RP, CC, RF, BES, DC, EPB, JWM, LK, MS-F, GNN, and ZAF provided final approval of the version of the manuscript to be published and agree to be accountable for all aspects of the work. All authors approve the authorship list. All authors had full access to all the data in the manuscript and had final responsibility for the decision to submit for publication. RPH, ZAF, MS-F, MD, and LT have verified the underlying data.

Conflicts of Interest

RPH declares consulting fees from HealthMode, Inc, Janssen Pharmaceuticals, and Takeda Pharmaceuticals; research support from Intralytix Inc; a Crohn’s and Colitis Foundation Career Development Award (grant number 607934); and research support from a National Institutes of Health K23 Career Development Award (grant number 1K23DK129835-01). DC is a coinventor on patents filed by the Icahn School of Medicine at Mount Sinai (ISMMS), relating to the treatment for treatment-resistant depression, suicidal ideation, and other disorders. ISMMS has entered into a licensing agreement with Janssen Pharmaceuticals, Inc, and it has and will receive payments from Janssen under the license agreement related to these patents for the treatment of treatment-resistant depression and suicidal ideation. Consistent with the ISMMS Faculty Handbook (the medical school policy), DC is entitled to a portion of the payments received by the ISMMS. Since SPRAVATO has received regulatory approval for treatment-resistant depression, the ISMMS and thus, through the ISMMS, DC, will be entitled to additional payments, beyond those already received, under the license agreement. DC is a named coinventor on several patents filed by the ISMMS for a cognitive training intervention to treat depression and related psychiatric disorders. The ISMMS has entered into a licensing agreement with Click Therapeutics, Inc, and has and will receive payments related to the use of this cognitive training intervention for the treatment of psychiatric disorders. In accordance with the ISMMS Faculty Handbook, DC has received a portion of these payments and is entitled to a portion of any additional payments that the medical school might receive from this license with Click Therapeutics. DC is a named coinventor on a patent application filed by the ISMMS for the use of intranasally administered Neuropeptide Y for the treatment of mood and anxiety disorders. This intellectual property has not been licensed. DC is a named coinventor on a patent application in the United States and several issued patents outside the United States filed by the ISMMS related to the use of ketamine for the treatment of posttraumatic stress disorder. This intellectual property has not been licensed. EPB reports consultancy agreements with Deloitte and Roland Berger; ownership interest in Digital Medicine E Böttinger GmbH, EBCW GmbH, and Ontomics, Inc; receives honoraria from Bayer, Bosch Health Campus, Sanofi, and Siemens; and serves as a scientific advisor/member of Bosch Health Campus and Seer Biosciences Inc. LK declares research funding from Abbvie and Pfizer, consulting for Abbvie and Pfizer, and equity ownership/stock options in MetaMe Health and Trellus Health. GNN reports employment with, consultancy agreements with, and ownership interest in Pensieve Health and Renalytix AI; receives consulting fees from AstraZeneca, BioVie, GLG Consulting, and Reata; and serves as a scientific advisor/member of Pensieve Health and Renalytix AI. ZAF discloses consulting fees from Alexion, GlaxoSmithKline, and Trained Therapeutix Discovery; and research

funding from Daiichi Sankyo, Amgen, Bristol Myers Squibb, and Siemens Healthineers. ZAF receives financial compensation as a board member and advisor to Trained Therapeutix Discovery and owns equity in Trained Therapeutix Discovery as a cofounder. In the past 5 years, JWM has provided consultation services and/or served on advisory boards for Allergan, Boehreinger Ingelheim, Clexio Biosciences, Fortress Biotech, FSV7, Global Medical Education, Otsuka, Sage Therapeutics, and Engrail Therapeutics. JWM is named on a patent pending for Neuropeptide Y as a treatment for mood and anxiety disorders and on a patent pending for the use of KCNQ channel openers to treat depression and related conditions. The remaining authors have no competing interests to declare.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 26 KB - jmir_v23i9e31295_app1.docx]

Multimedia Appendix 2
Supplementary figures. Plots A and C show changes in longitudinal stress following a positive COVID-19 nasal test in participants with low, medium, and high emotional support (A) or resilience (C); stars indicate that change in longitudinal stress was significantly different from zero. Line plots B and D show the relationship between New York City's COVID-19 case count and mean longitudinal stress (CIs) for participants with low, medium, and high emotional support (B) or resilience (D); stars indicate a significant trend between case count and longitudinal stress. +<italic>P</italic><.10, *<italic>P</italic><.05, **<italic>P</italic><.01, ***<italic>P</italic><.001.
[PNG File , 350 KB - jmir_v23i9e31295_app2.png]

References

1. Preti E, Di Mattei V, Perego G, Ferrari F, Mazzetti M, Taranto P, et al. The Psychological Impact of Epidemic and Pandemic Outbreaks on Healthcare Workers: Rapid Review of the Evidence. Curr Psychiatry Rep 2020 Jul 10;22(8):43 [FREE Full text] [doi: 10.1007/s11920-020-01166-z] [Medline: 32651717]
2. Bai Y, Lin CC, Lin CY, Chen JY, Chue CM, Chou P. Survey of stress reactions among health care workers involved with the SARS outbreak. Psychiatr Serv 2004 Sep;55(9):1055-1057. [doi: 10.1176/appi.ps.55.9.1055] [Medline: 15345768]
3. Charney AW, Katz C, Southwick SM, Charney DS. A Call to Protect the Health Care Workers Fighting COVID-19 in the United States. Am J Psychiatry 2020 Oct 01;177(10):900-901. [doi: 10.1176/appi.ajp.2020.20040535] [Medline: 32731814]
4. Lai J, Ma S, Wang Y, Cai Z, Hu J, Wei N, et al. Factors Associated With Mental Health Outcomes Among Health Care Workers Exposed to Coronavirus Disease 2019. JAMA Netw Open 2020 Mar 02;3(3):e203976 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3976] [Medline: 32202646]
5. Liang Y, Wu K, Zhou Y, Huang X, Zhou Y, Liu Z. Mental Health in Frontline Medical Workers during the 2019 Novel Coronavirus Disease Epidemic in China: A Comparison with the General Population. Int J Environ Res Public Health 2020 Sep 09;17(18):6550 [FREE Full text] [doi: 10.3390/ijerph17186550] [Medline: 32916836]
6. Rossi R, Socci V, Pacitti F, Di Lorenzo G, Di Marco A, Siracusano A, et al. Mental Health Outcomes Among Frontline and Second-Line Health Care Workers During the Coronavirus Disease 2019 (COVID-19) Pandemic in Italy. JAMA Netw Open 2020 May 01;3(5):e2010185 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.10185] [Medline: 32463467]
7. Zerbini G, Ebigbo A, Reicherts P, Kunz M, Messman H. Psychosocial burden of healthcare professionals in times of COVID-19 - a survey conducted at the University Hospital Augsburg. Ger Med Sci 2020;18:Doc05 [FREE Full text] [doi: 10.3205/000281] [Medline: 32595421]
8. Morgantini LA, Naha U, Wang H, Francavilla S, Acar Ö, Flores JM, et al. Factors contributing to healthcare professional burnout during the COVID-19 pandemic: A rapid turnaround global survey. PLoS One 2020 Sep 3;15(9):e0238217 [FREE Full text] [doi: 10.1371/journal.pone.0238217] [Medline: 32881887]
9. Shaffer F, Ginsberg JP. An Overview of Heart Rate Variability Metrics and Norms. Front Public Health 2017 Sep;5:258 [FREE Full text] [doi: 10.3389/fpubh.2017.00258] [Medline: 29034226]
10. McEwen B. Stress, adaptation, and disease. Allostasis and allostatic load. Ann N Y Acad Sci 1998 May 01;840:33-44. [doi: 10.1111/j.1749-6632.1998.tb09546.x] [Medline: 9629234]
11. Cohen S, Kamarck T, Mermelstein R. A Global Measure of Perceived Stress. Journal of Health and Social Behavior 1983 Dec;24(4):385. [doi: 10.2307/2136404]
12. Connor KM, Davidson JR. Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). Depress Anxiety 2003 Sep 02;18(2):76-82. [doi: 10.1002/da.10113] [Medline: 12964174]
13. Hahn EA, Devellis RF, Bode RK, Garcia SF, Castel LD, Eisen SV, et al. Measuring social health in the patient-reported outcomes measurement information system (PROMIS): item bank development and testing. Qual Life Res 2010 Sep;19(7):1035-1044 [FREE Full text] [doi: 10.1007/s11136-010-9654-0] [Medline: 20419503]
14. Hays RD, Schalet BD, Spritzer KL, Cella D. Two-item PROMIS® global physical and mental health scales. J Patient Rep Outcomes 2017 Sep 12;1(1):2 [FREE Full text] [doi: 10.1186/s41687-017-0003-8] [Medline: 29757325]

15. Scheier MF, Carver CS, Bridges MW. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. Journal of Personality and Social Psychology 1994;67(6):1063-1078. [doi: 10.1037/0022-3514.67.6.1063]
16. Monitor your heart rate with Apple Watch. Apple Inc. URL: https://support.apple.com/en-us/HT204666 [accessed 2020-10-22]
17. Salahuddin L, Cho J, Jeong MG, Kim D. Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. 2007 Presented at: 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Aug 22-26; Lyon, France p. 4656-4659. [doi: 10.1109/IEMBS.2007.4353378]
18. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Eur Heart J 1996 Mar;17(3):354-381. [Medline: 8737210]
19. New York City Case Count. GitHub. URL: https://github.com/nychealth/coronavirus-data/tree/master [accessed 2020-12-01]
20. Hirten R, Danieletto M, Tomalin L, Choi KH, Zweig M, Golden E, et al. Use of Physiological Data From a Wearable Device to Identify SARS-CoV-2 Infection and Symptoms and Predict COVID-19 Diagnosis: Observational Study. J Med Internet Res 2021 Feb 22;23(2):e26107 [FREE Full text] [doi: 10.2196/26107] [Medline: 33529156]
21. Rutter M. Resilience as a dynamic concept. Dev Psychopathol 2012 Apr 17;24(2):335-344. [doi: 10.1017/s0954579412000028]
22. Rutter M. Annual Research Review: Resilience--clinical implications. J Child Psychol Psychiatry 2013 Apr;54(4):474-487. [doi: 10.1111/j.1469-7610.2012.02615.x] [Medline: 23017036]
23. Rutter M. Implications of resilience concepts for scientific understanding. Ann N Y Acad Sci 2006 Dec 01;1094(1):1-12. [doi: 10.1196/annals.1376.002] [Medline: 17347337]
24. Herrman H, Stewart DE, Diaz-Granados N, Berger EL, Jackson B, Yuen T. What is resilience? Can J Psychiatry 2011 May 01;56(5):258-265. [doi: 10.1177/070674371105600504] [Medline: 21586191]
25. Holz NE, Tost H, Meyer-Lindenberg A. Resilience and the brain: a key role for regulatory circuits linked to social stress and support. Mol Psychiatry 2020 Feb 18;25(2):379-396. [doi: 10.1038/s41380-019-0551-9] [Medline: 31628419]
26. Kim H, Cheon E, Bai D, Lee YH, Koo B. Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. Psychiatry Investig 2018 Mar;15(3):235-245 [FREE Full text] [doi: 10.30773/pi.2017.08.17] [Medline: 29486547]
27. Averill LA, Averill CL, Kelmendi B, Abdallah CG, Southwick SM. Stress Response Modulation Underlying the Psychobiology of Resilience. Curr Psychiatry Rep 2018 Mar 28;20(4):27. [doi: 10.1007/s11920-018-0887-x] [Medline: 29594808]
28. Lehrer HM, Steinhardt MA, Dubois SK, Laudenslager ML. Perceived stress, psychological resilience, hair cortisol concentration, and metabolic syndrome severity: A moderated mediation model. Psychoneuroendocrinology 2020 Mar;113:104510 [FREE Full text] [doi: 10.1016/j.psyneuen.2019.104510] [Medline: 31911349]
29. Prell R, Opatz O, Merati G, Gesche B, Gunga H, Maggioni MA. Heart Rate Variability, Risk-Taking Behavior and Resilience in Firefighters During a Simulated Extinguish-Fire Task. Front Physiol 2020 Jul 10;11:482 [FREE Full text] [doi: 10.3389/fphys.2020.00482] [Medline: 32754042]
30. Dienstbier RA. Arousal and physiological toughness: implications for mental and physical health. Psychol Rev 1989 Jan;96(1):84-100. [doi: 10.1037/0033-295x.96.1.84] [Medline: 2538855]
31. Southwick SM, Bremner J, Rasmusson A, Morgan CA, Arnsten A, Charney DS. Role of norepinephrine in the pathophysiology and treatment of posttraumatic stress disorder. Biological Psychiatry 1999 Nov;46(9):1192-1204. [doi: 10.1016/s0006-3223(99)00219-x]
32. Fox S, Lydon S, Byrne D, Madden C, Connolly F, O'Connor P. A systematic review of interventions to foster physician resilience. Postgrad Med J 2018 Mar 10;94(1109):162-170. [doi: 10.1136/postgradmedj-2017-135212] [Medline: 29018095]
33. Mealer M, Conrad D, Evans J, Jooste K, Solyntjes J, Rothbaum B, et al. Feasibility and acceptability of a resilience training program for intensive care unit nurses. Am J Crit Care 2014 Nov 31;23(6):e97-105. [doi: 10.4037/ajcc2014747] [Medline: 25362680]
34. Lemaire J, Wallace J, Lewin A, de Grood J, Schaefer J. The effect of a biofeedback-based stress management tool on physician stress: a randomized controlled clinical trial. Open Med 2011;5(4):e154-e163 [FREE Full text] [Medline: 22567069]

Abbreviations

ANOVA: analysis of variance
ANS: autonomic nervous system
CD-RISC-2: Connor-Davidson Resilience Scale-2
HCW: health care worker
HRV: heart rate variability
LED: light-emitting diode
MESOR: midline statistic of rhythm
PCR: polymerase chain reaction

PROMIS: Patient-Reported Outcomes Measurement Information System
PSS-4: Perceived Stress Scale-4
QoL: quality of life
SDNN: standard deviation of NN intervals

Edited by G Eysenbach; submitted 16.06.21; peer-reviewed by K Aguirre, IN Gomez; comments to author 08.07.21; revised version received 19.07.21; accepted 07.08.21; published 13.09.21.

Please cite as:
Hirten RP, Danieletto M, Tomalin L, Choi KH, Zweig M, Golden E, Kaur S, Helmus D, Biello A, Pyzik R, Calcagno C, Freeman R, Sands BE, Charney D, Bottinger EP, Murrough JW, Keefer L, Suarez-Farinas M, Nadkarni GN, Fayad ZA
Factors Associated With Longitudinal Psychological and Physiological Stress in Health Care Workers During the COVID-19 Pandemic: Observational Study Using Apple Watch Data
J Med Internet Res 2021;23(9):e31295
URL: https://www.jmir.org/2021/9/e31295
doi:10.2196/31295
PMID:34379602

Review

Overview of Technologies Implemented During the First Wave of the COVID-19 Pandemic: Scoping Review

Alaa Abd-Alrazaq[1], PhD; Asmaa Hassan[1], MSc; Israa Abuelezz[1], MSc; Arfan Ahmed[1], PhD; Mahmood Saleh Alzubaidi[1], MSc; Uzair Shah[1], MSc; Dari Alhuwail[2,3], PhD; Anna Giannicchi[4], MA; Mowafa Househ[1], PhD

[1]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

[2]Information Science Department, Kuwait University, Kuwait, Kuwait

[3]Health Informatics Unit, Dasman Diabetes Institute, Kuwait, Kuwait

[4]School of Professional Studies, Berkeley College, New York, NY, United States

Corresponding Author:
Mowafa Househ, PhD
Division of Information and Computing Technology, College of Science and Engineering
Hamad Bin Khalifa University
Qatar Foundation
P.O. Box 5825, Doha Al Luqta St, Ar-Rayyan
Doha
Qatar
Phone: 974 55708549
Email: mhouseh@hbku.edu.qa

Abstract

Background: Technologies have been extensively implemented to provide health care services for all types of clinical conditions during the COVID-19 pandemic. While several reviews have been conducted regarding technologies used during the COVID-19 pandemic, they were limited by focusing either on a specific technology (or features) or proposed rather than implemented technologies.

Objective: This review aims to provide an overview of technologies, as reported in the literature, implemented during the first wave of the COVID-19 pandemic.

Methods: We conducted a scoping review using PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) Extension for Scoping Reviews. Studies were retrieved by searching 8 electronic databases, checking the reference lists of included studies and relevant reviews (backward reference list checking), and checking studies that cited included studies (forward reference list checking). The search terms were chosen based on the target intervention (ie, technologies) and the target disease (ie, COVID-19). We included English publications that focused on technologies or digital tools implemented during the COVID-19 pandemic to provide health-related services regardless of target health condition, user, or setting. Two reviewers independently assessed the eligibility of studies and extracted data from eligible papers. We used a narrative approach to synthesize extracted data.

Results: Of 7374 retrieved papers, 126 were deemed eligible. Telemedicine was the most common type of technology (107/126, 84.9%) implemented in the first wave of the COVID-19 pandemic, and the most common mode of telemedicine was synchronous (100/108, 92.6%). The most common purpose of the technologies was providing consultation (75/126, 59.5%), followed by following up with patients (45/126, 35.7%), and monitoring their health status (22/126, 17.4%). Zoom (22/126, 17.5%) and WhatsApp (12/126, 9.5%) were the most commonly used videoconferencing and social media platforms, respectively. Both health care professionals and health consumers were the most common target users (103/126, 81.7%). The health condition most frequently targeted was COVID-19 (38/126, 30.2%), followed by any physical health conditions (21/126, 16.7%), and mental health conditions (13/126, 10.3%). Technologies were web-based in 84.1% of the studies (106/126). Technologies could be used through 11 modes, and the most common were mobile apps (86/126, 68.3%), desktop apps (73/126, 57.9%), telephone calls (49/126, 38.9%), and websites (45/126, 35.7%).

Conclusions: Technologies played a crucial role in mitigating the challenges faced during the COVID-19 pandemic. We did not find papers describing the implementation of other technologies (eg, contact-tracing apps, drones, blockchain) during the first

wave. Furthermore, technologies in this review were used for other purposes (eg, drugs and vaccines discovery, social distancing, and immunity passport). Future research on studies on these technologies and purposes is recommended, and further reviews are required to investigate technologies implemented in subsequent waves of the pandemic.

(J Med Internet Res 2021;23(9):e29136) doi:10.2196/29136

KEYWORDS

technologies; digital tools; COVID-19; novel coronavirus; scoping review; digital health; telemedicine

Introduction

Background

Since late 2019, COVID-19 has spread rapidly across the globe, leading to 175 million infections and nearly 3.8 million deaths worldwide [1]. Health care systems quickly became overwhelmed with the influx of patients, which led to overcrowded or inaccessible hospitals, delayed interventions for time-sensitive conditions and chronic illnesses, and overworked treatment teams that lacked proper personal protective equipment [2].

While developing strategic implementations to eradicate the disease, treatment providers, government entities, and individuals relied on technologies to mitigate some of the disastrous effects of COVID-19. Specifically, technologies have been used by treatment providers working with patients via telehealth; for contact tracing, screening, diagnosing COVID-19, establishing vaccination or immunity passports, and delivering foods, medications, and equipment; for predicting the trend of the pandemic, drug and vaccine discovery, social distancing; and for monitoring quarantined individuals [3]. Technological support played an integral role in guiding these forms of treatment and newly implemented policies likely prevented an even greater global crisis.

Despite the technological advancements introduced as a result of COVID-19, which some researchers have referred to as a "digital revolution," more technologies are still required to alleviate the impact of COVID-19. For instance, hospital-at-home care has yet to be widely introduced, despite many patients who may benefit from this treatment [4]. While less-urgent telehealth visits have played a large role in keeping health care facilities at an appropriate patient capacity, the hospital-at-home model would allow for patients to receive treatment comfortably at home. Patients treated at home would likely not contribute to overcrowding or take up valuable personal protective equipment resources that treatment providers may need for more serious uses. Other researchers have commented that the use of artificial intelligence—which has grown instrumentally throughout the COVID-19 pandemic—has been sparing in assisting surgeons [5]. There is hope that, in the future, artificial intelligence can facilitate patient recovery by something as advanced as implementation in surgery or as simple as patient ventilatory parameter adjustment.

Overall, technologies have played an integral part in keeping communities safe and informed, giving treatment providers innovative ways to effectively perform their duties, and providing government administrations real-time information on the virus' effects.

Research Problem and Aim

Since the first COVID-19 infection was recorded in December 2019 [6], researchers have written extensively on how technology has played a crucial role in the fight against COVID-19—several reviews have been conducted about technologies used during the pandemic. For instance, Davalbhakta et al [7] reviewed mobile apps and assessed their quality, with the primary purpose of mitigating issues related to COVID-19. Gao et al [8] conducted a rapid review about the use of telemedicine for COVID-19 for any purpose. Another review [9] explored the COVID-19–related uses of the Internet of Things, unmanned aerial vehicles, blockchain, artificial intelligence, and 5G.

While these reviews [7-9] play a substantial role in informing how technologies have been used to control the spread of COVID-19, they have been limited by either their topic or their nature. Specifically, most reviews [3,7,8,10] only focused on one or a few technologies at a time, which is an approach that does not allow for gauging the all-encompassing scope of technology use during this pandemic. Moreover, some reviews were literature reviews; therefore, they were not systematic or comprehensive (ie, as systematic reviews and scoping reviews are) [11]. Lastly, some reviews included proposed technologies, rather than technologies that had been implemented [3,10]. In contrast, we aimed to provide an overview of technologies, as reported in the literature, that were implemented during the first wave of the COVID-19 pandemic.

Methods

Overview

To achieve the above-mentioned objective, while ensuring both replicable and transparent methods, we conducted a scoping review using PRISMA ((Preferred Reporting Items for Systematic Reviews and Meta-analyses) Extension for Scoping [12].

Search Strategy

Search Sources

We searched the following electronic databases on August 14, 2020: MEDLINE (via Ovid), EMBASE (via Ovid), PsycInfo (via Ovid), Scopus, Web of Science, IEEE Xplore, ACM Library, and Google Scholar. To account for Google Scholar's vast number of results, only the first 100 citations (sorted by relevance) were screened. To identify additional studies, eligible studies' reference lists were checked, and papers that cited eligible studies were checked.

Search Terms

To develop the search queries (Multimedia Appendix 1), 2 experts in digital health were consulted and other systematic reviews of relevance to the review were checked. Terms were chosen based on the target intervention (ie, technologies) and the target disease (ie, COVID-19).

Study Eligibility Criteria

We included papers that focused on technologies or digital tools implemented during COVID-19 to provide health-related services (eg, consultations, diagnosis, and follow-up) regardless of the target health condition, user, or setting. Studies that only proposed technologies (ie, that did not implement technologies) or focused on technologies for providing non–health-related services (eg, shopping, food delivery) were excluded. Peer-reviewed journal articles were considered, along with conference proceedings, theses, dissertations, preprints, and reports. We restricted our search to studies published between January 1, 2020 to August 14, 2020. Conference abstracts, proposals, editorials, commentaries, and non–English-language papers were excluded. No restrictions related to country of publication, outcome measure, and study design were applied.

Study Selection

We followed 3 steps to identify the relevant papers: (1) duplicates were identified and removed using EndNote (Clarivate Analytics). (2) Reviewers (AH and IA) independently checked the titles and abstracts of all identified studies. (3) The 2 reviewers independently read the full texts of papers included from the second step. Any disagreements between the 2 reviewers were resolved by consulting a third reviewer (AA).

Data Extraction and Synthesis

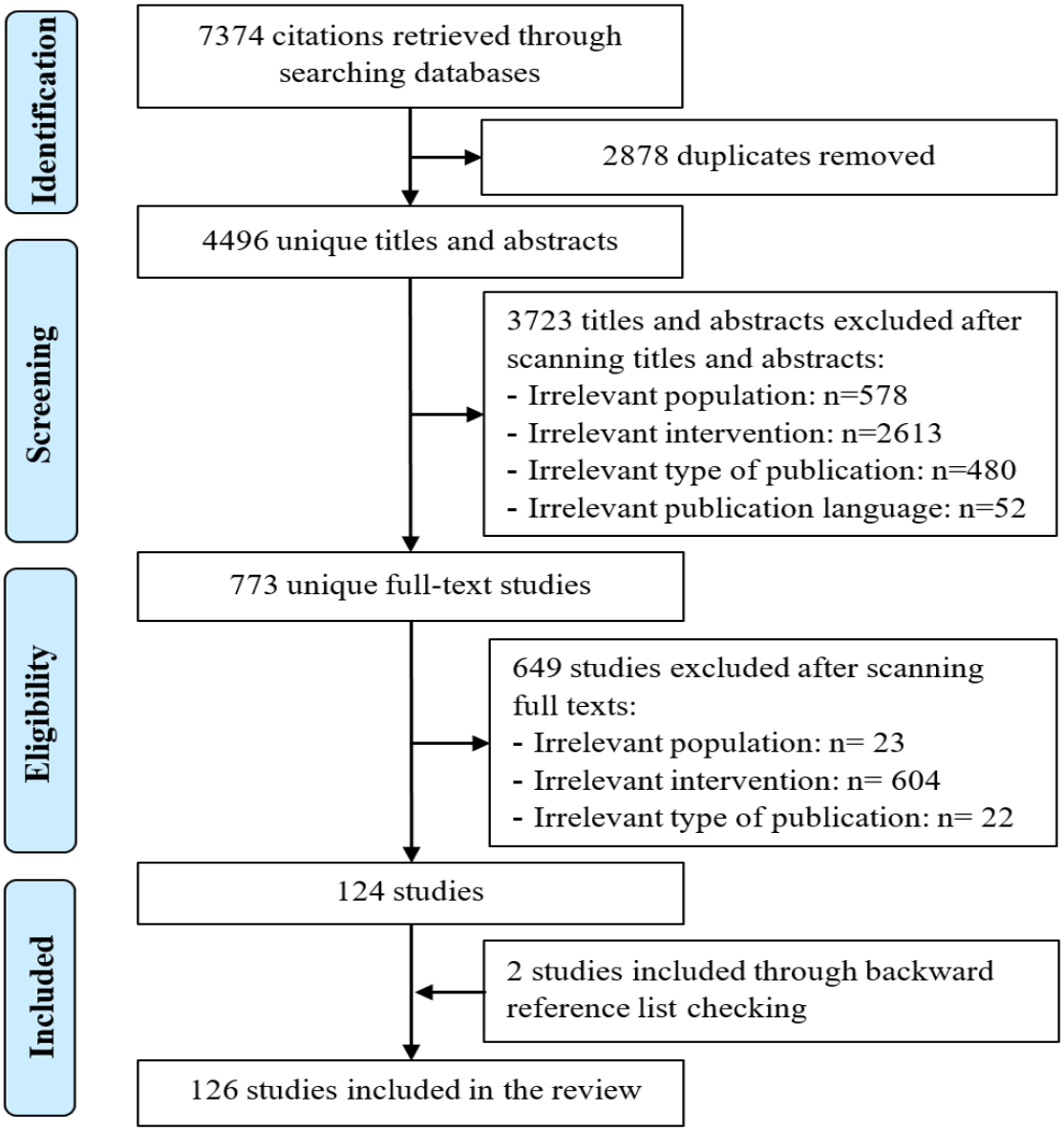
We designed a form to extract the data precisely and systematically (Multimedia Appendix 2). The form was pilot-tested using 10 eligible papers. Given the large number of eligible papers, the list was divided equally into 2 lists. While AA and MA independently extracted the data from the first list, AAA and US independently extracted data from the second list. Any disagreements between the reviewers were resolved by discussion until consensus was reached. A narrative approach was then used to synthesize the extracted data; tables were also used to describe technology types, aims, target users, target health conditions, settings, and modes. Excel (Microsoft Inc) was utilized to manage data extraction and synthesis.

Results

Search Results

Of 7374 citations retrieved through the search (Figure 1), 2878 duplicates were identified and removed. The remaining 4496 citation titles and abstracts were screened, of which 3723 citations were excluded. In total, 126 publications were included in this review [13-138].

Figure 1. Flowchart of the study selection process.



Characteristics of the Included Studies

The first paper [98] was published approximately 2 months after of the emergence of COVID-19. While only 1 paper [98] was published in the first 3 months, there was a sharp increase in the number published in the 5 subsequent months (Table 1); the largest number of studies (n=39) was published in June. The majority of studies (123/126, 97.6%) were papers in peer-reviewed journals, while the remaining studies were reports (n=2) and preprints (n=1). Studies were conducted in 23 different countries—the largest number of studies was carried out in the United States (51/126, 40.5%), followed by China (15/126, 11.9%), Italy (11/126, 8.7%), and India (10/126, 7.9%). Multimedia Appendix 3 shows the characteristics of each study.

Table 1. Characteristics of studies.

Characteristic	Studies (n=126), n (%)	Reference
Month of publication		
February	1 (0.8)	[98]
April	14 (11.1)	[26,41,52,54,55,62,69,77,91,102,105,114,117,122]
May	31 (24.6)	[17,18,21,30,38,42,43,51,53,61,63-65,68,70,73,75,76,81,82,86,87,90,95,100,101,116,125,129,130,138]
June	39 (31.0)	[13,20,22,24,25,27-29,31,32,34,36,37,40,44,46,48,50,56,58,67,72,78-80,83,84,89,94,96, 99,103,104,108,113,120,124,132,137]
July	30 (23.8)	[15,16,23,33,35,39,45,47,49,57,59,66,71,74,85,88,93,97,106,107,109,112,123,126,128,131,133-136]
August	11 (8.7)	[14,19,60,92,110,111,115,118,119,121,127]
Type of publication		
Journal article	123 (97.6)	[13-74,76-110,112-120,122-138]
Report	2 (1.6)	[75,111]
Preprint	1 (0.8)	[121]
Country of publication		
United States	51 (40.5)	[16,18,20,25,30,31,36,37,39,41,46,47,49,50,56,58,67,69,70,72-76,79,84,85,87,88,92,93, 97,102,105,108-115,117,121,125,126,129,133,134,136,137]
China	15 (11.9)	[14,15,44,51,52,71,77,81,96,98,107,122,127,128,138]
Italy	11 (8.7)	[19,27,34,45,55,60,62,65,90,99,100]
India	10 (7.9)	[23,24,28,38,53,63,64,101,103,124]
United Kingdom	7 (5.6)	[43,66,82,104,119,120,130]
Canada	4 (3.2)	[42,78,91,95]
Spain	4 (3.2)	[29,33,48,106]
Germany	3 (2.4)	[32,123,132]
Japan	3 (2.4)	[59,80,131]
Australia	2 (1.6)	[35,61]
Denmark	2 (1.6)	[21,40]
Netherlands	2 (1.6)	[86,94]
Poland	2 (1.6)	[22,68]
Argentina	1 (0.8)	[54]
Austria	1 (0.8)	[17]
Bahrain	1 (0.8)	[13]
Brazil	1 (0.8)	[116]
Egypt	1 (0.8)	[57]
Mexico	1 (0.8)	[26]
Serbia	1 (0.8)	[83]
Singapore	1 (0.8)	[118]
South Africa	1 (0.8)	[135]
Taiwan	1 (0.8)	[89]

Characteristics of Technologies

The 126 studies featured 11 different types of technologies used during the first wave of COVID-19 (Table 2). The most common type was telemedicine, which was used in approximately 84.9% of the studies (107/126). The mode of telemedicine was synchronous in 83 studies, asynchronous in 8 studies, and both in 17 studies. Technologies (Multimedia Appendix 4) were used for 20 different purposes. The most common aims of the technologies were providing consultations (75/126, 59.5%), following up with patients (45/126, 35.7%), and monitoring patient health status (22/126, 17.4%).

Table 2. Characteristics of technologies in the studies.

Characteristics	Studies (n=126), n (%)	Reference
Technology type[a]		
Telemedicine	107 (84.9)	[13-119]
Clinical decision support tools	11 (8.7)	[78,80-83,88,89,117,120-122]
Robotic systems	6 (4.8)	[123-128]
Symptom trackers	6 (4.8)	[78,122,129-132]
Dashboards	5 (4.0)	[78,93,117,133,134]
Electronic health records	4 (3.2)	[44,117,132,135]
Patient portals	3 (2.4)	[76,83,117]
Educational platforms	2 (1.6)	[136,137]
Triage tools	1 (0.8)	[93]
Reporting system	1 (0.8)	[83]
Low-dose computed tomography	1 (0.8)	[138]
Mode of telemedicine[b]		
Synchronous	83 (76.9)	[13-22,25,26,28,30,31,33,34,36-45,47,49-51,54-59,62-64,66-76,78,79,82-89,92,93,96,97,99-103,105,107-116,118,119]
Asynchronous	8 (7.4)	[23,24,48,81,91,94,98,129]
Both	17 (15.7)	[27,29,32,35,46,52,53,60,61,65,77,80,90,95,104,106,117]
Technology aim[a]		
Consultation	75 (59.5)	[13-87]
Follow up	45 (35.7)	[50-81,90,95-102,106,107,116,138]
Monitoring health status	22 (17.4)	[78,85-87,93,94,103-108,117,122,126-132,134]
Education	14 (11.1)	[46,47,49-52,77,91-94,107,136,137]
Triage	12 (9.5)	[52,78,79,82-84,87-90,93,117]
Treatment	12 (9.5)	[47,79,85,86,110-116,125]
Diagnosing	8 (6.3)	[45-47,89,117,126-128]
Screening	4 (3.2)	[80,81,109,124]
Accessing patient records	4 (3.2)	[44,117,132,135]
Monitoring health services	4 (3.2)	[93,117,133,134]
Decision making	3 (2.4)	[117,120,122]
Generating reports	3 (2.4)	[78,83,117]
Clinical assessment	2 (1.6)	[92,119]
Administrative support	2 (1.6)	[118,119]
Medical data exchange	2 (1.6)	[83,117]
Connecting patients and families	2 (1.6)	[93,108]
Booking appointments	2 (1.6)	[78,117]
Drug delivery	1 (0.8)	[48]

Characteristics	Studies (n=126), n (%)	Reference
Prognosis	1 (0.8)	[121]
Personal assistant	1 (0.8)	[123]
Technology development[a,c]		
Built for purpose	46 (58.2)	[20,25,29,34,41,44,45,47,49,51,56,58,61,69,70,73,76,80,81,83,86,89-91,93-95,101,104-107,114,115,119,120,122,123,125-128,131,132,135,136]
Purpose-shifted	45 (57.0)	[14-16,18,20,23-26,31,34,37-39,47,50,53,54,57,62-64,67,69-72,74-77,87,90,96,97,100,101,107,108,111,113,115,117,120,137]
Social media and videoconferencing platforms[a]		
Zoom	22 (17.5)	[16,25,31,34,37,50,57,62,67,69,70,72,74-77,87,108,111,113,115,137]
WhatsApp	12 (9.5)	[23-25,38,53,54,57,62,63,90,100,120]
FaceTime	6 (4.8)	[20,72,76,97,101,108]
WeChat	5 (4.0)	[14,15,71,96,107]
WebEx	5 (4.0)	[47,64,108,111,115]
Google Duo	5 (4.0)	[20,25,62,76,101]
Skype	4 (3.2)	[26,62,76,100]
Telegram	3 (2.4)	[25,62,90]
Messenger	2 (1.6)	[25,90]
Target users		
Health consumers and health care professionals	103 (81.7)	[13-32,34-41,43,45-60,62-84,86-93,95-102,104-118,123,131,132,138]
Health care professionals	15 (11.9)	[33,42,44,59,85,103,119,121,125-128,135-137]
Health consumers	5 (4.0)	[61,94,124,129,130]
Decision makers	3 (2.4)	[122,133,134]
Target conditions[a]		
COVID-19	38 (30.2)	[26,29,44,45,52,75,78,81,85,88,89,93-95,103,104,106-109,117-119,121,122,124,126-136,138]
Any conditions	21 (16.7)	[15,18,25,28,35,40,48,49,54,55,73,79,80,83,84,90,100,102,105,123,137]
Mental conditions	13 (10.3)	[31,33,42,50,74,76,87,91,111-115]
Cardiovascular conditions	9 (7.1)	[14,61,65,68,86,96,97,120,125]
Cancer conditions	6 (4.8)	[16,32,46,66,77,82]
Neurological conditions	6 (4.8)	[20,23,30,58,62,101]
Prenatal and postnatal conditions	6 (4.8)	[41,47,56,59,60,110]
Diabetes	5 (4.0)	[13,17,21,37,99]
Orthodontic conditions	4 (3.2)	[39,43,69,70]
Rheumatic conditions	4 (3.2)	[22,27,38,63]
Ophthalmic conditions	3 (2.4)	[24,53,71]
Urologic conditions	3 (2.4)	[72,92,134]
Dermatological conditions	3 (2.4)	[19,57,64]

Characteristics	Studies (n=126), n (%)	Reference
Ear, nose, and throat conditions	2 (1.6)	[67,116]
Liver conditions	2 (1.6)	[36,98]
Gastrointestinal conditions	1 (0.8)	[36,91]
Orthodontic conditions	1 (0.8)	[34]
Reproductive conditions	1 (0.8)	[87]
Transplant conditions	1 (0.8)	[51]
Setting[a]		
Hospitals	81 (64.3)	[13-15,17,18,21-23,27-35,37,41-44,48,50-60,64,66,69-73,77,79,80,82,85,88-90,92-94,96-100,102-108,110,116-123,126-128,133-135,137,138]
Medical clinics	46 (36.5)	[16,19,20,24-26,36,38-41,45-47,49,52,57,61-63,65,67,68,74-76,83,84,86-88,93,95,101,109,111-115,117,123,125,131,132]
Community	7 (5.6)	[78,91,123,124,129,130,136]
Internet connectivity[a]		
Web-based	106 (84.1)	[13-16,18-20,23-32,34-39,41,43-54,56-58,60-65,67,69-81,83-90,92-98,100-115,117,119,120,122,125-133,135-137]
Non–web-based	63 (50)	[13,17,20-24,26-33,35,40,42,43,46,49,52,53,55,59-62,65-68,74,76,77,79,81,82,84,88-93,95,97,99,101,102,104,106,112,115-118,120,121,123,124,134,138]
Modes[a]		
Mobile apps	86 (68.3)	[13-16,18-20,23-26,28,30-32,34,36-39,41,43-45,47,49,50,53,54,56,57,61-65,67,69-78,80,81,84-88,90,92-94,96-98,100-115,117,120,122,129-132,135,137]
Desktop apps	73 (57.9)	[14,16,19,20,23-26,28-32,34-38,47,50,52-54,57,58,62-65,67,69-78,80,82,83,85,87-90,92,93,96,100,101,105,107-111,113-115,119-121,131,133,134,137,138]
Telephone calls	49 (38.9)	[13,17,20-24,27-33,35,40,42,43,46,49,52,53,55,59-62,65-68,76,77,79,82,84,90,92,97,99,101,102,104,106,112,115-118]
Websites	45 (35.7)	[14-16,20,23-26,31,34,37,38,46,47,50,51,53,57,62-64,67,69-72,74-77,83,87,90,95,96,100,101,107,108,111,113,115,120,136,137]
Emails	13 (10.3)	[27,29,32,35,46,53,57,60,65,77,90,95,117]
Robot	6 (4.8)	[123-128]
Text messages	2 (1.6)	[48,91]
Interphone	1 (0.8)	[26]
Automated vital-sign monitor	1 (0.8)	[93]
Closed-circuit television cameras	1 (0.8)	[103]
Headset	1 (0.8)	[119]

[a]Numbers do not add to total.

[b]Number of telemedicine studies (n=108) was used to calculate percentages.

[c]We were able to identify the type of technology development in 79 studies; therefore, we used this number to calculate percentages.

The type of technology development was identified in 79 studies (Table 2). Technologies in these 79 studies were designed to provide the respective purpose from the beginning (46/79, 58.2%), or used for other purposes rather than the purpose for which they had been developed (45/79, 57.0%). Studies utilized 8 social media and videoconferencing platforms. The most common platforms were Zoom (22/126, 17.5%) and WhatsApp (12/126, 9.5%). The target users in the studies were health care professionals (15/126, 11.9%), health consumers (5/126, 4%), health care professionals as well as health consumers (103/126, 81.7%), and decision makers (3/126, 2.4%). Technologies in the studies targeted 19 groups of health conditions. The most targeted health condition by the technologies was COVID-19 (38/126, 30.2%), followed by any physical health conditions

(21/126, 16.7%) and mental health conditions (13/126, 10.3%). Multimedia Appendix 5 shows characteristics of the technologies used for COVID-19.

Technologies provided services to individuals in hospitals in 64.3% (81/126), medical clinics in 36.5% (46/126), and the community in 5.6% (7/126) of the studies. Technologies were web-based in 84.1% of the studies (106/126) and non–web-based in half of the included studies. Technologies were used through 11 modes. The most common mode was mobile apps (86/126, 68.3%), followed by desktop apps (73/126, 57.9%), telephone calls (49/126, 38.9%), and websites (45/126, 35.7%).

Discussion

Principal Findings

This review features all technologies utilized in the first efforts to provide health care services for all kinds of clinical conditions during the pandemic, which ideally provides a holistic perspective of technology use during the initial stages of the COVID-19 pandemic. Only 1 paper [98] was published in the first 3 months of the pandemic, whereas the number of the studies significantly increased in subsequent months. This may be attributed to the fact that the process of developing technologies, writing a report on its use, and publishing the report likely require at least 3 months. Furthermore, experts suggest that evidence-based technology adoption is critical for designing, using, and implementing digital tools in the fight against COVID-19 [139]. Most technologies were implemented in the United States. This is expected given that the United States is both the country most infected by COVID-19 [1] as well as highly advanced in technologies.

As evident from the results, telemedicine was the most frequently utilized technology in papers published from January to August 2020. One of the main challenges during the first few months of the pandemic was providing necessary health care services while simultaneously mitigating the risk of infection for patients and health care workers from inadequate supplies of personal protective equipment. The majority of studies included in this review reported the use of synchronous telemedicine, which can improve access to specialized care services at familiar settings (ie, homes) in a more time- and cost-effective manner [140,141]. Similar to earlier reviews, our results showed that telemedicine was used for COVID-19 screening, providing health care advice, and mental health therapy [8]. The remainder of the technologies reported in the studies involved the use of clinical decision support tools, robotic systems, and symptom trackers. While many other technologies were implemented during the first wave of COVID-19 (eg, contact-tracing apps and drones [11]), our review did not report any. This could be attributed to the fact that (1) there are no publications on each technology implemented during COVID-19, (2) only studies published in English were included in this review, and (3) only academic literature was searched.

Several social media and videoconferencing platforms were used in the first wave. Zoom was the most commonly used; this may be attributed to the fact that Zoom is one of the Health Insurance Portability and Accountability Act–compliant teleconferencing platforms and has been integrated into electronic health record systems in many health care settings [31,67,87]. This review also found that most of the implemented technologies targeted both health care providers and health consumers. This is likely because extreme measures (eg, lockdowns; curfews; and closures of many businesses, spaces, and organizations) imposed to prevent the spread of COVID-19 immensely disrupted the routine delivery of health care services to many patients—especially those with chronic conditions—therefore, technologies were required to keep patients and health care providers connected.

COVID-19 was the most targeted health condition by the technologies. This can be attributed to the fact that COVID-19 was given priority over other health conditions in the first wave, when there were extremely large numbers of daily infections, and no treatment or vaccine was available. Mobile apps were the most common venue used to target illness during this time period, as well. This is reasonable given the widespread use of mobile devices over the world [142].

Research and Practical Implications

Modern medicine relies on evidence-based practices and interventions. Many researchers argue that digital health tools and technologies are not exempt and should be evaluated and validated to demonstrate their efficacy [143]. Widespread technology adoption relies heavily upon the availability of evidence regarding technology's effectiveness [144]. While there is a sense of urgency because of the COVID-19 pandemic and its impact on every aspect of daily living, evaluation of digital tools' and technologies' effectiveness remains paramount. However, since the COVID-19 pandemic is still ongoing, there are copious amounts of digital tools and technologies that need to be peer-reviewed, undergo rigorous testing, and be integrated into public health systems [11].

Many digital interventions have failed to be integrated into national systems due to low levels of acceptance at that scale; many digital interventions have been successful at pilot scales only [145,146]. Therefore, it is integral to carefully evaluate digital health tools and technologies by providing validated, documented, and reproducible clinically meaningful evidence [143,147]. There is a need for more peer-reviewed evidence-based research into the effectiveness of these digital tools and technologies.

Overall, there is emerging consensus that digital tools and technologies play a central role in public health interventions during public health outbreaks (eg, the COVID-19 pandemic) by complementing and augmenting traditional or nondigital interventions [11]. However, despite the promise of technologies' benefits, such as alleviating the impact of COVID-19 (such as providing access to health care services during lockdowns through telemedicine), solely relying on these technologies may widen disparities and exacerbate existing vulnerabilities of those who are unable to adopt or afford these technologies [148]. These digital tools and technologies pose the risk of generating informatics-based (or digital) inequalities that disproportionately favor groups with socioeconomic advantages [149]. Therefore, health system leaders,

informaticians, and policy makers need to consider and act upon underlying factors contributing to inequalities in terms of access to resources, digital literacy, and usability [150].

To ensure sustainable investments, use, and adoption in digital tools and technologies after the COVID-19 pandemic has been resolved, issues related to regulating and reimbursing the use of these tools and technologies need to be addressed [151]. Additionally, it is paramount that it is not assumed that digital tools and technologies will naturally become integrated into routine clinical practice post–COVID-19; participatory approaches that include all stakeholders are essential to sustaining the use and adoption of digital tools and technologies [152].

Strengths and Limitations

Strengths

To the best of our knowledge, this paper is the first review to explore all technologies implemented to provide health care services during COVID-19. This review can be considered comprehensive as it does not focus on specific technologies, diseases, users, settings, or countries. Therefore, the review provides a holistic view of the role of technologies during COVID-19 to help decision makers, health care providers, and health care consumers understand the potentials of technologies.

Given that we followed well-recommended guidelines in developing, executing, and reporting this review, it can be considered a robust and high-quality review. The search was sensitive and precise because the most popular databases in health and information technology were searched using a well-developed search query.

The risk of publication bias is minimal in this review because of the strategies that were used, such as searching grey literature databases (ie, Google Scholar) and backward and forward reference list checking. Furthermore, this review has a low risk of selection bias because study selection and data extraction were conducted by 2 reviewers independently.

Limitations

Although the search query consisted of 89 terms related to technologies, it is possible that other terms related to technologies were missed. Therefore, it is likely that several relevant studies were not included. Because of practical constraints, the search was restricted to English studies. Therefore, several studies written in other languages were not included.

This review focused on implemented technologies and excluded proposed technologies, which perhaps could have been subsequently implemented. Thus, several important technologies were not considered in our review. Although this review shows the potential of technologies, it cannot comment on their effectiveness, as it is beyond the scope.

Conclusion

Technologies played a crucial role in mitigating challenges arising from the COVID-19 pandemic, and in the first wave, numerous technologies were used for various purposes. However, we did not find papers on other technologies such as contact-tracing apps, drones, blockchain, unmanned aerial vehicles, wearable devices, and personal protective equipment that were implemented during the first wave of the pandemic. Furthermore, technologies reported in the studies were used for other purposes, such as drug and vaccine discovery, social distancing, and immunity passports. Therefore, it is recommended that researchers conduct studies on these technologies and purposes. Further reviews are required about technologies implemented in subsequent waves of the COVID-19 pandemic. There is also a need for more peer-reviewed evidence-based research into the effectiveness of these digital tools and technologies and users' satisfaction.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 52 KB - jmir_v23i9e29136_app1.docx]

Multimedia Appendix 2
Data extraction form.
[DOCX File , 18 KB - jmir_v23i9e29136_app2.docx]

Multimedia Appendix 3
Characteristics of each included study.
[DOCX File , 32 KB - jmir_v23i9e29136_app3.docx]

Multimedia Appendix 4
Characteristics of technologies in each included study.
[DOCX File , 61 KB - jmir_v23i9e29136_app4.docx]

Multimedia Appendix 5
Characteristics of technologies used for COVID-19, as reported in 38 studies.
[DOCX File , 395 KB - jmir_v23i9e29136_app5.docx]

References

1. WHO Coronavirus Disease (COVID-19) dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2001-04-21]
2. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: a surveillance study. J Med Internet Res 2020 Apr 09;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
3. Abd-alrazaq A, Alajlani M, Alhuwail D, Erbad A, Giannicchi A, Shah Z. Blockchain technologies to mitigate COVID-19 challenges: a scoping review. Comput Methods Programs Biomed Update 2021;1:1-8 [FREE Full text] [doi: 10.1016/j.cmpbup.2020.100001]
4. Pericàs JM, Cucchiari D, Torrallardona-Murphy O, Calvo J, Serralabós J, Alvés E, Hospital Clínic 4H Team (Hospital at Home-Health Hotel). Hospital at home for the management of COVID-19: preliminary experience with 63 patients. Infection 2021 Apr;49(2):327-332 [FREE Full text] [doi: 10.1007/s15010-020-01527-z] [Medline: 32995970]
5. Madurai Elavarasan R, Pugazhendhi R. Restructured society and environment: a review on potential technological strategies to control the COVID-19 pandemic. Sci Total Environ 2020 Jul 10;725:138858 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.138858] [Medline: 32336562]
6. Kong WH, Li Y, Peng MW, Kong DG, Yang XB, Wang L, et al. SARS-CoV-2 detection in patients with influenza-like illness. Nat Microbiol 2020 May;5(5):675-678. [doi: 10.1038/s41564-020-0713-1] [Medline: 32265517]
7. Davalbhakta S, Advani S, Kumar S, Agarwal V, Bhoyar S, Fedirko E, et al. A systematic review of smartphone applications available for corona virus disease 2019 (COVID19) and the assessment of their quality using the Mobile Application Rating Scale (MARS). J Med Syst 2020 Aug 10;44(9):164 [FREE Full text] [doi: 10.1007/s10916-020-01633-3] [Medline: 32779002]
8. Gao Y, Liu R, Zhou Q, Wang X, Huang L, Shi Q, COVID-19 Evidence and Recommendations Working Group. Application of telemedicine during the coronavirus disease epidemics: a rapid review and meta-analysis. Ann Transl Med 2020 May;8(10):626 [FREE Full text] [doi: 10.21037/atm-20-3315] [Medline: 32566563]
9. Chamola V, Hassija V, Gupta V, Guizani M. A comprehensive review of the COVID-19 pandemic and the role of IoT, drones, AI, blockchain, and 5G in managing its impact. IEEE Access 2020;8:90225-90265. [doi: 10.1109/access.2020.2992341]
10. Abd-Alrazaq A, Alajlani M, Alhuwail D, Schneider J, Al-Kuwari S, Shah Z, et al. Artificial intelligence in the fight against COVID-19: scoping review. J Med Internet Res 2020 Dec 15;22(12):e20756 [FREE Full text] [doi: 10.2196/20756] [Medline: 33284779]
11. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020 Aug 07;26(8):1183-1192. [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]
12. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
13. Alromaihi D, Alamuddin N, George S. Sustainable diabetes care services during COVID-19 pandemic. Diabetes Res Clin Pract 2020 Aug;166:108298 [FREE Full text] [doi: 10.1016/j.diabres.2020.108298] [Medline: 32623031]
14. Li L, Liu G, Xu W, Zhang Y, He M. Effects of internet hospital consultations on psychological burdens and disease knowledge during the early outbreak of COVID-19 in China: cross-sectional survey study. J Med Internet Res 2020 Aug 04;22(8):e19551 [FREE Full text] [doi: 10.2196/19551] [Medline: 32687061]
15. Liu L, Gu J, Shao F, Liang X, Yue L, Cheng Q, et al. Application and preliminary outcomes of remote diagnosis and treatment during the COVID-19 outbreak: retrospective cohort study. JMIR Mhealth Uhealth 2020 Jul 03;8(7):e19417 [FREE Full text] [doi: 10.2196/19417] [Medline: 32568722]
16. Lonergan PE, Washington Iii SL, Branagan L, Gleason N, Pruthi RS, Carroll PR, et al. Rapid utilization of telehealth in a comprehensive cancer center as a response to COVID-19. J Med Internet Res 2020 Jun 21;22(7):e19322 [FREE Full text] [doi: 10.2196/19322] [Medline: 32568721]
17. Mader JK. Personal experiences with coronavirus disease 2019 and diabetes: the time for telemedicine is now. J Diabetes Sci Technol 2020 Jul;14(4):752-753 [FREE Full text] [doi: 10.1177/1932296820930289] [Medline: 32443942]
18. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
19. Marasca C, De Rosa A, Fabbrocini G, Cantelli M, Patrì A, Vastarella M, et al. Psychological teleconsultations in patients suffering from chronic skin diseases during the COVID-19 era: a service to improve patients' quality of life. J Dermatolog Treat 2020 Oct 12:1-2. [doi: 10.1080/09546634.2020.1809625] [Medline: 32783664]

20. McGinley MP, Ontaneda D, Wang Z, Weber M, Shook S, Stanton M, et al. Teleneurology as a solution for outpatient care during the COVID-19 pandemic. Telemed J E Health 2020 Dec;26(12):1537-1539. [doi: 10.1089/tmj.2020.0137] [Medline: 32552509]
21. Nørgaard K. Telemedicine consultations and diabetes technology during COVID-19. J Diabetes Sci Technol 2020 Jul;14(4):767-768 [FREE Full text] [doi: 10.1177/1932296820929378] [Medline: 32429702]
22. Opinc A, Łukasik Z, Makowska J. The attitude of Polish rheumatology patients towards telemedicine in the age of the COVID-19 pandemic. Reumatologia 2020;58(3):134-141 [FREE Full text] [doi: 10.5114/reum.2020.96665] [Medline: 32684645]
23. Panda PK, Dawman L, Panda P, Sharawat IK. Feasibility and effectiveness of teleconsultation in children with epilepsy amidst the ongoing COVID-19 pandemic in a resource-limited country. Seizure 2020 Oct;81:29-35 [FREE Full text] [doi: 10.1016/j.seizure.2020.07.013] [Medline: 32712376]
24. Pandey N, Srivastava RM, Kumar G, Katiyar V, Agrawal S. Teleconsultation at a tertiary care government medical university during COVID-19 lockdown in India - a pilot study. Indian J Ophthalmol 2020 Jul;68(7):1381-1384 [FREE Full text] [doi: 10.4103/ijo.IJO_1658_20] [Medline: 32587169]
25. Peden CJ, Mohan S, Pagán V. Telemedicine and COVID-19: an observational study of rapid scale up in a US academic medical system. J Gen Intern Med 2020 Sep 04;35(9):2823-2825 [FREE Full text] [doi: 10.1007/s11606-020-05917-9] [Medline: 32500329]
26. Perez-Alba E, Nuzzolo-Shihadeh L, Espinosa-Mora JE, Camacho-Ortiz A. Use of self-administered surveys through QR code and same center telemedicine in a walk-in clinic in the era of COVID-19. J Am Med Inform Assoc 2020 Jun 01;27(6):985-986 [FREE Full text] [doi: 10.1093/jamia/ocaa054] [Medline: 32282922]
27. Perniola S, Alivernini S, Varriano V, Paglionico A, Tanti G, Rubortone P, et al. Telemedicine will not keep us apart in COVID-19 pandemic. Ann Rheum Dis 2020 Jun 05:1. [doi: 10.1136/annrheumdis-2020-218022] [Medline: 32503851]
28. Poulose BK, Phieffer LS, Mayerson J, Like D, Forrest LA, Rahmanian A, et al. Responsible return to essential and non-essential surgery during the COVID-19 pandemic. J Gastrointest Surg 2020 Jun 04:1105-1107 [FREE Full text] [doi: 10.1007/s11605-020-04673-9] [Medline: 32500416]
29. Rabuñal R, Suarez-Gil R, Golpe R, Martínez-García M, Gómez-Méndez R, Romay-Lema E, et al. Usefulness of a telemedicine tool TELEA in the management of the COVID-19 pandemic. Telemed J E Health 2020 Nov;26(11):1332-1335. [doi: 10.1089/tmj.2020.0144] [Medline: 32501747]
30. Rametta SC, Fridinger SE, Gonzalez AK, Xian J, Galer PD, Kaufman M, et al. Analyzing 2,589 child neurology telehealth encounters necessitated by the COVID-19 pandemic. Neurology 2020 Sep 01;95(9):e1257-e1266 [FREE Full text] [doi: 10.1212/WNL.0000000000010010] [Medline: 32518152]
31. Ramtekkar U, Bridge JA, Thomas G, Butter E, Reese J, Logan E, et al. Pediatric telebehavioral health: a transformational shift in care delivery in the era of COVID-19. JMIR Ment Health 2020 Sep 18;7(9):e20157 [FREE Full text] [doi: 10.2196/20157] [Medline: 32525485]
32. Rodler S, Apfelbeck M, Schulz GB, Ivanova T, Buchner A, Staehler M, et al. Telehealth in uro-oncology beyond the pandemic: toll or lifesaver? Eur Urol Focus 2020 Sep 15;6(5):1097-1103 [FREE Full text] [doi: 10.1016/j.euf.2020.05.010] [Medline: 32534969]
33. Roncero C, García-Ullán L, de la Iglesia-Larrad JI, Martín C, Andrés P, Ojeda A, Montejo. The response of the mental health network of the Salamanca area to the COVID-19 pandemic: the role of the telemedicine. Psychiatry Res 2020 Sep;291:113252 [FREE Full text] [doi: 10.1016/j.psychres.2020.113252] [Medline: 32623263]
34. Saccomanno S, Quinzi V, Sarhan S, Laganà D, Marzo G. Perspectives of tele-orthodontics in the COVID-19 emergency and as a future tool in daily practice. Eur J Paediatr Dent 2020 Jun;21(2):157-162 [FREE Full text] [doi: 10.23804/ejpd.2020.21.02.12] [Medline: 32567949]
35. Schulz T, Long K, Kanhutu K, Bayrak I, Johnson D, Fazio T. Telehealth during the coronavirus disease 2019 pandemic: rapid expansion of telehealth outpatient use during a pandemic is possible if the programme is previously established. J Telemed Telecare 2020 Jul 19:1357633X20942045. [doi: 10.1177/1357633X20942045] [Medline: 32686556]
36. Serper M, Nunes F, Ahmad N, Roberts D, Metz DC, Mehta SJ. Positive early patient and clinician experience with telemedicine in an academic gastroenterology practice during the COVID-19 pandemic. Gastroenterology 2020 Oct;159(4):1589-1591.e4 [FREE Full text] [doi: 10.1053/j.gastro.2020.06.034] [Medline: 32565015]
37. Shelton C, Demidowich AP, Zilbermint M. Inpatient diabetes management during the COVID-19 crisis: experiences from two community hospitals. J Diabetes Sci Technol 2020 Jul;14(4):780-782 [FREE Full text] [doi: 10.1177/1932296820930268] [Medline: 32486900]
38. Shenoy P, Ahmed S, Paul A, Skaria TG, Joby J, Alias B. Switching to teleconsultation for rheumatology in the wake of the COVID-19 pandemic: feasibility and patient response in India. Clin Rheumatol 2020 Sep;39(9):2757-2762 [FREE Full text] [doi: 10.1007/s10067-020-05200-6] [Medline: 32474883]
39. Siow MY, Walker JT, Britt E, Kozy JP, Zanzucchi A, Girard PJ, et al. What was the change in telehealth usage and proportion of no-show visits for an orthopaedic trauma clinic during the COVID-19 pandemic? Clin Orthop Relat Res 2020 Oct;478(10):2257-2263. [doi: 10.1097/CORR.0000000000001396] [Medline: 32639309]

40. Wolthers TO, Wolthers OD. Telephone consultation as a substitute for face-to-face consultation during the COVID-19 pandemic. Dan Med J 2020 Jun 23;67(7):A04200300 [FREE Full text] [Medline: 32734880]
41. Aziz A, Zork N, Aubey JJ, Baptiste CD, D'Alton ME, Emeruwa UN, et al. Telehealth for high-risk pregnancies in the setting of the COVID-19 pandemic. Am J Perinatol 2020 Jun;37(8):800-808 [FREE Full text] [doi: 10.1055/s-0040-1712121] [Medline: 32396948]
42. Geoffroy PA, Le Goanvic V, Sabbagh O, Richoux C, Weinstein A, Dufayet G, et al. Psychological support system for hospital workers during the Covid-19 outbreak: rapid design and implementation of the Covid-Psy Hotline. Front Psychiatry 2020 May 28;11:511 [FREE Full text] [doi: 10.3389/fpsyt.2020.00511] [Medline: 32670100]
43. Gilbert AW, Billany JCT, Adam R, Martin L, Tobin R, Bagdai S, et al. Rapid implementation of virtual clinics due to COVID-19: report and early evaluation of a quality improvement initiative. BMJ Open Qual 2020 May 21;9(2):e000985 [FREE Full text] [doi: 10.1136/bmjoq-2020-000985] [Medline: 32439740]
44. Ren X, Zhai Y, Song X, Wang Z, Dou D, Li Y. The application of mobile telehealth system to facilitate patient information presentation and case discussion. Telemed J E Health 2020 Jun;26(6):725-733. [doi: 10.1089/tmj.2020.0084] [Medline: 32298208]
45. Sossai P, Uguccioni S, Casagrande S. Telemedicine and the 2019 coronavirus (SARS-CoV-2). Int J Clin Pract 2020 Oct 08;74(10):e13592 [FREE Full text] [doi: 10.1111/ijcp.13592] [Medline: 32563197]
46. DiGiovanni G, Mousaw K, Lloyd T, Dukelow N, Fitzgerald B, D'Aurizio H, et al. Development of a telehealth geriatric assessment model in response to the COVID-19 pandemic. J Geriatr Oncol 2020 Jun;11(5):761-763 [FREE Full text] [doi: 10.1016/j.jgo.2020.04.007] [Medline: 32327321]
47. Dosaj A, Thiyagarajan D, Ter Haar C, Cheng J, George J, Wheatley C, et al. Rapid implementation of telehealth services during the COVID-19 pandemic. Telemed J E Health 2021 Feb;27(2):116-120. [doi: 10.1089/tmj.2020.0219] [Medline: 32706616]
48. Margusino-Framiñán L, Illarro-Uranga A, Lorenzo-Lorenzo K, Monte-Boquet E, Márquez-Saavedra E, Fernández-Bargiela N, et al. Pharmaceutical care to hospital outpatients during the COVID-19 pandemic. Telepharm Farm Hosp 2020 Jun 13;44(7):61-65 [FREE Full text] [doi: 10.7399/fh.11498] [Medline: 32533674]
49. McElroy JA, Day TM, Becevic M. The influence of telehealth for better health across communities. Prev Chronic Dis 2020 Jul 16;17:E64 [FREE Full text] [doi: 10.5888/pcd17.200254] [Medline: 32678060]
50. Datta N, Derenne J, Sanders M, Lock JD. Telehealth transition in a comprehensive care unit for eating disorders: challenges and long-term benefits. Int J Eat Disord 2020 Nov;53(11):1774-1779. [doi: 10.1002/eat.23348] [Medline: 32715512]
51. Zhao Y, Wei L, Liu B, Du D. Management of transplant patients outside hospital during COVID-19 epidemic: a Chinese experience. Transpl Infect Dis 2020 Oct;22(5):e13327 [FREE Full text] [doi: 10.1111/tid.13327] [Medline: 32407003]
52. Song X, Liu X, Wang C. The role of telemedicine during the COVID-19 epidemic in China-experience from Shandong province. Crit Care 2020 Apr 28;24(1):178 [FREE Full text] [doi: 10.1186/s13054-020-02884-9] [Medline: 32345359]
53. Das AV, Rani PK, Vaddavalli PK. Tele-consultations and electronic medical records driven remote patient care: responding to the COVID-19 lockdown in India. Indian J Ophthalmol 2020 Jun;68(6):1007-1012 [FREE Full text] [doi: 10.4103/ijo.IJO_1089_20] [Medline: 32461415]
54. Di Tommasso F, Fitz Maurice M, Sastre P, Hirschson Prado A, Dominé E, Agüero P, et al. WhatsApp consultations in the department of electrophysiology of a public hospital of the city of Buenos Aires in times of COVID-19. Rev Argent Cardiol 2020 Jun;88(3):225-228. [doi: 10.7775/rac.v88.i3.17976]
55. Luciani LG, Mattevi D, Cai T, Giusti G, Proietti S, Malossini G. Teleurology in the time of Covid-19 pandemic: here to stay? Urology 2020 Jun;140:4-6 [FREE Full text] [doi: 10.1016/j.urology.2020.04.004] [Medline: 32298686]
56. Madden N, Emeruwa UN, Friedman AM, Aubey JJ, Aziz A, Baptiste CD, et al. Telehealth uptake into prenatal care and provider attitudes during the COVID-19 pandemic in New York City: a quantitative and qualitative analysis. Am J Perinatol 2020 Aug;37(10):1005-1014 [FREE Full text] [doi: 10.1055/s-0040-1712939] [Medline: 32516816]
57. Mostafa PIN, Hegazy AA. Dermatological consultations in the COVID-19 era: is teledermatology the key to social distancing? an Egyptian experience. J Dermatolog Treat 2020 Jul 07:1-6. [doi: 10.1080/09546634.2020.1789046] [Medline: 32602763]
58. Mouchtouris N, Lavergne P, Montenegro TS, Gonzalez G, Baldassari M, Sharan A, et al. Telemedicine in neurosurgery: lessons learned and transformation of care during the COVID-19 pandemic. World Neurosurg 2020 Aug;140:e387-e394 [FREE Full text] [doi: 10.1016/j.wneu.2020.05.251] [Medline: 32512241]
59. Nakagawa K, Umazume T, Mayama M, Chiba K, Saito Y, Kawaguchi S, et al. Feasibility and safety of urgently initiated maternal telemedicine in response to the spread of COVID-19: a 1-month report. J Obstet Gynaecol Res 2020 Oct;46(10):1967-1971 [FREE Full text] [doi: 10.1111/jog.14378] [Medline: 32691488]
60. Pagliazzi A, Mancano G, Forzano G, di Giovanni F, Gori G, Traficante G, et al. Genetic counseling during COVID-19 pandemic: Tuscany experience. Mol Genet Genomic Med 2020 Oct;8(10):e1433 [FREE Full text] [doi: 10.1002/mgg3.1433] [Medline: 32743952]
61. Pluymaekers NA, Hermans AN, van der Velden RM, den Uijl DW, Vorstermans B, Buskes S, et al. On-demand app-based rate and rhythm monitoring to manage atrial fibrillation through teleconsultations during COVID-19. Int J Cardiol Heart Vasc 2020 Jun;28:100533 [FREE Full text] [doi: 10.1016/j.ijcha.2020.100533] [Medline: 32391412]

62. Prada V, Bellone E, Schenone A, Grandis M. The suspected SARS-Cov-2 infection in a Charcot-Marie-Tooth patient undergoing postsurgical rehabilitation: the value of telerehabilitation for evaluation and continuing treatment. Int J Rehabil Res 2020 Sep;43(3):285-286 [FREE Full text] [doi: 10.1097/MRR.0000000000000418] [Medline: 32317558]
63. Rastogi S, Singh N, Pandey P. On the brighter side of COVID-19 induced lockdown: devising the collateral methods to provide Ayurveda consultation during impasse. J Ayurveda Integr Med 2020 May 08:1-3 [FREE Full text] [doi: 10.1016/j.jaim.2020.05.001] [Medline: 32390696]
64. Rismiller K, Cartron AM, Trinidad JCL. Inpatient teledermatology during the COVID-19 pandemic. J Dermatolog Treat 2020 Aug 13;31(5):441-443. [doi: 10.1080/09546634.2020.1762843] [Medline: 32364809]
65. Salzano A, D'Assante R, Stagnaro FM, Valente V, Crisci G, Giardino F, et al. Heart failure management during the COVID-19 outbreak in Italy: a telemedicine experience from a heart failure university tertiary referral centre. Eur J Heart Fail 2020 Jun;22(6):1048-1050 [FREE Full text] [doi: 10.1002/ejhf.1911] [Medline: 32463534]
66. Smrke A, Younger E, Wilson R, Husson O, Farag S, Merry E, et al. Telemedicine During the COVID-19 Pandemic: Impact on Care for Rare Cancers. JCO Glob Oncol 2020 Jul;6:1046-1051 [FREE Full text] [doi: 10.1200/GO.20.00220] [Medline: 32639877]
67. Strohl MP, Dwyer CD, Ma Y, Rosen CA, Schneider SL, Young VN. Implementation of telemedicine in a laryngology practice during the COVID-19 pandemic: lessons learned, experiences shared. J Voice 2020 Jun 23:1-7 [FREE Full text] [doi: 10.1016/j.jvoice.2020.06.017] [Medline: 32778359]
68. Świerad M, Dyrbuś K, Szkodziński J, Zembala MO, Kalarus Z, Gąsior M. Telehealth visits in a tertiary cardiovascular center as a response of the healthcare system to the severe acute respiratory syndrome coronavirus 2 pandemic in Poland. Pol Arch Intern Med 2020 Aug 27;130(7-8):700-703 [FREE Full text] [doi: 10.20452/pamw.15370] [Medline: 32426953]
69. Tanaka MJ, Oh LS, Martin SD, Berkson EM. Telemedicine in the era of COVID-19: the virtual orthopaedic examination. J Bone Joint Surg Am 2020 Jun 17;102(12):e57 [FREE Full text] [doi: 10.2106/JBJS.20.00609] [Medline: 32341311]
70. Tenforde AS, Iaccarino MA, Borgstrom H, Hefner JE, Silver J, Ahmed M, et al. Telemedicine during COVID-19 for outpatient sports and musculoskeletal medicine physicians. PM R 2020 Sep;12(9):926-932 [FREE Full text] [doi: 10.1002/pmrj.12422] [Medline: 32424977]
71. Wang Y, Zhao L, Liu Z, Li X. Perioperative management by WeChat under the haze of COVID-19. Int J Ophthalmol 2020;13(7):1161-1163 [FREE Full text] [doi: 10.18240/ijo.2020.07.22] [Medline: 32685407]
72. Watts KL, Abraham N. "Virtually perfect" for some but perhaps not for all: launching telemedicine in the Bronx during the COVID-19 pandemic. J Urol 2020 Nov;204(5):903-904. [doi: 10.1097/JU.0000000000001185] [Medline: 32519903]
73. Wosik J, Fudim M, Cameron B, Gellad ZF, Cho A, Phinney D, et al. Telehealth transformation: COVID-19 and the rise of virtual care. J Am Med Inform Assoc 2020 Jun 01;27(6):957-962 [FREE Full text] [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]
74. Yellowlees P, Nakagawa K, Pakyurek M, Hanson A, Elder J, Kales HC. Rapid conversion of an outpatient psychiatric clinic to a 100% virtual telepsychiatry clinic in response to COVID-19. Psychiatr Serv 2020 Jul 01;71(7):749-752. [doi: 10.1176/appi.ps.202000230] [Medline: 32460683]
75. Zweig S. Patient-doctor telemedicine: virtual care in the era of COVID-19 and beyond. Mo Med 2020;117(3):175-176 [FREE Full text] [Medline: 32636539]
76. Sullivan AB, Kane A, Roth AJ, Davis BE, Drerup ML, Heinberg LJ. The COVID-19 crisis: a mental health perspective and response using telemedicine. J Patient Exp 2020 Jun;7(3):295-301 [FREE Full text] [doi: 10.1177/2374373520922747] [Medline: 32821785]
77. Lee AKF, Cho RHW, Lau EHL, Cheng HK, Wong EWY, Ku PKM, et al. Mitigation of head and neck cancer service disruption during COVID-19 in Hong Kong through telehealth and multi-institutional collaboration. Head Neck 2020 Jul;42(7):1454-1459 [FREE Full text] [doi: 10.1002/hed.26226] [Medline: 32357277]
78. Krausz M, Westenberg JN, Vigo D, Spence RT, Ramsey D. Emergency response to COVID-19 in Canada: platform development and implementation for eHealth in crisis management. JMIR Public Health Surveill 2020 May 15;6(2):e18995 [FREE Full text] [doi: 10.2196/18995] [Medline: 32401218]
79. Lau J, Knudsen J, Jackson H, Wallach AB, Bouton M, Natsui S, et al. Staying connected in the COVID-19 pandemic: telehealth at the largest ssafety-net system in the United States. Health Aff (Millwood) 2020 Aug;39(8):1437-1442. [doi: 10.1377/hlthaff.2020.00903] [Medline: 32525705]
80. Sampa MB, Hoque MR, Islam R, Nishikitani M, Nakashima N, Yokota F, et al. Redesigning portable health clinic platform as a remote healthcare system to tackle COVID-19 pandemic situation in unreached communities. Int J Environ Res Public Health 2020 Jun 30;17(13):4709 [FREE Full text] [doi: 10.3390/ijerph17134709] [Medline: 32629963]
81. Liu Y, Wang Z, Ren J, Tian Y, Zhou M, Zhou T, et al. A COVID-19 risk assessment decision support system for general practitioners: design and development study. J Med Internet Res 2020 Jun 29;22(6):e19786 [FREE Full text] [doi: 10.2196/19786] [Medline: 32540845]
82. Paleri V, Hardman J, Tikka T, Bradley P, Pracy P, Kerawala C. Rapid implementation of an evidence-based remote triaging system for assessment of suspected referrals and patients with head and neck cancer on follow-up after treatment during the COVID-19 pandemic: Model for international collaboration. Head Neck 2020 Jul;42(7):1674-1680 [FREE Full text] [doi: 10.1002/hed.26219] [Medline: 32374942]

83. Milenkovic A, Jankovic D, Rajkovic P. Extensions and adaptations of existing medical information system in order to reduce social contacts during COVID-19 pandemic. Int J Med Inform 2020 Sep;141:104224 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104224] [Medline: 32570196]
84. Ayoub P, Chang DD, Hussein N, Stewart K, Wise A, Malik I, et al. Medical student mobilization during a Pandemic: the Ochsner clinical school response to COVID-19. Ochsner J 2020;20(2):146-150 [FREE Full text] [doi: 10.31486/toj.20.0069] [Medline: 32612468]
85. Harris DA, Archbald-Pannone L, Kaur J, Cattell-Gordon D, Rheuban KS, Ombres RL, et al. Rapid telehealth-centered response to COVID-19 outbreaks in postacute and long-term care facilities. Telemed J E Health 2021 Jan;27(1):102-106. [doi: 10.1089/tmj.2020.0236] [Medline: 32644899]
86. Linz D, Pluymaekers NAHA, Hendriks JM. TeleCheck-AF for COVID-19. Eur Heart J 2020 Jun 01;41(21):1954-1955 [FREE Full text] [doi: 10.1093/eurheartj/ehaa404] [Medline: 32379309]
87. Barney A, Buckelew S, Mesheriakova V, Raymond-Flesch M. The COVID-19 pandemic and rapid implementation of adolescent and young adult telemedicine: challenges and opportunities for innovation. J Adolesc Health 2020 Aug;67(2):164-171 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.05.006] [Medline: 32410810]
88. Obeid JS, Davis M, Turner M, Meystre SM, Heider PM, O'Bryan EC, et al. An artificial intelligence approach to COVID-19 infection risk assessment in virtual visits: A case report. J Am Med Inform Assoc 2020 Aug 01;27(8):1321-1325 [FREE Full text] [doi: 10.1093/jamia/ocaa105] [Medline: 32449766]
89. Lin C, Tseng W, Wu J, Tay J, Cheng M, Ong H, et al. A double triage and telemedicine protocol to optimize infection control in an emergency department in Taiwan during the COVID-19 pandemic: retrospective feasibility study. J Med Internet Res 2020 Jun 23;22(6):e20586 [FREE Full text] [doi: 10.2196/20586] [Medline: 32544072]
90. Pignatti M, Pinto V, Miralles MEL, Giorgini FA, Cannamela G, Cipriani R. How the COVID-19 pandemic changed the Plastic Surgery activity in a regional referral center in Northern Italy. J Plast Reconstr Aesthet Surg 2020 Jul;73(7):1348-1356 [FREE Full text] [doi: 10.1016/j.bjps.2020.05.002] [Medline: 32499187]
91. Agyapong VIO. Coronavirus disease 2019 pandemic: health system and community response to a text message (Text4Hope) program supporting mental health in Alberta. Disaster Med Public Health Prep 2020 Oct;14(5):e5-e6 [FREE Full text] [doi: 10.1017/dmp.2020.114] [Medline: 32317038]
92. Yadav A, Caldararo K, Singh P. Optimising the use of telemedicine in a kidney transplant programme during the coronavirus disease 2019 pandemic. J Telemed Telecare 2020 Aug 06:1357633X20942632. [doi: 10.1177/1357633X20942632] [Medline: 32762269]
93. Salway RJ, Silvestri D, Wei EK, Bouton M. Using information technology to improve COVID-19 care at New York City health + hospitals. Health Aff (Millwood) 2020 Sep;39(9):1601-1604. [doi: 10.1377/hlthaff.2020.00930] [Medline: 32673131]
94. Timmers T, Janssen L, Stohr J, Murk JL, Berrevoets MAH. Using eHealth to support COVID-19 education, self-assessment, and symptom monitoring in the Netherlands: observational study. JMIR Mhealth Uhealth 2020 Jun 23;8(6):e19822 [FREE Full text] [doi: 10.2196/19822] [Medline: 32516750]
95. Lam PW, Sehgal P, Andany N, Mubareka S, Simor AE, Ozaldin O, et al. A virtual care program for outpatients diagnosed with COVID-19: a feasibility study. CMAJ Open 2020;8(2):E407-E413 [FREE Full text] [doi: 10.9778/cmajo.20200069] [Medline: 32447283]
96. Li H, Chan YC, Huang J, Cheng SW. Pilot study using telemedicine video consultation for vascular patients' care during the COVID-19 period. Ann Vasc Surg 2020 Oct;68:76-82 [FREE Full text] [doi: 10.1016/j.avsg.2020.06.023] [Medline: 32562832]
97. Lin JC, Humphries MD, Shutze WP, Aalami OO, Fischer UM, Hodgson KJ. Telemedicine platforms and their use in the coronavirus disease-19 era to deliver comprehensive vascular care. J Vasc Surg 2021 Feb;73(2):392-398 [FREE Full text] [doi: 10.1016/j.jvs.2020.06.051] [Medline: 32622075]
98. Ma X, Chen Z, Zhu J, Shen X, Wu M, Shi L, et al. Management strategies of neonatal jaundice during the coronavirus disease 2019 outbreak. World J Pediatr 2020 Jun;16(3):247-250 [FREE Full text] [doi: 10.1007/s12519-020-00347-3] [Medline: 32112336]
99. Meloni M, Izzo V, Giurato L, Gandini R, Uccioli L. Management of diabetic persons with foot ulceration during COVID-19 health care emergency: effectiveness of a new triage pathway. Diabetes Res Clin Pract 2020 Jul;165:108245 [FREE Full text] [doi: 10.1016/j.diabres.2020.108245] [Medline: 32497745]
100. Palomba G, Dinuzzi VP, De Palma GD, Aprea G. Management strategies and role of telemedicine in a surgery unit during COVID-19 outbreak. Int J Surg 2020 Jul;79:189-190 [FREE Full text] [doi: 10.1016/j.ijsu.2020.05.081] [Medline: 32479914]
101. Punia V, Nasr G, Zagorski V, Lawrence G, Fesler J, Nair D, et al. Evidence of a rapid shift in outpatient practice during the COVID-19 pandemic using telemedicine. Telemed J E Health 2020 Oct;26(10):1301-1303. [doi: 10.1089/tmj.2020.0150] [Medline: 32429769]
102. Qualliotine JR, Orosco RK. Self-removing passive drain to facilitate postoperative care via telehealth during the COVID-19 pandemic. Head Neck 2020 Jun;42(6):1305-1307 [FREE Full text] [doi: 10.1002/hed.26203] [Medline: 32347997]
103. Naik BN, Gupta R, Singh A, Soni SL, Puri GD. Real-time smart patient monitoring and assessment amid COVID-19 pandemic - an alternative approach to remote monitoring. J Med Syst 2020 Jun 13;44(7):131 [FREE Full text] [doi: 10.1007/s10916-020-01599-2] [Medline: 32533379]

104. Thornton J. The "virtual wards" supporting patients with covid-19 in the community. BMJ 2020 Jun 04;369:m2119. [doi: 10.1136/bmj.m2119] [Medline: 32499317]
105. Umoren RA, Gray MM, Handley S, Johnson N, Kunimura C, Mietzsch U, et al. In-hospital telehealth supports care for neonatal patients in strict isolation. Am J Perinatol 2020 Jun;37(8):857-860 [FREE Full text] [doi: 10.1055/s-0040-1709687] [Medline: 32268382]
106. Martínez-García M, Bal-Alvarado M, Santos Guerra F, Ares-Rico R, Suárez-Gil R, Rodríguez-Álvarez A, et al. Monitoring of COVID-19 patients via telemedicine with telemonitoring. Revista Clínica Española (English Edition) 2020 Nov;220(8):472-479. [doi: 10.1016/j.rceng.2020.07.001]
107. Xu H, Huang S, Qiu C, Liu S, Deng J, Jiao B, et al. Monitoring and management of home-quarantined patients with covid-19 using a WeChat-based telemedicine system: retrospective cohort study. J Med Internet Res 2020 Jul 02;22(7):e19514 [FREE Full text] [doi: 10.2196/19514] [Medline: 32568727]
108. Vilendrer S, Patel B, Chadwick W, Hwa M, Asch S, Pageler N, et al. Rapid deployment of inpatient telemedicine in response to COVID-19 across three health systems. J Am Med Inform Assoc 2020 Jun 04;27(7):1102-1109. [doi: 10.1093/jamia/ocaa077] [Medline: 32495830]
109. Ratwani RM, Brennan D, Sheahan W, Fong A, Adams K, Gordon A, et al. A descriptive analysis of an on-demand telehealth approach for remote COVID-19 patient screening. J Telemed Telecare 2020 Jul 23:1357633X20943339. [doi: 10.1177/1357633X20943339] [Medline: 32698650]
110. McKiever ME, Cleary EM, Schmauder T, Talley A, Hinely KA, Costantine MM, et al. Unintended consequences of the transition to telehealth for pregnancies complicated by opioid use disorder during the coronavirus disease 2019 pandemic. Am J Obstet Gynecol 2020 Nov;223(5):770-772 [FREE Full text] [doi: 10.1016/j.ajog.2020.08.003] [Medline: 32771380]
111. Medalia A, Lynch DA, Herlands T. Telehealth conversion of serious mental illness recovery services during the COVID-19 crisis. Psychiatr Serv 2020 Aug 01;71(8):872. [doi: 10.1176/appi.ps.71705] [Medline: 32741337]
112. Miu AS, Vo HT, Palka JM, Glowacki CR, Robinson RJ. Teletherapy with serious mental illness populations during COVID-19: telehealth conversion and engagement. Couns Psychol Q 2020 Jul 12:1-18. [doi: 10.1080/09515070.2020.1791800]
113. Sequeira A, Alozie A, Fasteau M, Lopez AK, Sy J, Turner KA, et al. Transitioning to virtual programming amidst COVID-19 outbreak. Couns Psychol Q 2020 Jun 12:1-16. [doi: 10.1080/09515070.2020.1777940]
114. Stewart RW, Orengo-Aguayo R, Young J, Wallace MM, Cohen JA, Mannarino AP, et al. Feasibility and effectiveness of a telehealth service delivery model for treating childhood posttraumatic stress: A community-based, open pilot trial of trauma-focused cognitive–behavioral therapy. J Psychother Integr 2020 Jun;30(2):274-289. [doi: 10.1037/int0000225]
115. Sharma A, Sasser T, Schoenfelder Gonzalez E, Vander Stoep A, Myers K. Implementation of home-based telemental health in a large child psychiatry department during the COVID-19 crisis. J Child Adolesc Psychopharmacol 2020 Sep;30(7):404-413. [doi: 10.1089/cap.2020.0062] [Medline: 32639849]
116. Dimer NA, Canto-Soares ND, Santos-Teixeira LD, Goulart BNGD. The COVID-19 pandemic and the implementation of telehealth in speech-language and hearing therapy for patients at home: an experience report. Codas 2020;32(3):e20200144 [FREE Full text] [doi: 10.1590/2317-1782/20192020144] [Medline: 32578694]
117. Reeves JJ, Hollandsworth HM, Torriani FJ, Taplitz R, Abeles S, Tai-Seale M, et al. Rapid response to COVID-19: health informatics support for outbreak management in an academic health system. J Am Med Inform Assoc 2020 Mar 24;27(6):853-859 [FREE Full text] [doi: 10.1093/jamia/ocaa037] [Medline: 32208481]
118. Low TY, Mathews I, Lau JW, Ngiam KY. Close air support: enhancing emergency care in the COVID-19 pandemic. Emerg Med J 2020 Oct;37(10):642-643 [FREE Full text] [doi: 10.1136/emermed-2020-210148] [Medline: 32753393]
119. Martin G, Koizia L, Kooner A, Cafferkey J, Ross C, Purkayastha S, PanSurg Collaborative. Use of the HoloLens2 mixed reality headset for protecting health care workers during the COVID-19 pandemic: prospective, observational evaluation. J Med Internet Res 2020 Aug 14;22(8):e21486 [FREE Full text] [doi: 10.2196/21486] [Medline: 32730222]
120. Nagaratnam K, Harston G, Flossmann E, Canavan C, Geraldes RC, Edwards C. Innovative use of artificial intelligence and digital communication in acute stroke pathway in response to COVID-19. Future Healthc J 2020 Jun;7(2):169-173 [FREE Full text] [doi: 10.7861/fhj.2020-0034] [Medline: 32550287]
121. Shamout FE, Shen Y, Wu N, Kaku A, Park J, Makino T, et al. An artificial intelligence system for predicting the deterioration of COVID-19 patients in the emergency department. NPJ Digit Med 2021 May 12;4(1):80 [FREE Full text] [doi: 10.1038/s41746-021-00453-0] [Medline: 33980980]
122. Gong M, Liu L, Sun X, Yang Y, Wang S, Zhu H. Cloud-based system for effective surveillance and control of COVID-19: useful experiences from Hubei, China. J Med Internet Res 2020 Apr 22;22(4):e18948 [FREE Full text] [doi: 10.2196/18948] [Medline: 32287040]
123. Miseikis J, Caroni P, Duchamp P, Gasser A, Marko R, Miseikiene N, et al. Lio-a personal robot assistant for human-robot interaction and care applications. IEEE Robot Autom Lett 2020 Oct;5(4):5339-5346. [doi: 10.1109/lra.2020.3007462]
124. Rane KP. Design and development of low cost humanoid robot with thermal temperature scanner for COVID-19 virus preliminary identification. Int J Adv Trend Comput Sci Eng 2020 Jun 25;9(3):3485-3493. [doi: 10.30534/ijatcse/2020/153932020]

125. Tabaza L, Virk HUH, Janzer S, George JC. Robotic-assisted percutaneous coronary intervention in a COVID-19 patient. Catheter Cardiovasc Interv 2021 Feb 15;97(3):E343-E345 [FREE Full text] [doi: 10.1002/ccd.28982] [Medline: 32433796]
126. Wang J, Peng C, Zhao Y, Ye R, Hong J, Huang H, et al. Application of a robotic tele-echography system for COVID-19 pneumonia. J Ultrasound Med 2021 Feb;40(2):385-390. [doi: 10.1002/jum.15406] [Medline: 32725833]
127. Ye R, Zhou X, Shao F, Xiong L, Hong J, Huang H, et al. Feasibility of a 5G-based robot-assisted remote ultrasound system for cardiopulmonary assessment of patients with coronavirus disease 2019. Chest 2021 Jan;159(1):270-281 [FREE Full text] [doi: 10.1016/j.chest.2020.06.068] [Medline: 32653568]
128. Yu R, Li Y, Peng C, Ye R, He Q. Role of 5G-powered remote robotic ultrasound during the COVID-19 outbreak: insights from two cases. Eur Rev Med Pharmacol Sci 2020 Jul;24(14):7796-7800 [FREE Full text] [doi: 10.26355/eurrev_202007_22283] [Medline: 32744706]
129. Drew DA, Nguyen LH, Steves CJ, Menni C, Freydin M, Varsavsky T, COPE Consortium. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. Science 2020 Jun 19;368(6497):1362-1367 [FREE Full text] [doi: 10.1126/science.abc0473] [Medline: 32371477]
130. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 Jul;26(7):1037-1040 [FREE Full text] [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]
131. Yamamoto K, Takahashi T, Urasaki M, Nagayasu Y, Shimamoto T, Tateyama Y, et al. Health observation app for COVID-19 symptom tracking integrated with personal health records: proof of concept and practical use study. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e19902 [FREE Full text] [doi: 10.2196/19902] [Medline: 32568728]
132. Schinköthe T, Gabri MR, Mitterer M, Gouveia P, Heinemann V, Harbeck N, et al. A web- and app-based connected care solution for COVID-19 in- and outpatient care: qualitative study and application development. JMIR Public Health Surveill 2020 Jun 01;6(2):e19033 [FREE Full text] [doi: 10.2196/19033] [Medline: 32406855]
133. Dixit RA, Hurst S, Adams KT, Boxley C, Lysen-Hendershot K, Bennett SS, et al. Rapid development of visualization dashboards to enhance situation awareness of COVID-19 telehealth initiatives at a multihospital healthcare system. J Am Med Inform Assoc 2020 Jul 01;27(9):1456-1461 [FREE Full text] [doi: 10.1093/jamia/ocaa161] [Medline: 32618999]
134. Stevens JS, Toma K, Tanzi-Pfeifer S, Rao MK, Mohan S, Gharavi AG, et al. Dashboards to facilitate nephrology disaster planning in the COVID-19 era. Kidney Int Rep 2020 Aug;5(8):1298-1302 [FREE Full text] [doi: 10.1016/j.ekir.2020.06.033] [Medline: 32775829]
135. Ntshalintshali S, Mnqwazi C. Affordable digital innovation to reduce SARS-CoV-2 transmission among healthcare workers. S Afr Med J 2020 Jun 02;110(7):605-606. [doi: 10.7196/SAMJ.2020.v110i7.14879] [Medline: 32880330]
136. Lai L, Sato R, He S, Ouchi K, Leiter R, deLima Thomas J, et al. Usage patterns of a web-based palliative care content platform (PalliCOVID) during the COVID-19 pandemic. J Pain Symptom Manage 2020 Oct;60(4):e20-e27 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.07.016] [Medline: 32730951]
137. Li CH, Rajamohan AG, Acharya PT, Liu CJ, Patel V, Go JL, et al. Virtual read-out: radiology education for the 21st century during the COVID-19 pandemic. Acad Radiol 2020 Jun;27(6):872-881 [FREE Full text] [doi: 10.1016/j.acra.2020.04.028] [Medline: 32386950]
138. Li J, Wang X, Huang X, Chen F, Zhang X, Liu Y, et al. Application of CareDose 4D combined with Karl 3D technology in the low dose computed tomography for the follow-up of COVID-19. BMC Med Imaging 2020 May 24;20(1):56 [FREE Full text] [doi: 10.1186/s12880-020-00456-5] [Medline: 32448136]
139. Madhavan S, Bastarache L, Brown J, Butte A, Dorr D, Embi P, et al. Use of electronic health records to support a public health response to the COVID-19 pandemic in the United States: a perspective from 15 academic medical centers. J Am Med Inform Assoc 2021 Feb 15;28(2):393-401 [FREE Full text] [doi: 10.1093/jamia/ocaa287] [Medline: 33260207]
140. Tenforde AS, Borgstrom H, Polich G, Steere H, Davis IS, Cotton K, et al. Outpatient physical, occupational, and speech therapy synchronous telemedicine: a survey study of patient satisfaction with virtual visits during the COVID-19 pandemic. Am J Phys Med Rehabil 2020 Nov;99(11):977-981 [FREE Full text] [doi: 10.1097/PHM.0000000000001571] [Medline: 32804713]
141. Thomas I, Siew LQC, Rutkowski K. Synchronous telemedicine in allergy: lessons learned and transformation of care during the COVID-19 pandemic. J Allergy Clin Immunol Pract 2021 Jan;9(1):170-176.e1 [FREE Full text] [doi: 10.1016/j.jaip.2020.10.013] [Medline: 33091636]
142. Silver L. Smartphone ownership is growing rapidly around the world, but not always equally. Pew Research Center. 2019 May 01. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2001-04-21]
143. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851. [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
144. Venkatesh, Thong, Xu. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012 Sep 01;36(1):157-141. [doi: 10.2307/41410412]
145. Huang F, Blaschke S, Lucas H. Beyond pilotitis: taking digital health interventions to the national level in China and Uganda. Global Health 2017 Jul 31;13(1):49 [FREE Full text] [doi: 10.1186/s12992-017-0275-z] [Medline: 28756767]

146. Labrique A, Vasudevan L, Weiss W, Wilson K. Establishing standards to evaluate the impact of integrating digital health into health systems. Glob Health Sci Pract 2018 Oct 10;6(Supplement 1):S5-S17. [doi: 10.9745/ghsp-d-18-00230]
147. Jandoo T. WHO guidance for digital health: what it means for researchers. Digit Health 2020;6:2055207619898984 [FREE Full text] [doi: 10.1177/2055207619898984] [Medline: 31949918]
148. Katzow MW, Steinway C, Jan S. Telemedicine and Health Disparities During COVID-19. Pediatrics 2020 Aug;146(2):e20201586 [FREE Full text] [doi: 10.1542/peds.2020-1586] [Medline: 32747592]
149. Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc 2018 Aug 01;25(8):1080-1088. [doi: 10.1093/jamia/ocy052] [Medline: 29788380]
150. Davies AR, Honeyman M, Gann B. Addressing the Digital Inverse Care Law in the Time of COVID-19: Potential for Digital Technology to Exacerbate or Mitigate Health Inequalities. J Med Internet Res 2021 Apr 07;23(4):e21726 [FREE Full text] [doi: 10.2196/21726] [Medline: 33735096]
151. Kichloo A, Albosta M, Dettloff K, Wani F, El-Amir Z, Singh J, et al. Telemedicine, the current COVID-19 pandemic and the future: a narrative review and perspectives moving forward in the USA. Fam Med Community Health 2020 Aug;8(3):e000530 [FREE Full text] [doi: 10.1136/fmch-2020-000530] [Medline: 32816942]
152. Gunasekeran DV, Tham Y, Ting DSW, Tan GSW, Wong TY. Digital health during COVID-19: lessons from operationalising new models of care in ophthalmology. Lancet Digit Health 2021 Feb;3(2):e124-e134 [FREE Full text] [doi: 10.1016/S2589-7500(20)30287-9] [Medline: 33509383]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses

Edited by C Basch; submitted 02.04.21; peer-reviewed by M Alajlani, K Denecke, E Al-Jafar, K Fadahunsi; comments to author 19.04.21; revised version received 28.04.21; accepted 20.06.21; published 14.09.21.

Please cite as:
Abd-Alrazaq A, Hassan A, Abuelezz I, Ahmed A, Alzubaidi MS, Shah U, Alhuwail D, Giannicchi A, Househ M
Overview of Technologies Implemented During the First Wave of the COVID-19 Pandemic: Scoping Review
J Med Internet Res 2021;23(9):e29136
URL: https://www.jmir.org/2021/9/e29136
doi:10.2196/29136
PMID:34406962

Original Paper

Characterization and Identification of Variations in Types of Primary Care Visits Before and During the COVID-19 Pandemic in Catalonia: Big Data Analysis Study

Francesc Lopez Segui[1], PhD; Guillem Hernandez Guillamet[1,2], MSc; Héctor Pifarré Arolas[1], PhD; Francesc X Marin-Gomez[3,4,5], MD, PhD; Anna Ruiz Comellas[3,5,6], MD, PhD; Anna Maria Ramirez Morros[2,4], RNU, MPH; Cristina Adroher Mas[7], MSc; Josep Vidal-Alaball[3,4,5], MD, PhD

[1]Centre de Recerca en Economia i Salut, Pompeu Fabra University, Barcelona, Spain

[2]Gerència Territorial de la Catalunya Central, Institut Català de la Salut, Sant Fruitós de Bages, Spain

[3]Health Promotion in Rural Areas Research Group, Gerència Territorial de la Catalunya Central, Institut Català de la Salut, Sant Fruitós de Bages, Spain

[4]Unitat de Suport a la Recerca de la Catalunya Central, Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina, Sant Fruitós de Bages, Spain

[5]Faculty of Medicine, University of Vic, Central University of Catalonia, Vic, Spain

[6]Centre d'Atenció Primària Sant Joan de Vilatorrada, Gerència Territorial de la Catalunya Central, Institut Català de la Salut, Sant Fruitós de Bages, Spain

[7]Sant Joan de Déu Hospital, Barcelona, Spain

Corresponding Author:
Josep Vidal-Alaball, MD, PhD
Health Promotion in Rural Areas Research Group
Gerència Territorial de la Catalunya Central, Institut Català de la Salut
Pica d'Estats 36
Sant Fruitós de Bages
Spain
Phone: 34 93 693 0040
Email: jvidal.cc.ics@gencat.cat

Abstract

Background: The COVID-19 pandemic has turned the care model of health systems around the world upside down, causing the abrupt cancellation of face-to-face visits and redirection of the model toward telemedicine. Digital transformation boosts information systems—the more robust they are, the easier it is to monitor the health care system in a highly complex state and allow for more agile and reliable analysis.

Objective: The purpose of this study was to analyze diagnoses from primary care visits and distinguish between those that had higher and lower variations, relative to the 2019 and 2020 periods (roughly before and during COVID-19), to identify clinical profiles that may have been most impaired from the least-used diagnostic codes for visits during the pandemic.

Methods: We used a database from the Primary Care Services Information Technologies Information System of Catalonia. We analyzed the register of visits (n=2,824,185) and their *International Classification of Diseases* (*ICD-10*) diagnostic codes (n=3,921,974; mean 1.38 per visit), as approximations of the reasons for consultations, at 3 different grouping levels. The data were represented by a term frequency matrix and analyzed recursively in different partitions aggregated according to date.

Results: The increase in non–face-to-face visits (+267%) did not counterbalance the decrease in face-to-face visits (–47%), with an overall reduction in the total number of visits of 1.36%, despite the notable increase in nursing visits (10.54%). The largest increases in 2020 were visits with diagnoses related to COVID-19 (*ICD-10* codes Z20-Z29: 2.540%), along with codes related to economic and housing problems (*ICD-10* codes Z55-Z65: 44.40%). Visits with most of the other diagnostic codes decreased in 2020 relative to those in 2019. The largest reductions were chronic pathologies such as arterial hypertension (*ICD-10* codes I10-I16: –32.73%) or diabetes (*ICD-10* codes E08-E13: –21.13%), but also obesity (E65-E68: –48.58%) and bodily injuries (*ICD-10* code T14: –33.70%). Visits with mental health–related diagnostic codes decreased, but the decrease was less than the average decrease. There was a decrease in consultations—for children, adolescents, and adults—for respiratory infections (*ICD-10* codes J00-J06: –40.96%). The results show large year-on-year variations (in absolute terms, an average of 12%), which is representative of the strong shock to the health system.

Conclusions: The disruption in the primary care model in Catalonia has led to an explosive increase in the number of non–face-to-face visits. There has been a reduction in the number of visits for diagnoses related to chronic pathologies, respiratory infections, obesity, and bodily injuries. Instead, visits for diagnoses related to socioeconomic and housing problems have increased, which emphasizes the importance of social determinants of health in the context of this pandemic. Big data analytics with routine care data yield findings that are consistent with those derived from intuition in everyday clinical practice and can help inform decision making by health planners in order to use the next few years to focus on the least-treated diseases during the COVID-19 pandemic.

(J Med Internet Res 2021;23(9):e29622) doi:10.2196/29622

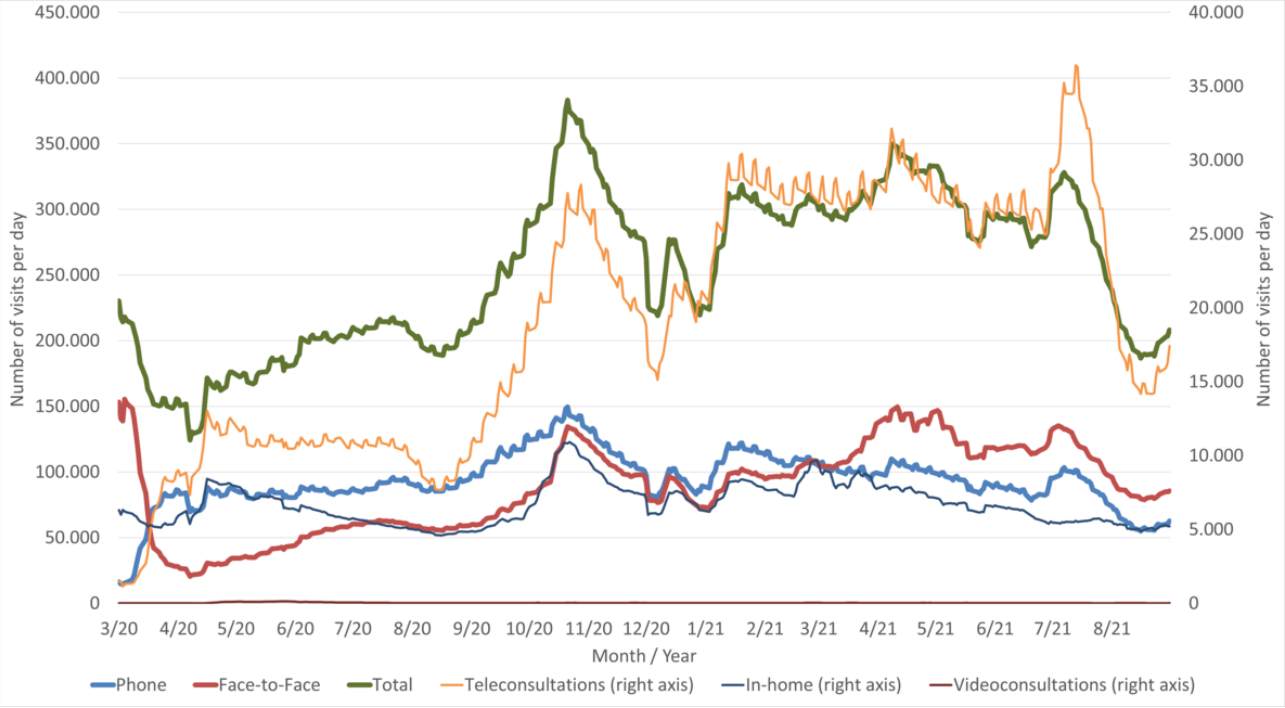
KEYWORDS

COVID-19; primary care; diagnose variations; big data; ICD10; health system; big data; primary care; healthcare system

Introduction

The COVID-19 pandemic has turned the care model of health systems around the world upside down, with the abrupt cancellation of face-to-face visits, redirection of the model toward non–face-to-face care, and then the gradual recovery of services, first for priority population sectors and later for the general population [1-4]. Telemedicine in primary care in Catalonia includes provider–patient or provider–provider communications and can take place synchronously (telephone or video calls) or asynchronously (teleconsultations). By the end of February 2021, all face-to-face activity had not yet recovered, although the total number of visits was approximately 15% higher than that at the start of the pandemic (Figure 1), in part because of the increasing use of telemedicine (mostly in the form of telephone consultations).

Figure 1. A 7-day rolling average (with weekends and holidays not included) of primary care daily visits in Catalonia, by type, from March 2020 to August 2021 (unpublished data, Catalan Ministry of Health).



Because of the disruptions, health systems have begun to identify population profiles that may be most adversely affected by changes and underdiagnosed diseases during the pandemic [5-8]. Some patients have had to wait longer than before to receive health care. It has been observed, for example, that cardiac arrests outside hospital have increased, with patients having worse prognoses [9], or that patients with heart failure have had fewer admissions, but when admitted, heart failure has been more severe and admission to the intensive care unit is required more often [10]. It has also been noted that access to telemedicine visits has not been the same for the whole population and that, when this type of care becomes predominant, inequalities in access to care may arise [11,12].

According to a recent study [13], COVID-19 has resulted in loss of years of life over 7 times greater than that in a typical influenza season in Spain. The health emergency has precipitated policy maker decision-making that has often been without any existing evidence and with long delays [14], but it has also highlighted the value of information systems and, more generally, the role of health policy evaluation agencies [15]. Robust information systems make it easier to assess and monitor the health system in a highly complex state and allow for agile and reliable analysis. In this context, we explored the potential for the use of a massive amount of primary care data from the Catalan public health system to inform quick, granular decision making regarding which areas have been most neglected as a

result of the change in the care model. Specifically, we aimed to identify which diagnoses in primary care have increased and decreased from 2019 to 2020 (approximately before and during the COVID-19 pandemic), in order to determine which areas need to be strengthened in view of the gradual recovery of health system activity.

Methods

Ethics

The study was conducted according to the guidelines of the Declaration of Helsinki and in accordance with European General Data Protection Regulations and the Spanish Organic Law on Data Protection and Guarantee of Digital Rights. No ethical approval was required; analyses were conducted only on aggregated data.

Data

The Catalan health system provides universal coverage to 7.6 million inhabitants. Administratively, it is composed of a single public payer and several publicly or privately owned service providers, in an integrated system and with an important role of community and primary health care and the use of information technology. We used a database from the Primary Care Services Information Technologies System of the Health Region of Central Catalonia (Catalonia, Spain) belonging to the Catalan Institute of Health. In this region, the Health Information System is used by the majority (32/37, 86.48%) of Primary Care centers, which facilitated analysis. The variable of interest was primary care visits according to the *International Statistical Classification of Diseases tenth revision* (*ICD-10*) [16] diagnostic codes. These codes were used as a proxy for the reason for the visit.

Data from primary care visits in 2020 (the period covering most of the COVID-19 pandemic) and 2019 (the period used as a comparator)—both face-to-face (at the primary care center or at home) and non–face-to-face (telemedicine tools including telephone and teleconsultations)—were analyzed. One visit could result in several diagnostic codes and not all visits were associated with at least one, as doing so is not mandatory in the Catalan primary care electronic health record. Consequently, the data set was reduced to visits in which at least one diagnosis had been identified.

Statistical Analysis

Diagnostic codes attributed to visits were analyzed at 3 levels of aggregation: aggregation level 1 (21 codes, with format "Neoplasms, C00-D49"), which corresponds to the concept of *chapters* according to the reference *ICD-10* dictionary [16]; aggregation level 2, which was a second, more specific grouping (284 different codes, formatted "Malignant neoplasms of digestive organs, C15-C26"), which corresponded to the concept of *blocks* in the above-mentioned dictionary; and aggregation level 3, which was a third grouping with even more specificity (1532 different codes, with the format "Malignant neoplasm of anus and anal canal, C21").

The variety of diagnostic codes used in clinical practice suggested that there would be a need for the use of big data analysis techniques. Visit data were recursively analyzed in 4 different partitions, and the diagnoses were aggregated according to date. *ICD-10* groupings were represented with a term frequency matrix [17,18], where matrix columns indicated the absolute number of each diagnostic code.

The variation index was used to represent the relative weight of the variation of a diagnostic code over the absolute total interannual variation of all codes. Thus, in addition to the frequency of use of each diagnostic code, this index takes into account the importance of each on a linear scale to allow visualization of the diagnoses on 2 scales (code frequency and variation index). Data were analyzed using R (version 3.4.3; R Foundation for Statistical Computing).

Results

There were 3,555,799 primary care visits for 376,486 citizens (a total of 404,245 citizens are in the reference population of this health region) in 2019 and 2020, and at least one diagnosis was identified in 79.43% of visits (2824371/3,555,799). The degree of coding is slightly higher in the face-to-face visits compared to the non-face-to-face visits Regarding the effects of the pandemic on missing data, during 2019, 18.46% (262,460/1,421,779) of visits had no associated diagnoses, while in 2020, missing instances reached 22.50% (315,541/1,402,406), representing a limited impact on reported outcomes. The final database was composed of 2,824,185 visits for 358,419 different citizens, with 3,921,974 diagnostic visit codes (1.38 codes per visit on average; 3.5% of visits had more than 3 associated codes; Figure 2 and Figure 3).

Figure 2. Aggregation level 1 cumulative frequency of the 21 most used *International Statistical Classification of Diseases, tenth revision*, diagnostic codes (2019 and 2020). behav: behavioral, endoc: endocrine, malf: malformation, metabol: metabolic, nutrit: nutritional.

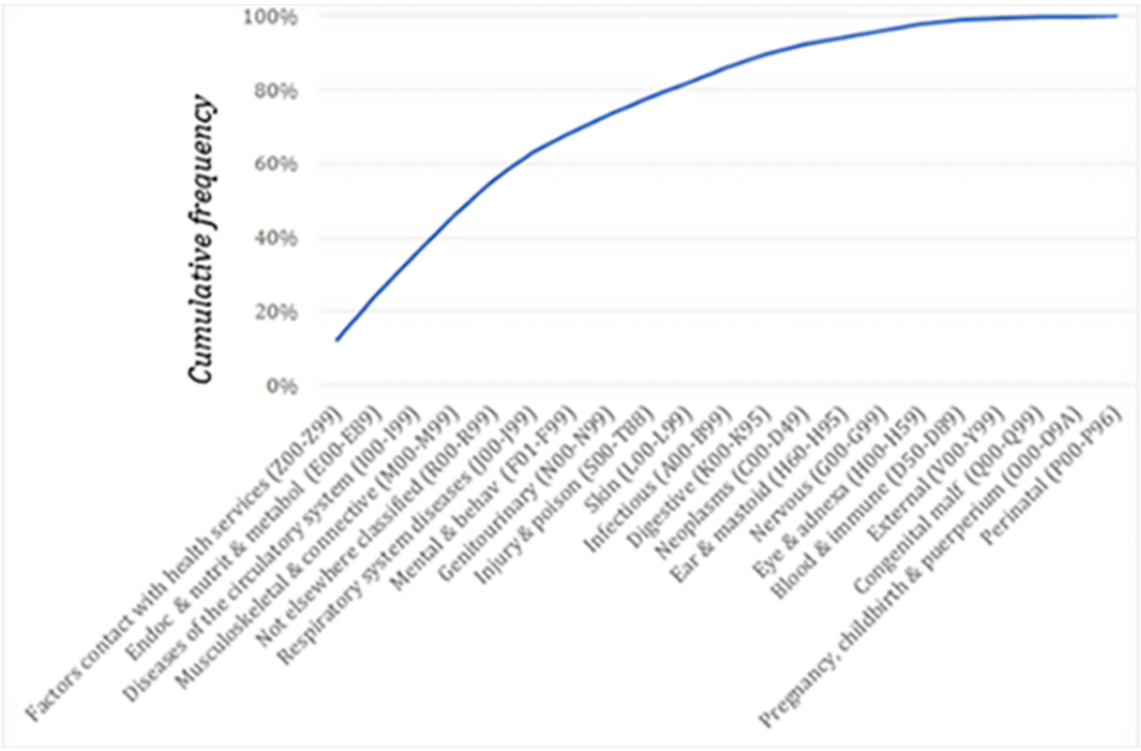
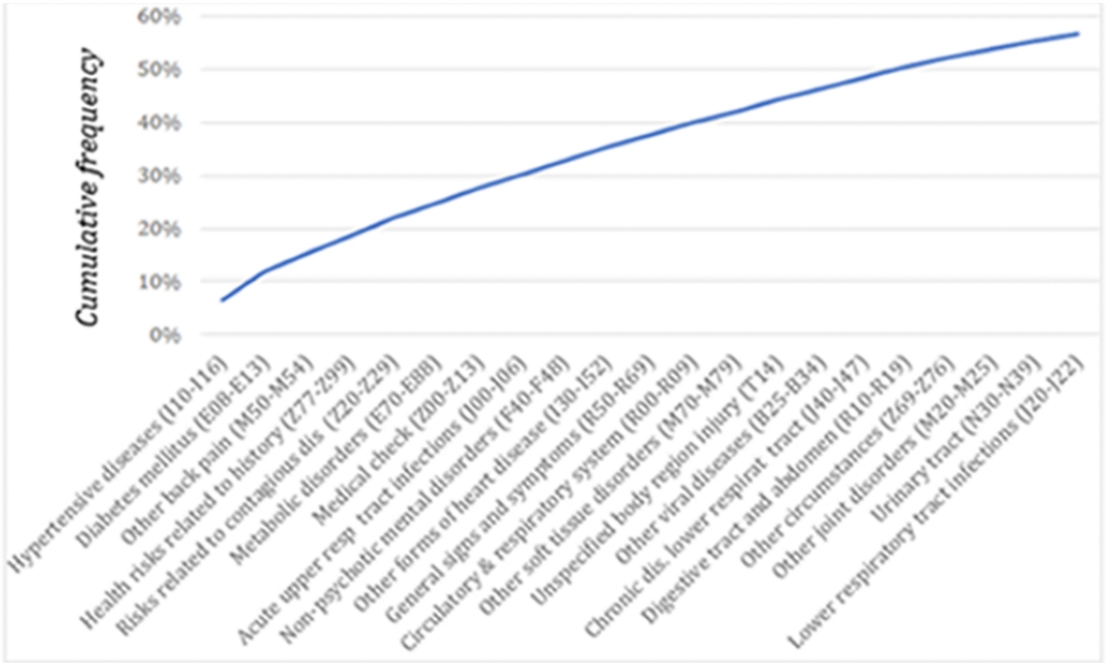


Figure 3. Aggregation level 2 cumulative frequency of the 21 most used *International Statistical Classification of Diseases, tenth revision*, diagnostic codes (2019 and 2020). dis: disease, resp/respirat: respiratory.



The total number of visits in 2020 decreased slightly (–1.36%) compared to those in 2019; if COVID-19–related visits are excluded, this figure is even lower (–7.56%). The type of visit changed markedly (Table 1). In 2019, most visits (1,212,800/1,421,779, 85.30%) were face-to-face, while in 2020 more than half of visits (768,548/1,402,406, 54.80%) were telemedicine. The increase in this type of visit, therefore, has not been able to compensate for the decrease in face-to-face visits.

Table 1. Evolution of visits.

Characteristic	2019	2020	Difference, %
Total patients attended	287,936	287,965	0.01
Total visits, n	1,421,779	1,402,406	–1.36
Face-to-face, n (%)	1,212,800 (85.30)	633,858 (45.20)	–47.74
Telemedicine, n (%)	208,979 (14.70)	768,548 (54.80)	267.76
General medicine, n (%)	859,153 (60.43)	822,433 (58.64)	–4.27
Nursing, n (%)	427,959 (30.10)	473,051 (33.73)	10.54
Pediatrics, n (%)	134,667 (9.47)	106,922 (7.62)	–20.60
Early childhood[a], n (%)	95,098 (6.69)	76,460 (5.45)	–24.38
Childhood[b], n (%)	57,436 (4.04)	52,101 (3.72)	–10.24
Adolescence[c], n (%)	55,768 (3.92)	61,693 (4.40)	9.60
Youth[d], n (%)	55,943 (3.93)	69,287 (4.94)	19.26
Adulthood[e], n (%)	475,366 (33.43)	514,327 (36.67)	7.58
Older adult[f], n (%)	682,168 (47.98)	628,538 (44.82)	–8.53
Total diagnostic codes, n	2,079,680	1,912,342	–10.57
Diagnostic codes per visit, mean (SD)	1.46 (1.00)	1.36 (0.89)	–6.84
Sex, n (%)			
Male	641,602 (45.13)	627,806(44.77)	–2.15
Female	780,074 (54.87)	774,486 (55.23)	–0.72
Age of visits, mean (SD)	53.54 (26.25)	52.89 (25.72)	–1.21

[a]Early childhood: 0 to 5 years old.

[b]Childhood: infancy to 11 years old.

[c]Adolescence: 12 to 18 years old.

[d]Youth: 14-26 years old.

[e]Adulthood: 27 to 59 years old.

[f]Older adult: over 60 years old.

In relation to the specialty, the number of visits with doctors decreased slightly (–4.27%), the number of visits with nurses increased (10.54%), and the number of pediatric visits decreased (–20.60%). Despite this decrease, the mean age of patients decreased by approximately 1 year (from 53.54 in 2019 to 52.89 years in 2020). There was a variation in visits according to age groups: there was a decrease in visits for patients in early childhood and infancy (from 0 to 5 years old), an increase in patients in adolescence (12 to 18 years old) to adulthood (27 to 59 years old), and a decrease in older adult patients. The proportion of visits made by women was 55.04% (1,554,560/2,824,185); the magnitude remained fairly unchanged between 2019 and 2020.

The cumulative frequency of the 21 most used diagnostic codes for the 2 first levels of grouping were analyzed (Figure 3). In the case of aggregation level 1, the 5 code groupings comprised more than half of the total number of codes (Factors influencing health status and contact with health services; Endocrine, nutritional, and metabolic diseases; Diseases of the circulatory system; Diseases of the musculoskeletal system and connective tissue; Signs, symptoms, and abnormal test results not classified elsewhere). For aggregation level 2, the 5 most-used codes comprised approximately one-quarter of the total (Hypertensive diseases; Diabetes mellitus; Other back disorders; Persons with potential health risks related to personal and family history and certain conditions influencing health status; Persons with potential health risks related to communicable diseases).

For aggregation level 1, large interannual variations were observed (Table 2), especially for diagnostic codes related to *factors influencing health status and contact with health services* (*ICD-10* codes Z00-Z99), codes that were initially used for diagnoses related to COVID-19, and the diagnosis of *certain infectious and parasitic diseases*, which nearly doubled in 2020. The rest of the codes decreased in 2020 compared to 2019; however, those related to mental health decreased less than the rest (*ICD-10* codes F01-F99: –3.39%). Diagnostic codes related to injuries were reduced by almost one-quarter (*ICD-10* codes S00-T88: –26.85%), possibly due to a drastic reduction in accidents resulting from people engaging in less activity during confinement. The variation index shows the weight of each of these variations on overall disruption (Table 3).

Table 2. Use of *International Statistical Classification of Diseases, tenth revision*, diagnostic codes (aggregation level 1).

Diagnostic code	Total, n (%)	2019, n (%)	2020, n (%)	Interannual variation, %	Variation index
Factors influencing health status and contact with health services (Z00-Z99)	483,368 (12.32)	186,604 (9.06)	296,764 (15.93)	59.03	0.22
Endocrine, nutritional, and metabolic diseases (E00-E89)	464,009 (11.83)	261,076 (12.68)	202,933 (10.90)	–22.27	0.11
Diseases of the circulatory system (I00-I99)	440,256 (11.23)	252,555 (12.26)	187,701 (10.08)	–25.68	0.13
Diseases of the musculoskeletal system and connective tissue (M00-M99)	405,813 (10.35)	224,901 (10.92)	180,912 (9.71)	–19.56	0.09
Signs, symptoms, and abnormal test results not elsewhere classified (R00-R99)	385,121 (9.82)	200,711 (9.75)	184,410 (9.90)	–8.12	0.03
Diseases of the respiratory system (J00-J99)	297,074 (7.57)	176,405 (8.57)	120,669 (6.48)	–31.60	0.11
Mental, behavioral, and neurodevelopmental disorders (F01-F99)	214,709 (5.47)	109,203 (5.30)	105,506 (5.66)	–3.39	0.01
Diseases of the genitourinary system (N00-N99)	182,642 (4.66)	96,692 (4.70)	85,950 (4.61)	–11.11	0.02
Injuries, poisonings, and other consequences of external causes (S00-T88)	182,045 (4.64)	105,135 (5.11)	76,910 (4.13)	–26.85	0.06
Diseases of the skin and subcutaneous tissue (L00-L99)	164,759 (4.20)	89,119 (4.33)	75,640 (4.06)	–15.12	0.03
Certain infectious and parasitic diseases (A00-B99)	151,999 (3.88)	52,220 (2.54)	99,779 (5.36)	91.07	0.09
Diseases of the digestive system (K00-K95)	143,793 (3.67)	78,968 (3.83)	64,825 (3.48)	–17.91	0.03
Neoplasms (C00-D49)	102,910 (2.62)	55,201 (2.68)	47,709 (2.56)	–13.57	0.01
Diseases of the ear and mastoid process (H60-H95)	81,495 (2.08)	47,699 (2.32)	33,796 (1.81)	–29.15	0.03
Diseases of the nervous system (G00-G99)	72,150 (1.84)	38,057 (1.85)	34,093 (1.83)	–10.42	0.01
Diseases of the eye and adnexa (H00-H59)	61,810 (1.58)	36,837 (1.79)	24,973 (1.34)	–32.21	0.02
Diseases of the blood and blood-forming organs and disorders affecting the immune mechanism (D50-D89)	45,551 (1.16)	25,305 (1.23)	20,246 (1.09)	–19.99	0.01
External causes of morbidity (V00-Y99)	23,659 (0.60)	12,481 (0.61)	11,178 (0.60)	–10.44	0.00
Congenital malformations. congenital deformities, and congenital chromosomal anomalies (Q00-Q99)	10,034 (0.26)	5493 (0.27)	4541 (0.24)	–17.33	0.00
Pregnancy. childbirth, and puerperium (O00-O9A)	7023 (0.18)	3788 (0.18)	3235 (0.17)	–14.60	0.00
Certain conditions originating in the perinatal period (P00-P96)	1754 (0.04)	991 (0.05)	763 (0.04)	–23.01	0.00

Table 3. Top 10 *International Statistical Classification of Diseases, tenth revision*, diagnostic codes according to the variation index.

Diagnostic code	Variation index
Aggregation level 2	
Persons with potential health risks related to communicable diseases (Z20-Z29)	0.20
Other viral diseases (B25-B34)	0.10
Hypertensive diseases (I10-I16)	0.09
Acute upper respiratory infections (J00-J06)	0.05
Diabetes mellitus (E08-E13)	0.04
Metabolic disorders (E70-E88)	0.03
Injury to unspecified body region (T14)	0.03
Other acute lower respiratory tract infections (J20-J22)	0.03
Other soft tissue disorders (M70-M79)	0.03
Other back disorders (M50-M54)	0.02
Aggregation level 3	
Contact and exposure (suspected): communicable diseases (Z20)	0.17
Viral infection of unspecified location (B34)	0.08
Essential (primary) hypertension (I10)	0.07
COVID (U07)	0.05
Diabetes mellitus type 2 (E11)	0.03
Injury to unspecified body region (T14)	0.03
Disorders of lipoprotein metabolism and other hyperlipidemias (E78)	0.02
Acute rhinopharyngitis (J00)	0.02
Dorsalgia (M54)	0.02
Overweight and obesity (E66)	0.02

The number of diagnostic codes decreased slightly (Figure 4), and those for telemedicine and face-to-face visits increased and decreased, respectively. Aggregation level 2 analysis shows that the diagnostic codes with the largest increase in 2020 were those related to COVID-19 (*ICD-10* codes Z20-Z29: 2.540%), along with codes related to economic and housing problems (*ICD-10* codes Z55-Z65: 44.40%). Visits with most of the other diagnostic codes decreased in 2020 relative to those in 2019. Visits for chronic pathologies such as arterial hypertension (*ICD-10* codes I10-I16: –32.73%) or diabetes (*ICD-10* codes E08-E13: –21.13%), but also those for obesity (*ICD-10* codes E65-E68: –48.58%) and bodily injuries (T14: –33.70%) were reduced. The results show year-on-year variations (in absolute terms, an average of 12%).

The frequencies of diagnoses (listed according to the highest variation index) in face-to-face and non–face-to-face visits were systematically negative (Table 4) and positive (Table 5), respectively. It should also be noted that none of the face-to-face visit code groups were offset by those of the non–face-to-face group (ie, the largest reductions in the former were not offset by the largest increases in the latter).

Figure 4. Primary care consultations before (in 2019) vs during (in 2020) COVID-19 (absolute frequency, log scale) according to *International Statistical Classification of Diseases, tenth revision*, codes. Each point corresponds to an *International Statistical Classification of Diseases, tenth revision*, diagnostic code. The 45° line indicates the identical frequency of diagnostic code use between periods; points above indicate the set of diagnostic codes that have been visited relatively more in 2020 and vice versa. The color of the dot indicates the value of the variation index (the higher, the more important the year-on-year variation).

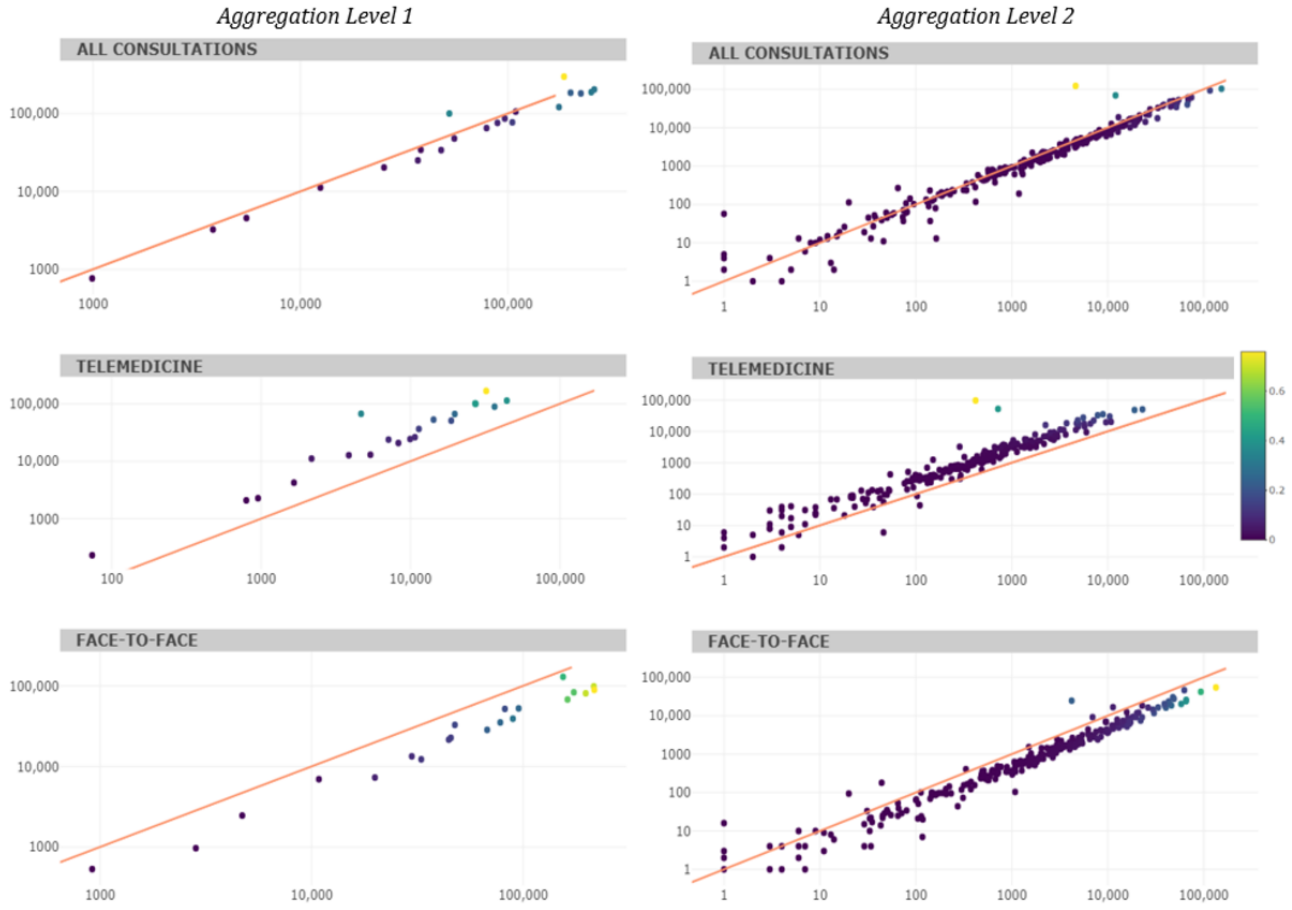


Table 4. Top 5 in variation index of diagnostic use in face-to-face visits (aggregation level 2).

ICD-10[a] diagnostic code	Total, n (%)	2019, n (%)	2020, n (%)	Interannual variation, %	Variation index
Diseases of the blood and blood-forming organs and disorders of the immune mechanism (D50-D89)	27,197 (1.03)	19,912 (1.12)	7285 (0.83)	–63.41	0.14
Mental, behavioral, and neurodevelopmental disorders (F01-F99)	128,611 (4.85)	89,415 (5.05)	39,196 (4.47)	–56.16	0.13
Diseases of the ear and mastoid process (H60-H95)	68,215 (2.57)	45,515 (2.57)	22,700 (2.59)	–50.13	0.13
Diseases of the digestive system (K00-K95)	96,028 (3.62)	67,589 (3.81)	28,439 (3.24)	–57.92	0.11
Congenital malformations, congenital deformities, and congenital chromosomal anomalies (Q00-Q99)	7107 (0.27)	4668 (0.26)	2439 (0.28)	–47.75	0.10

[a]*ICD-10*: *International Statistical Classification of Diseases, tenth revision*

Table 5. Top 5 in variation index of diagnostic use in non–face-to-face visits (aggregation level 1).

ICD-10[a] diagnostic code	Total, n (%)	2019, n (%)	2020, n (%)	Interannual variation, %	Variation index
Factors influencing health status and contact with health services (Z00-Z99)	199,464 (15.68)	32,146 (11.20)	167,318 (16.99)	420.49	0.19
Signs, symptoms, and abnormal test results not elsewhere classified (R00-R99)	128,214 (10.08)	27,092 (9.44)	101,122 (10.27)	273.25	0.11
Diseases of the musculoskeletal system and connective tissue (M00-M99)	127,293 (10.01)	27,328 (9.52)	99,965 (10.15)	265.80	0.10
Endocrine, nutritional, and metabolic diseases (E00-E89)	157,663 (12.40)	44,155 (15.38)	113,508 (11.53)	157.07	0.10
Certain infectious and parasitic diseases (A00-B99)	71,620 (5.63)	4670 (1.63)	167,318 (16.99)	1333.62	0.09

[a]*ICD-10*: *International Statistical Classification of Diseases, tenth revision*

Finally, for children, adolescents, and adults, there were decreases in consultations for respiratory infections; ear, nose and throat pathologies; and skin problems. In adults and older adults, there were overall decreases in visits for chronic pathologies, especially those for arterial hypertension and diabetes. Decreases in visits for some pathologies were counterbalanced by an increase in telemedicine visits—among adults, for back disorders, soft tissue disorders, upper respiratory tract infections, and anxiety, and among older adults, different types of neoplasms. By sex, no differences were observed in the frequency of diagnoses (Multimedia Appendices 1-3).

Discussion

Principal Findings

We analyzed the differences in distinctive characteristics of care before and after the outbreak of the COVID-19 pandemic in the context of primary care in Catalonia. As expected, the increase in the frequency of visits with diagnostic codes related to *contact with or suspected exposure to SARS-CoV-2* stands out, but those related to *housing, employment and psychosocial factors* (*ICD-10* codes Z55-Z65) are also noteworthy. On the eve of a foreseeable economic crisis, this fact underlines the importance of social determinants of health and population health management when planning reconstruction policy for the period after the pandemic, as it is well known that vulnerable population groups were the worst affected by the economic crisis [19]. The change in the pattern of visits is also visible in 2 other aspects. First, the number of visits in the nursing service increased (10.54%), while those in general medicine visits decreased (–4.27%). Second, the increasing use of non–face-to-face care tools since the start of the pandemic has served a wider and more representative cross-section of the population (traditionally underrepresented groups, including low-income individuals and patients residing in the poorest areas of the country, ie, rural areas) [12]. For this reason, since the outbreak of COVID-19, and due to the role of primary care in the management of sick leave, there seems to have been a reduction in the clinical complexity of patients seen, while patients with socioeconomic problems (eg, unemployment) have increased. Visits associated with diagnostic codes for chronic pathologies (mainly endocrine and cardiovascular diseases) have been the most affected, with a decrease in frequency that has not been counterbalanced by the increase in telemedicine.

Opportunities abound for quantifying the impact of the pandemic, in addition to analyses presented herein, on service delivery and citizens' health in order to anticipate and plan for changes in demand for health services in the coming months or years. It would be necessary to study whether, due to the lack of testing (eg, routine analysis), prescriptions performed by clinicians have been more pharmacological; which visits that were previously made in person are now conducted using telemedicine; to what extent the reduced intensity of care for chronic disease management has affected the health of citizens [6]; and whether mental health (note that the amount of this type of visit have been reduced less than the rest) will be the next pandemic [20] in order to reprioritize a care model in Catalonia which, in the opinion of the authors, was already suffering a lot of health care pressure before COVID-19.

Despite the crisis caused by the pandemic, the changes that have been made to address the crisis may improve future health systems. Big data analysis tools allow exhaustive, systematic, agile evaluation. Given the preeminence and role of primary care as the gateway to the public health system in Catalonia, the information derived from primary care information registers is very appropriate for monitoring variations in the provision of health services. This is demonstrated by the fact that the results of this study quantify reductions in the intensity of some types of visits that are consistent with our expectations. Future avenues of research are the sensitivity of information system records in predicting the severity of the situation.

Limitations

It is worth mentioning that the 2 samples that were compared were not perfectly symmetric, insofar as the start of the COVID-19 outbreak did not coincide exactly with end and beginning of calendar years. Thus, 2020 data reflect, in part, some weeks or months of normal health care, showing some bias in favor of non–COVID-19 visits. In addition, we used *ICD-10* diagnostic codes as a proxy for the reason for visits; though it is a useful variable due to its systematic collection, it is not perfect. Moreover, while it is true that in hospitals there is a fairly exhaustive control in the coding of these variables due to the implications for billing, this is not the case in primary care, so these data may be less well coded. The circumstances

caused by the pandemic may also have affected the overall quality of the data (eg, due to newly established teleworking tools among health professionals).

Conclusions

The disruption in the primary care model in Catalonia has led to an explosive increase in the number of non–face-to-face visits. There has been a reduction in the number of visits for diagnoses related to chronic pathologies, obesity, bodily injuries, and respiratory infections, possibly as an indirect effect of COVID-19 prevention measures. Instead, diagnostic visits for socioeconomic and housing problems have increased, which emphasizes the importance of social determinants of health in a pandemic context.

Big data analysis can help to inform health planners in their decision making to focus on the least treated diseases during the pandemic. The availability of information enables comprehensive and reliable monitoring of the health system to be undertaken. In extraordinary situations, such as those created by the COVID-19 pandemic, and in view of the necessary emergency plan that will have to be carried out to compensate for the onslaught of clinical activity caused by COVID-19, big data analysis is even more relevant.

Conflicts of Interest

None declared.

Multimedia Appendix 1
HTML (graphs).
[TXT File , 0 KB - jmir_v23i9e29622_app1.txt]

Multimedia Appendix 2
EXE file with interactive plots.
[TXT File , 0 KB - jmir_v23i9e29622_app2.txt]

Multimedia Appendix 3
Variation Index formula.
[PNG File , 90 KB - jmir_v23i9e29622_app3.png]

References

1. Wosik J, Fudim M, Cameron B, Gellad ZF, Cho A, Phinney D, et al. Telehealth transformation: COVID-19 and the rise of virtual care. J Am Med Inform Assoc 2020 Apr 20:1 [FREE Full text] [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]
2. Clarke G, Pariza P, Wolters A. How has COVID-19 affected service delivery in GP practices that offered remote consultations before the pandemic? The Health Foundation. 2020. URL: https://www.health.org.uk/news-and-comment/charts-and-infographics/how-has-covid-19-affected-service-delivery-in-gp-practices [accessed 2021-08-23]
3. Pérez Sust P, Solans O, Fajardo JC, Medina Peralta M, Rodenas P, Gabaldà J, et al. Turning the crisis into an opportunity: digital health strategies deployed during the COVID-19 outbreak. JMIR Public Health Surveill 2020 May 4;6(2):e19106. [doi: 10.2196/19106] [Medline: 32339998]
4. López Seguí F, Walsh S, Solans O, Adroher Mas C, Ferraro G, García-Altés A, et al. Teleconsultation between patients and health care professionals in the Catalan primary care service: message annotation analysis in a retrospective cross-sectional study. J Med Internet Res 2020 Sep 17;22(9):e19149 [FREE Full text] [doi: 10.2196/19149] [Medline: 32687477]
5. Pifarré i Arolas H, Vidal-Alaball J, Gil J, López F, Nicodemo C, Saez M. Missing Diagnoses during the COVID-19 Pandemic: A Year in Review. Int J Environ Res Public Health 2021 May 17;18(10):5335. [doi: 10.3390/ijerph18105335]
6. Coma E, Guiriguet C, Mora N, Marzo-Castillejo M, Benítez M, Méndez-Boo L, et al. Impact of the COVID-19 pandemic and related control measures on cancer diagnosis in Catalonia: a time-series analysis of primary care electronic health records covering about five million people. BMJ Open 2021 May 18;11(5):e047567 [FREE Full text] [doi: 10.1136/bmjopen-2020-047567] [Medline: 34006554]
7. Maringe C, Spicer J, Morris M, Purushotham A, Nolte E, Sullivan R, et al. The impact of the COVID-19 pandemic on cancer deaths due to delays in diagnosis in England, UK: a national, population-based, modelling study. The Lancet Oncology 2020 Aug;21(8):1023-1034. [doi: 10.1016/s1470-2045(20)30388-0] [Medline: 32702310]
8. Dinmohamed AG, Cellamare M, Visser O, de Munck L, Elferink MAG, Westenend PJ, et al. The impact of the temporary suspension of national cancer screening programmes due to the COVID-19 epidemic on the diagnosis of breast and colorectal cancer in the Netherlands. J Hematol Oncol 2020 Nov 04;13(1):147 [FREE Full text] [doi: 10.1186/s13045-020-00984-1] [Medline: 33148289]
9. Marijon E, Karam N, Jost D, Perrot D, Frattini B, Derkenne C, et al. Out-of-hospital cardiac arrest during the COVID-19 pandemic in Paris, France: a population-based, observational study. The Lancet Public Health 2020 Aug;5(8):e437-e443. [doi: 10.1016/s2468-2667(20)30117-1]

10. Colivicchi F, Di Fusco SA, Magnanti M, Cipriani M, Imperoli G. The impact of the coronavirus disease-2019 pandemic and Italian lockdown measures on clinical presentation and management of acute heart failure. J Card Fail 2020 Jun;26(6):464-465 [FREE Full text] [doi: 10.1016/j.cardfail.2020.05.007] [Medline: 32417376]
11. Eberly LA, Kallan MJ, Julien HM, Haynes N, Khatana SAM, Nathan AS, et al. Patient characteristics associated with telemedicine access for primary and specialty ambulatory care during the COVID-19 pandemic. JAMA Netw Open 2020 Dec 01;3(12):e2031640 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.31640] [Medline: 33372974]
12. Solans O, Vidal-Alaball J, Roig Cabo P, Mora N, Coma E, Bonet Simó JM, et al. Characteristics of citizens and their use of teleconsultations in primary care in the Catalan Public Health System before and during the COVID-19 pandemic: retrospective descriptive cross-sectional study. J Med Internet Res 2021 May 27;23(5):e28629 [FREE Full text] [doi: 10.2196/28629] [Medline: 33970867]
13. Pifarré I Arolas H, Acosta E, López-Casasnovas G, Lo A, Nicodemo C, Riffe T, et al. Years of life lost to COVID-19 in 81 countries. Sci Rep 2021 Feb 18;11(1):3504 [FREE Full text] [doi: 10.1038/s41598-021-83040-3] [Medline: 33603008]
14. García-Basteiro A, Alvarez-Dardet C, Arenas A, Bengoa R, Borrell C, Del Val M, et al. The need for an independent evaluation of the COVID-19 response in Spain. Lancet 2020 Aug 22;396(10250):529-530 [FREE Full text] [doi: 10.1016/S0140-6736(20)31713-X] [Medline: 32771082]
15. CRES-UPF. Por un Hispa-NICE: Ahora o Nunca (Manifesto). Universitat Pompeu Fabra Barcelona. URL: https://www.upf.edu/es/web/cres/noticies/-/asset_publisher/R0ylBfnCHIIE/content/id/240892331/maximized#.YHaTzugzZPY [accessed 2021-04-14]
16. International classification of diseases, tenth revision, clinical modification (ICD-10-CM). Centers for Disease Control and Prevention. 2019. URL: http://www.cdc.gov/nchs/icd/icd10cm.htm [accessed 2021-08-23]
17. Rajaraman A, Ullman JD. Data mining. In: Mining of Massive Datasets. Cambridge: Cambridge University Press; 2011:1-17.
18. Manning C. Scoring, term weighting, and the vector space model. In: Introduction to Information Retrieval. Cambridge: Cambridge University Press; 2008.
19. Lopez-Valcarcel BG, Barber P. Economic crisis, austerity policies, health and fairness: lessons learned in Spain. Appl Health Econ Health Policy 2017 Feb;15(1):13-21. [doi: 10.1007/s40258-016-0263-0] [Medline: 27461007]
20. Xiong J, Lipsitz O, Nasri F, Lui LM, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: a systematic review. J Affect Disord 2020 Dec 01;277:55-64 [FREE Full text] [doi: 10.1016/j.jad.2020.08.001] [Medline: 32799105]

Abbreviations

***ICD-10*:** *International Statistical Classification of Diseases, tenth revision*

Edited by C Basch; submitted 14.04.21; peer-reviewed by A Roghani, W Kraaij, X Cheng, Z Reis; comments to author 11.05.21; revised version received 18.05.21; accepted 12.06.21; published 14.09.21.

Please cite as:
Lopez Segui F, Hernandez Guillamet G, Pifarré Arolas H, Marin-Gomez FX, Ruiz Comellas A, Ramirez Morros AM, Adroher Mas C, Vidal-Alaball J
Characterization and Identification of Variations in Types of Primary Care Visits Before and During the COVID-19 Pandemic in Catalonia: Big Data Analysis Study
J Med Internet Res 2021;23(9):e29622
URL: https://www.jmir.org/2021/9/e29622
doi:10.2196/29622
PMID:34313600

Original Paper

Public Adoption of and Trust in the NHS COVID-19 Contact Tracing App in the United Kingdom: Quantitative Online Survey Study

Liz Dowthwaite[1], PhD; Joel Fischer[2], PhD; Elvira Perez Vallejos[3], PhD; Virginia Portillo[1], PhD; Elena Nichele[1], PhD; Murray Goulden[4], PhD; Derek McAuley[1], PhD

[1]Horizon Digital Economy Research, University of Nottingham, Nottingham, United Kingdom

[2]School of Computer Science, University of Nottingham, Nottingham, United Kingdom

[3]National Intitute for Health Research Biomedical Research Centre, Institute of Mental Health, Division of Psychiatry and Applied Psychology, University of Nottingham, Nottingham, United Kingdom

[4]School of Sociology and Social Policy, University of Nottingham, Nottingham, United Kingdom

Corresponding Author:
Liz Dowthwaite, PhD
Horizon Digital Economy Research
University of Nottingham
Innovation Park
Triumph Road
Nottingham, NG7 2TU
United Kingdom
Phone: 44 07943583035
Email: liz.dowthwaite@gmail.com

Abstract

Background: Digital contact tracing is employed to monitor and manage the spread of COVID-19. However, to be effective the system must be adopted by a substantial proportion of the population. Studies of mostly hypothetical contact tracing apps show generally high acceptance, but little is known about the drivers and barriers to adoption of deployed systems.

Objective: The aim of this study was to investigate adoption of and attitudes toward the NHS (National Health Service) COVID-19 smartphone app, the digital contact tracing solution in the United Kingdom.

Methods: An online survey based on the extended Technology Acceptance Model with the added factor of trust was carried out with a representative sample of the UK population. Statistical analysis showed adoption rates, attitudes toward and trust in the app, and compliance with self-isolation advice and highlighted differences for vulnerable populations (ie, older adults aged 65 years and over and members of Black, Asian, and minority ethnic [BAME] communities).

Results: A total of 1001 participants took part in the study. Around half of the participants who had heard of the NHS COVID-19 mobile phone app (490/963, 50.9%; 95% CI 47.8%-54.0%) had downloaded and kept the app, but more than one-third (345/963, 35.8%; 95% CI 32.8%-38.8%) either did not intend to download it or had deleted it. Significantly more BAME respondents than White respondents had deleted the app (16/115, 13.9%; 95% CI 11.8%-16.0%, vs 65/876, 7.4%; 95% CI 5.8%-9.0%), and significantly more older adults 65 years and over than those under 65 years did not intend to download it (44/127, 34.6%; 95% CI 31.7%-37.5%, vs 220/874, 25.2%; 95% CI 22.5%-27.9%). Broadly, one of the reasons for uptake was to help the NHS and other people, especially among older adults, although significantly fewer BAME participants agreed that they did so to help the NHS. Reported compliance with received notifications to self-isolate was high but was significantly lower than reported intended compliance without received notifications. Only one-fifth (136/699, 19.5%; 95% CI 17.0%-22.0%) of participants understood that the decision to send self-isolation notifications was automated by the app. There were a range of significantly more negative views among BAME participants, including lower trust in the NHS, while older adults were often significantly more positive. Respondents without the app reported significantly lower trust and more negative views toward the app and were less likely to report that they understood how the app works.

Conclusions: While compliance on the part of the approximately 50% of participants who had the app was fairly high, there were issues surrounding trust and understanding that hindered adoption and, therefore, the effectiveness of digital contact tracing,

particularly among BAME communities. This study highlights that more needs to be done to improve adoption among groups who are more vulnerable to the effects of the virus in order to enhance uptake and acceptance of contact tracing apps.

(J Med Internet Res 2021;23(9):e29085) doi:10.2196/29085

KEYWORDS

trust; technology adoption; COVID-19; digital contact tracing; coronavirus; vulnerable populations; attitudes; SARS-CoV-2; digital proximity tracing; compliance

Introduction

Digital contact tracing solutions are employed to monitor and manage the spread of disease during pandemics [1]. Public acceptance of app-based contact tracing in the United Kingdom, the European Union, and the United States is high [2-4]; however, to make a difference they must be adopted by a substantial proportion of the population [5,6]. Engaging users in the development, implementation, and evaluation of contact tracing can help maximize engagement and technology acceptance, according to the Responsible Research and Innovation (RRI) framework [7,8]. Therefore, this paper reports on research to understand the drivers and barriers to adoption of a COVID-19 contact tracing app in the United Kingdom to help increase the effectiveness of such systems and inform future design.

An earlier study of hypothetical digital contact tracing in the United Kingdom suggested that people would adopt it to protect family, friends, and the community as well as to stop the pandemic, while potential barriers were reported risks of postpandemic surveillance, increasing anxiety, and fear of hacking [2,4]. In a study by Velicia-Martin et al, the Technology Acceptance Model (TAM) was used to show that the intention to use a contact tracing app was determined by how useful it was perceived to be; the study also showed that concern about privacy would be overridden by concerns about health [9]. Similarly, a study in Germany looked at the difference between the perceived utility of a contact tracing app and a data donation app. Motivations for using and accepting a contact tracing app were higher, and the data donation app was perceived as having less utility for the user [10]. However, there may also be wider social implications, such as having no choice but to download the app for work or venue check-ins, and real-world uptake might differ from usage within a trial.

Trust may also significantly impact the adoption of contact tracing apps, [11]; for example, a study conducted in Switzerland suggested that higher levels of trust in government and health authorities may also lead to increased uptake [12]—perceived effectiveness of a contact tracing app and the overall app user experience depended on the app being embedded within the health system. In the UK context, this may mean that trust or confidence in the National Health Service (NHS) might influence people's attitudes toward and usage of the app. In Germany, it was also shown that general trust in official app providers as well as social trust played important roles, highlighting the importance of both data-securing issues and interpersonal solidarity [10]. A study across five countries—France, Germany, Italy, the United Kingdom, and the United States—also found that a lack of trust was one of the main barriers for adoption of a hypothetical contact tracing app [4].

Security and privacy have also been shown to be important. In the Netherlands, a study designed to determine the potential uptake of a contact tracing app in the Dutch population showed an adoption rate as high as 64% for apps that had a higher number of security and privacy-respecting features [13]. In Ireland, there was shown to be a high level of willingness to download a public health–backed app to augment contact tracing, with 54% of respondents definitely willing to download an app and 30% who would consider downloading it [14]. The most common reason to download the app was linked to social altruism: helping family and friends and a sense of responsibility to the wider community. The most common reason not to download the app was linked to issues of trust, privacy and data security, and fear that technology companies or the government might use the app technology for greater surveillance after the pandemic. Another study across five countries—France, Germany, Italy, the United Kingdom, and the United States—found that there was strong support for an app, whether it was subject to voluntary or automatic installation, but the study identified concerns once more about cybersecurity and privacy [4]. In Australia, 37.3% of participants in a study of 1500 citizens had downloaded the COVIDSafe app, and 27.7% refused to do so [15]. Reasons for not downloading the app included privacy and technical concerns, the belief that that app was unnecessary due to social distancing, distrust in the government, and apathy. This study also highlighted the importance of public health messages for increasing the acceptability of apps and their correct use, while also addressing concerns around privacy, data storage, and technical ability needed to operate the app; it also emphasized the importance of identifying and understanding specific barriers to the use of contact tracing apps to improve their design.

Data suggest that specific groups of the population are more at risk of dying of COVID-19, including older adults [16-18] as well as those in Black, Asian, and minority ethnic (BAME) communities [19,20]. The main challenge among the Dutch population was to increase the uptake among older adults, who were least inclined to install and use a COVID-19 contact tracing app [13]. In line with this, in Germany, age was negatively associated with the motivation for using a data donation app [10]. The risk for BAME communities, in particular, has been linked to socioeconomic factors [21,22]. Furthermore, recent studies have shown increased COVID-19 vaccine hesitancy in certain BAME communities [23-25]; however, little is known about whether hesitancy also extends to attitudes toward digital contact tracing.

At the time the study reported in this paper closed on December 21, 2020, 2,183,506 people in the United Kingdom had tested positive for COVID-19, approximately 3.2% of the population [26]. The UK government released the NHS COVID-19 app on September 24, 2020. The app is entirely automated and decentralized, and it relies on Bluetooth proximity as well as self-reporting of symptoms and test results [27,28]. The app has been downloaded more than 21 million times, suggesting a 56% uptake among the population aged 16 years and older who own a smartphone [29,30]. A recent study also showed that the app has been effective in reducing the number of positive cases of COVID-19 in the United Kingdom; 1.7 million users were contact-traced by the app, with an estimated reduction in cases during the second wave of COVID-19 by one-quarter. However, only 28% of people had actively used the app in the period leading up to this study [31]. Little is known about the views driving or hindering adoption of the app; our research addresses this gap. The study surveys public trust in and adoption of digital contact tracing in the United Kingdom, in terms of reported reasons, compliance, and understanding of the app, especially highlighting significant differences for vulnerable groups.

Methods

Recruitment

Ethical approval was granted for the study by the Research Ethics Committee of the authors' institution. The recruitment was carried out by Ipsos MORI via email to a nationally representative sample, based on age, gender, and region, drawn from a randomly selected pool of participants who met the relevant criteria. There was also a 10% to 15% quota for BAME respondents, with the same process applied to ensure hitting the minimum required quota. As fieldwork progressed, they specifically targeted any quota groups that were still required to meet the final profile that was needed, again randomly selecting within those groups. Participants were incentivized for survey participation through monetary compensation paid into their panel account. As a market research agency, Ipsos MORI operates under the Market Research Society code of conduct and is General Data Protection Regulation compliant, so participants' privacy was guaranteed. Data were only received via their survey platform in an anonymized form, so no personally identifiable information on the participants was received. A total of 2575 invitations to take part in the study were sent out.

A representative sample of 1001 members of the UK population aged 16 to 75 years took part, weighted to the known offline proportions for age within gender, region, working status, and ethnicity. Participants were asked a series of demographic questions, the full details of which are provided in Table S1 in Multimedia Appendix 1. A summary of the main self-reported characteristics is provided in Table 1.

Table 1. Summary characteristics of participants.

Characteristic	Participants (N=1001), n (%)
Age (years)	
Under 65	874 (87.3)
65 and over	127 (12.7)
Gender	
Male	501 (50.0)
Female	500 (50.0)
Employment status	
Working	666 (66.5)
Not working	335 (33.5)
Education	
Up to General Certificate of Secondary Education (GSCE)	307 (30.7)
Post-GSCE	694 (69.3)
Ethnicity	
White	876 (87.5)
Black, Asian, and minority ethnic	115 (11.5)
Not stated	10 (1.0)
Country of residence	
England	847 (84.6)
Wales	48 (4.8)
Scotland	85 (8.5)
Northern Ireland	21 (2.1)

Materials and Procedure

An online survey was carried out between December 11 and 21, 2020, when the United Kingdom was between "lockdown 2" and "lockdown 3" and subject to a regional tier system. Questionnaire development was carried out in several stages. First, in the summer of 2020, a series of interviews were carried out with members of the public with regard to their opinions of, and intention to use, the United Kingdom's test and trace app when it would be released (paper forthcoming). From these interviews, a series of themes were identified, which led to the survey being based on the extended TAM (TAM2) [32]. The TAM2 identifies several key factors in the adoption of new technology and has been applied to explore acceptance of various technologies, including hypothetical COVID-19 tracing apps [9]. The conceptual model is extended with "trust" as a factor for acceptance, as it may significantly impact the adoption of contact tracing apps [11]. A list of pertinent questions was developed among the author team; the questions were tested and refined involving experts in questionnaire development from Ipsos MORI. Figure 1 illustrates how these questions relate to and extend the TAM2 framework.

Recruitment and data collection were carried out by Ipsos MORI, who also carried out piloting of the questionnaire. The questionnaire had an initial data review on Day 1 with 61 respondents; it was thoroughly checked to make sure that all data were being collected correctly (eg, checking routing and displaying of correct answer options) and was checked for anomalies and understanding. The data were reviewed again when 213 respondents had completed it to ensure data quality.

Participants were provided with information and privacy notices and gave informed consent to take part. All questions were closed-ended, either multiple choice or rated on Likert or Likert-like scales from 1 ("strongly disagree" or "not at all") to 5 ("strongly agree" or "entirely"), except for a single open-ended question that was included for further comments; participants were routed to appropriate questions based on previous answers.

Figure 1. The extended Technology Acceptance Model (TAM2) and its relation to the this questionnaire study. Items in bold in white boxes are existing factors in the TAM2 [8], followed by examples, as bullet points, from the questionnaire. Trust was added as an additional factor (peach box) affecting intention to use and includes examples from the questionnaire.

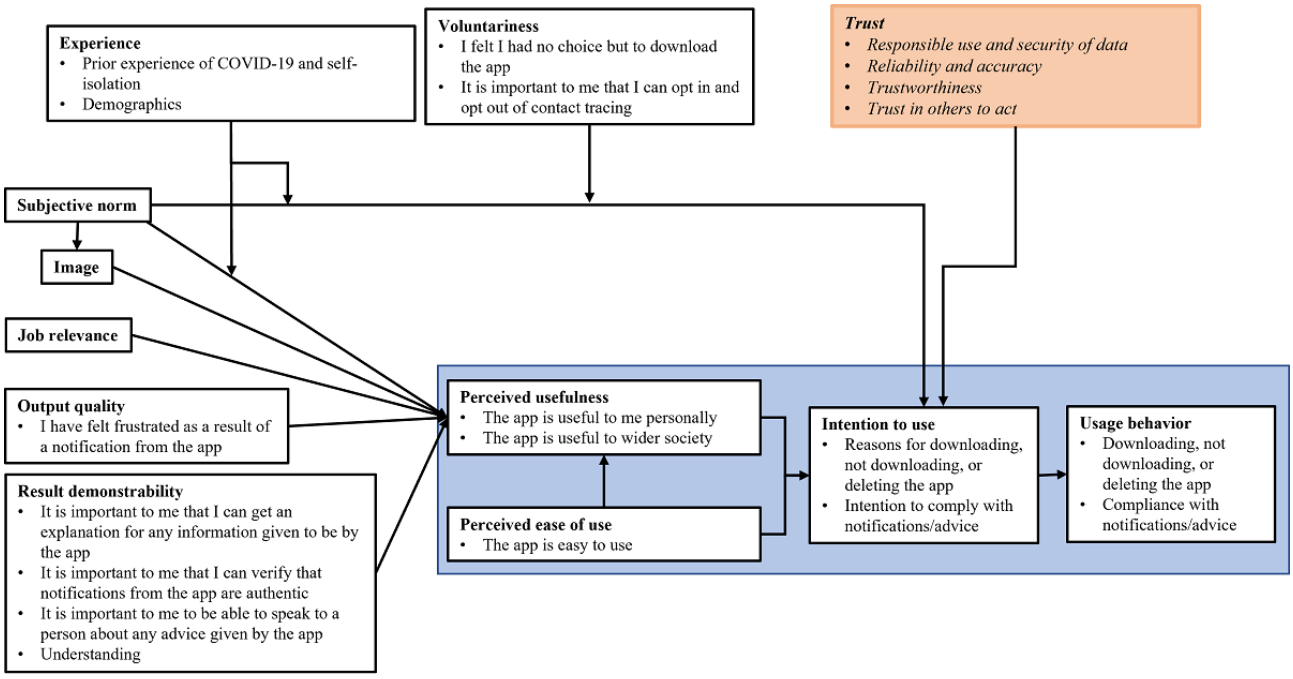
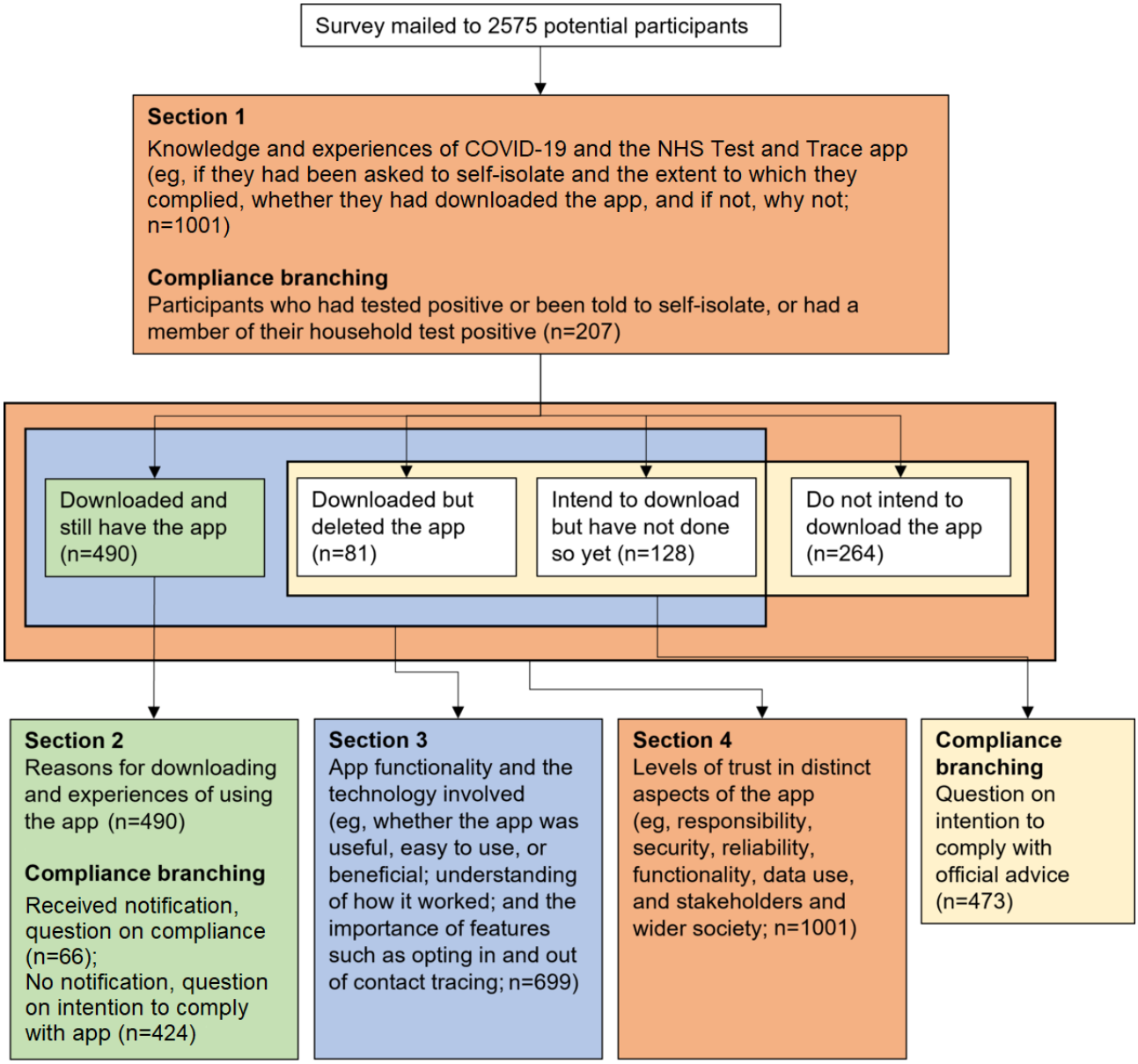


Figure 2 shows the survey flow and resulting major subpopulations that were used for branching. Section 1 of the survey asked participants to indicate what knowledge and experiences they had of COVID-19 and the NHS Test and Trace app; for example, if they had been asked to self-isolate and the extent to which they complied, whether they had downloaded the app, and if not, why not. Section 2 focused on those with the app and collected reasons for downloading and experiences of using the app. Section 3 asked about app functionality and the technology involved, including whether the app was useful, easy to use, or beneficial; understanding of how it worked; and the importance of features such as opting in and out of contact tracing. Section 4 asked about levels of trust in distinct aspects of the app, including responsibility, security, reliability, functionality, data use, and stakeholders and wider society.

Figure 2. Illustration of major sections of the survey and subgroups identified for branching. NHS: National Health Service.



Statistical Analysis

Responses were analyzed using SPSS software (version 26; IBM Corp) and Excel (Microsoft 365). Summary statistics (ie, mean, median, SD, and IQR) or frequencies were extracted for all questions. Confidence intervals for proportions are given at the 95% level. Most questions were significantly nonnormal as shown by skewness and kurtosis, so nonparametric tests are appropriate. In the text, we report the sample mean, given with statistical test results, and the median response on a Likert-like scale. All inferential statistical analysis was carried out with $P<.05$ as the threshold for statistical significance. Missing data were reported as "no response" and were included in frequency calculations; missing data for inferential statistical analysis and the calculation of means were excluded.

Subgroup analyses using independent-samples Mann-Whitney *U* or chi-square tests were used to compare those who had been told to self-isolate with those who had not, White participants with BAME participants, and participants under 65 years with participants 65 years and over. Independent-samples Kruskal-Wallis tests were used to compare those who had the app, did not have the app, had deleted the app, or intended to download the app but had not yet done so. Post hoc tests were carried out to identify which groups had significant differences between them, with Bonferroni corrections to account for multiple testing. Due to use of the nonparametric versions of inferential statistics, and because the weighted values for the data all rounded to 1, weighting was not used. However, exploratory analysis revealed no difference in significance when using the parametric versions of tests.

Results

Adoption of the Test and Trace App

A total of 1001 participants took part in the study. Most participants (n=963, 96.2%; 95% CI 95.0%-97.4%) had heard of the NHS COVID-19 mobile phone app, of which 50.9% (490/963; 95% CI 47.8%-54.0%) had downloaded the app and still had it on their phone. A further 13.3% (128/963; 95% CI

11.2%-15.4%) had not yet downloaded it but intended to, 27.4% (264/963; 95% CI 24.6%-30.2%) did not intend to download it, and 8.4% (81/963; 95% CI 6.7%-10.1%) had downloaded it but since deleted it.

Among the 27.4% (264/963) of participants who did not intend to download the app, the most common reasons were a desire not to be tracked, not thinking it would be effective, not wanting to take part in contact tracing in that way, and lack of trust in those who built the app (Table 2). Of the 8.4% (81/963) who had decided to delete the app, this was mostly because they did not think it was effective or did not want to be tracked. Reasons given by the 13.3% (128/963) of participants who intended to download the app were mostly to help the NHS or to help protect their friends and family or themselves, as well as to reduce the spread of the virus and to help protect broader society (Table 3).

Of the 50.9% (490/963) of participants who had downloaded the app, 92.0% (451/490; 95% CI 90.3%-93.7%) had opened the app and had a look around, 66.7% (327/490; 95% CI 63.8%-69.6%) had used it for a venue check-in, 58.4% (286/490; 95% CI 55.3%-61.5%) had made use of the “check symptoms” section, 71.2% (349/490; 95% CI 68.4%-74.0%) had contact tracing always switched on, 20.4% (100/490; 95% CI 17.9%-22.9%) sometimes had contact tracing switched on, 1.8% (9/490; 95% CI 1.0%-2.6%) never turned contact tracing on, and 6.5% (32/490; 95% CI 5.0%-8.0%) did not know if contact tracing was enabled or not. The strongest reasons given for downloading the app were helping the NHS and protecting friends and family (Table 3).

Table 2. Reasons for not having the NHS Test and Trace app for those who do not intend to download it and those who downloaded but deleted the app.

Reasons for not having the app[a]	Participants who do not intend to download the app (n=264)		Participants who downloaded but deleted the app (n=81)	
	n (%)	95% CI	n (%)	95% CI
I don’t want to be tracked	105 (39.8)	36.8-42.8	22 (27.2)	24.4-30.0
I don’t think it will be effective	80 (30.3)	27.5-33.1	28 (34.6)	31.7-37.5
I choose/chose not to take part in contact tracing in this way	78 (29.5)	26.7-32.3	12 (14.8)	12.6-17.0
I don’t trust the people who built the app	72 (27.3)	24.5-30.1	14 (17.3)	15.0-19.6
The app doesn’t/didn’t work on my mobile phone	28 (10.6)	8.7-12.5	14 (17.3)	15.0-19.6
I don’t have a smartphone	27 (10.2)	8.3-12.1	N/A[b]	N/A
I don’t want to be told to self-isolate	16 (6.1)	4.9-7.6	9 (11.1)	9.2-13.0
None of the above	25 (9.5)	7.7-11.3	7 (8.6)	6.9-10.3
I wouldn’t/didn’t know how to use it	12 (4.5)	3.2-5.8	12 (14.8)	12.6-17.0
Don’t know	1 (0.1)	0-0.3	0 (0)	__[c]

[a]Multiple answers were allowed.

[b]N/A: not applicable, because participants do not have a smartphone, which is needed to download the app.

[c]The 95% CI value could not be calculated since there were no participants.

Table 3. Reasons for intention to download and for downloading the NHS COVID-19 app.

Reasons for intention to download and for downloading the app	Participants who intended to download the app[a] (n=128)		The extent to which each reason was a motivation for downloading the app[b] (n=490)	
	n (%)	95% CI	Mean (SD)	Median (IQR)
To help the National Health Service	84 (65.6)	62.7-68.5	4.42 (0.753)	5 (1)
To help protect my friends and family	80 (62.5)	59.5-65.5	4.36 (0.792)	5 (1)
To help protect myself	70 (54.7)	51.6-57.8	4.27 (0.883)	4 (1)
Because it will reduce the spread of the virus	55 (43.0)	39.9-46.1	4.11 (0.966)	4 (1)
To help protect broader society	46 (35.9)	32.9-46.1	4.2 (0.897)	4 (1)
Because I need it to check into venues	24 (18.8)	16.4-21.2	3.54 (1.179)	4 (1)
Because the government told me to	15 (11.7)	9.7-13.7	3.46 (1.177)	4 (1)
Because everyone else is	12 (9.4)	7.6-11.2	3.14 (1.177)	3 (2)
Because it is a requirement for my job	7 (5.5)	4.1-6.9	2.53 (1.361)	2 (3)
None of the above	2 (1.6)	0.8-2.4	N/A[c]	N/A

[a]Participants who intended to download the app were asked to select all reasons that applied.

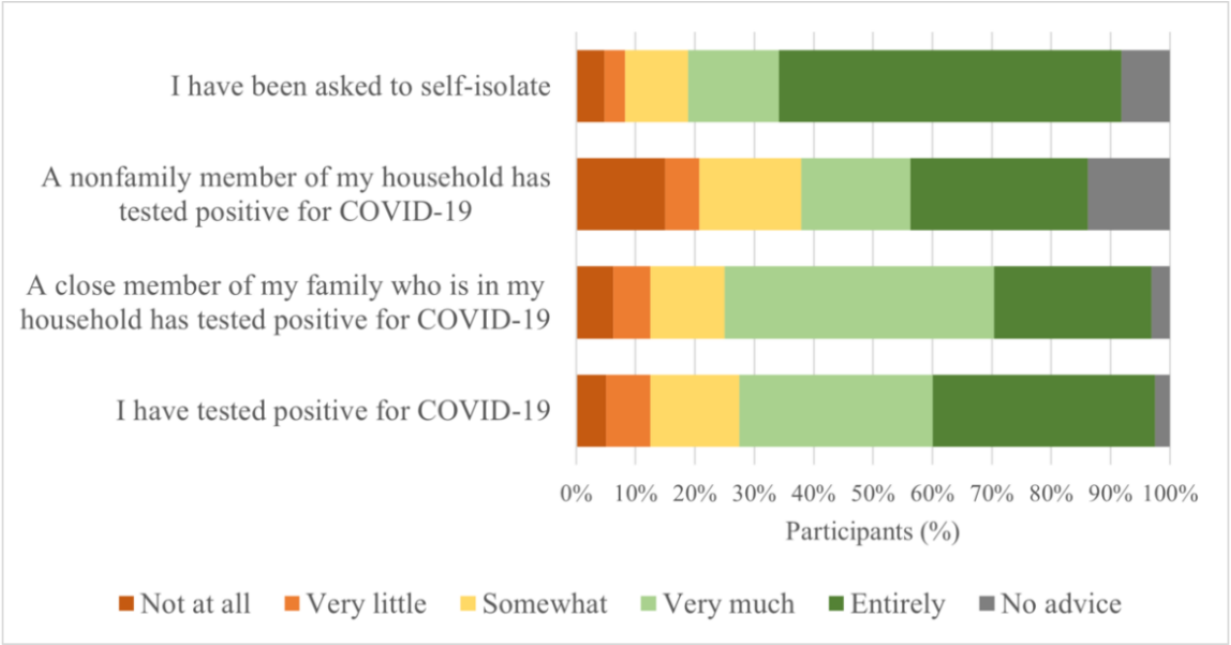
[b]Participants who had the app were asked to what extent each reason was a motivation for downloading the app; responses were rated on a 5-point Likert scale from 1 (“strongly disagree”) to 5 (“strongly agree”).

[c]N/A: not applicable; this entry was not rated.

Compliance With Official Advice and Self-isolation

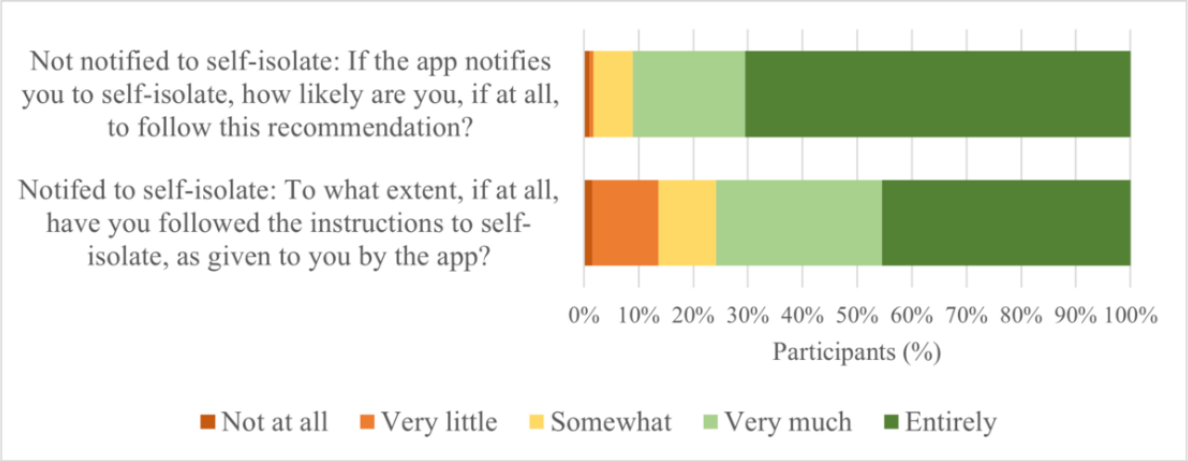
Out of 1001 participants, nearly half (n=434, 43.4%; 95% CI 40.3%-46.5%) had at least one of the following experiences: 4.0% (n=40; 95% CI 2.8%-5.2%) had tested positive for COVID-19, 14.1% (n=141; 95% CI 11.9%-16.3%) had a member of their household test positive, 27.9% (n=279; 95% CI 25.1%-30.7%) had another person close to them test positive, and 8.5% (n=85; 95% CI 6.8%-10.2%) had been asked to self-isolate in any form, whether via app or other means. Participants who had been affected or asked to self-isolate, or who had a member of their household who had (207/1001, 20.7%), were asked to indicate how much they had complied with any official advice they received, regardless of whether it was through the app or another source; the average response across all four experiences was “very much” (mean 3.88, SD 1.292). However, 10.6% (22/207; 95% CI 7.9%-11.5%) stated they did not receive any official advice at all, most often when a nonfamily member of their household had tested positive (12/87, 14%; 95% CI 11.7%-15.9%) (Figure 3). Note that this implies that some participants, therefore, did not consider being asked to self-isolate as “official advice.”

Figure 3. Compliance with official advice, dependent on the circumstances. Participants were asked "To what extent, if at all, did you follow any advice given to you?".



Of the 47.3% (473/1001) of participants who did not currently have the app, most stated they would either very much or entirely follow advice to self-isolate if they received a call (352/473, 74.4%; 95% CI 71.7%-77.1%). Of participants with the app (490/963, 50.9%), 13.5% (66/490; 95% CI 11.4%-15.6%) had been notified to self-isolate, and 45% (30/66; 95% CI 42.4%-48.6%) of those said that they had entirely followed the recommendation. However, of the 85.3% (418/490; 95% CI 83.1%-87.5%) who stated that they had not been notified by the app to self-isolate, 70.6% (295/418; 95% CI 73.4%-78.6%) said they would entirely follow a recommendation from the app if they received one (Figure 4). An independent-samples Mann-Whitney *U* test comparing those who had received advice to self-isolate from the app (66/490, 13.5%) to those who had not (418/490, 85.3%) shows that reported intention to comply with advice is significantly stronger than reported actual compliance ($mean_{intention}$ 4.59, SD 0.738; $mean_{actual}$ 4.06, SD 1.094; U=17673.0; P<.001).

Figure 4. Compliance and intention to comply with app notifications to self-isolate. An independent-samples Mann-Whitney *U* test comparing those who had received advice to self-isolate (n=66) from the app to those who had not (n=418) shows that reported intention to comply with advice is significantly stronger than reported actual compliance (mean intention 4.59, SD 0.738; mean actual 4.06, SD 1.094; U=17673.0; P<.001).



Understanding and Attitudes Toward the App

Of participants who currently have the app, have deleted it, or intended to download it (699/963, 72.6%), most stated that they thought that decisions to send a notification to self-isolate were made by both humans and the app (379/699, 54.2%; 95% CI 51.1%-57.3%); only 19.5% (136/699; 95% CI 17.0%-22.0%) thought they were made by the app only. Participants with the app agreed that they knew how the app worked, that it was easy to use, that it was useful to them and to wider society, that the regulations surrounding the app were sufficient, that it was important that they could get explanations and verify information from the app, that they could speak to a person about any advice they receive, and that they have the option to opt out of contact tracing if they chose to. They were neutral about data concerns and about whether they had a choice in downloading the app. They tended to disagree that they had been frustrated by a notification from the app. Independent-samples Kruskal-Wallis tests showed that participants who had deleted the app (81/963, 8.4%) felt significantly less than those who still have the app (490/963, 50.9%) that they understood the app (χ^2_2=16.1, P=.003), that it was useful to them (χ^2_2=26.2, P<.001) or wider society (χ^2_2=29.7, P<.001), that regulations were sufficient (χ^2_2=12.9, P=.003), or that it was easy to use (χ^2_1=41.2, P<.001). It was also significantly less important to them that they could verify app notifications (χ^2_2=18.3, P<.001). However, they showed significantly more concern about how their data were used (χ^2_2=25.7, P<.001) and were more likely to have been frustrated by a notification from the app (χ^2_1=18.3, P<.001). None of the other statements (Table 4) were significantly different.

Table 4. Levels of agreement with statements related to the technology and ecosystem surrounding the NHS COVID-19 app.

Statements[a]	Participants who still have the app (n=490)		Participants who deleted the app (n=81)		*P* value[b]
	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)	
I understand how the NHS COVID-19 app works	3.99 (0.850)	4 (1)	3.57 (1.036)	4 (1)	.003
I am concerned about how my data will be used by the app	3.04 (1.250)	3 (2)	3.73 (1.037)	4 (1)	<.001
The app is useful to me personally	3.84 (0.944)	4 (2)	3.19 (1.174)	3 (2)	<.001
The app is useful to wider society	4.11 (0.874)	4 (1)	3.47 (1.096)	4 (1)	<.001
It is important to me that I can get an explanation for any information given to me by the app	4.01 (0.801)	4 (1)	3.72 (1.028)	4 (2)	.052
It is important to me that I can verify that notifications from the app are authentic	4.14 (0.823)	4 (1)	3.60 (1.137)	4 (1)	<.001
The regulations governing the creation of the app are sufficient	3.72 (0.934)	4 (1)	3.37 (1.089)	3 (1)	.003
It is important to me to be able to speak to a person about any advice given by the app	3.71 (0.982)	4 (1)	3.64 (1.099)	4 (1)	>.99
It is important to me that I can opt in and opt out of contact tracing	3.50 (1.166)	4 (1)	3.58 (0.947)	4 (1)	>.99
The app is easy to use	4.18 (0.815)	4 (1)	3.42 (1.082)	4 (1)	<.001
I felt that I had no choice but to download the app	2.98 (1.273)	3 (2)	3.27 (1.162)	3 (2)	.06
I have felt frustrated as a result of a notification from the app	2.59 (1.279)	2 (2)	3.28 (1.154)	3 (2)	<.001

[a]Statements were rated on a 5-point Likert scale from 1 (“strongly disagree”) to 5 (“strongly agree”).

[b]Independent-samples Kruskal-Wallis tests were carried out with a significance level of $P<.05$ and post hoc tests to indicate which groups had significant differences between them, with Bonferroni correction to account for multiple tests. Additional differences, including participants who downloaded and still have the app and those who intend to download the app, are in Table S2 in Multimedia Appendix 1.

Trust in Test and Trace

While those who still have the app (490/963, 50.9%) tended to agree that they had trust in various aspects of the app (Table 5), independent-samples Kruskal-Wallis tests showed that those who chose not to download the app (264/963, 27.4%) had significantly less trust, feeling neutral regarding trusting that the data were used responsibly (χ^2_3=222.2, $P<.001$) and stored securely (χ^2_3=236.3, $P<.001$), that the app does what it is supposed to do (χ^2_3=273.9, $P<.001$), and that the app is basically trustworthy (χ^2_3=243.1, $P<.001$). They were also significantly less trusting of others, feeling neutral about whether they trusted others to download the app (χ^2_3= 128.8, $P<.001$) or to self-isolate if they were told to (χ^2_3=74.1, $P<.001$). Participants who chose not to download the app were also significantly more likely not to trust that their data would be deleted when the app said it would be (χ^2_3=251.1, $P<.001$) or that the app was reliable (χ^2_3=277.7, $P<.001$). Trust was thought to be important for all participants to feel comfortable using the app, although significantly less so for those who chose not to download it (χ^2_3=24.0, $P<.001$).

Table 5. Levels of agreement with statements related to trust in the NHS COVID-19 app.

Statements[a]	Participants who have the app (n=490)		Participants who do not intend to download the app (n=264)		*P* value[b]
	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)	
I trust that the data collected by the app are used responsibly	3.98 (0.888)	4 (1)	2.63 (1.224)	3 (2)	<.001
I trust that the data collected by the app are stored securely	3.93 (0.916)	4 (2)	2.56 (1.149)	3 (1)	<.001
I feel that the app is reliable	3.89 (0.919)	4 (2)	2.42 (1.062)	2 (1)	<.001
I trust that the app will do what it is supposed to do	3.97 (0.880)	4 (1)	2.53 (1.136)	3 (1)	<.001
I think the NHS COVID-19 app is basically trustworthy	4.03 (0.875)	4 (1)	2.73 (1.129)	3 (1)	<.001
I think that most other people will download the app	3.58 (1.026)	4 (1)	2.65 (1.086)	3 (1)	<.001
I trust that most other people will self-isolate if told to do so by the app	3.54 (1.113)	4 (1)	2.80 (1.106)	3 (2)	<.001
I trust that my data will be deleted when the app says they will be	3.93 (0.892)	4 (2)	2.48 (1.196)	2 (2)	<.001
It is important to me that I trust the app in order to use it	4.15 (0.736)	4 (1)	3.86 (1.145)	4 (2)	<.001

[a]Statements were rated on a scale from 1 (“strongly disagree”) to 5 (“strongly agree”).

[b]Independent-samples Kruskal-Wallis tests were carried out with a significance level of $P<.05$ and post hoc tests to indicate which groups had significant differences between them, with Bonferroni correction to account for multiple tests. Additional differences, including participants who deleted or intended to download the app, are in Table S3 in Multimedia Appendix 1.

While those who still had the app (490/963, 50.9%) agreed that they trusted most of the stakeholders involved in the NHS Test and Trace system (Table 6), independent-samples Kruskal-Wallis tests showed that those who chose not to download the app (264/963, 27.4%) were significantly more neutral about big tech companies ($\chi^2_3=82.1$, $P<.001$), small hospitality venues ($\chi^2_3=34.8$, $P<.001$), large hospitality venues ($\chi^2_3=56.1$, $P<.001$), and their local council ($\chi^2_3=61.7$, $P<.001$), and significantly more negative about their trust in the UK government ($\chi^2_3=61.7$, $P<.001$). Those with the app were neutral about their trust in private contractors, while those without the app were significantly more negative ($\chi^2_3=85.6$, $P<.001$). Finally, both groups tended to agree that they trusted the NHS, but those without the app significantly less so ($\chi^2_3=78.9$, $P<.001$).

Table 6. Levels of trust in stakeholders involved in the NHS Test and Trace system among participants.

Statements[a]	Participants who have the app (n=490)		Participants who do not intend to download the app (n=264)		*P* value[b]
	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)	
I trust the big tech companies, such as Google and Apple	3.42 (0.996)	4 (1)	2.65 (1.134)	3 (1)	<.001
I trust private contractors, such as Serco	3.07 (1.086)	3 (2)	2.29 (1.021)	2 (2)	<.001
I trust small hospitality venues, such as independent pubs and cafés	3.62 (0.864)	4 (1)	3.13 (1.125)	3 (2)	<.001
I trust larger hospitality venues, such as chain restaurants	3.52 (0.923)	4 (1)	2.91 (1.068)	3 (2)	<.001
I trust the UK Government	3.32 (1.213)	4 (2)	2.39 (1.181)	2 (2)	<.001
I trust my local council	3.51 (0.968)	4 (1)	2.86 (1.096)	3 (2)	<.001
I trust the National Health Service	4.33 (0.774)	4 (1)	3.72 (1.102)	4 (2)	<.001

[a]Statements were rated on a 5-point Likert scale from 1 (“strongly disagree”) to 5 (“strongly agree”).

[b]Independent-samples Kruskal-Wallis tests were carried out with a significance level of $P<.05$ and post hoc tests to indicate which groups had significant differences between them, with Bonferroni correction to account for multiple tests. Additional differences, including participants who deleted or intended to download the app, are in Table S4 in Multimedia Appendix 1.

Vulnerable Groups

Of the 11.5% (115/1001) of participants who identified as BAME, 53.0% (61/115; 95% CI 49.9%-56.1%) had any close experience of COVID-19, including friends and family receiving diagnoses, compared to 42.3% (371/876; 95% CI 39.2%-45.4%) of White participants (Table 7). Chi-square tests showed that significantly more BAME participants had a member of their household test positive, (χ^2_1=10.0, *P*<.05). While more BAME participants had tested positive or had another person close to them test positive than White participants, the differences were not significant; a similar proportion had been asked to self-isolate. Significantly fewer BAME participants than White participants had downloaded the app (χ^2_1=4.7, *P*<.05) and more had deleted the app (χ^2_1=4.5, *P*<.05); while fewer BAME participants did not intend to download the app and more intended to download it, neither difference was significant (Table 8). BAME participants agreed significantly less that they downloaded the app to help the NHS ($mean_{BAME}$ 4.02, SD 1.000; $mean_{White}$4.46, SD 0.710; *U*=7803.5; *P*=.001) and significantly more that it was a requirement for their job ($mean_{BAME}$ 3.19, SD 1.437; $mean_{White}$ 2.45, SD 1.341; *U*=13734.0; *P*<.001); no other reasons showed a difference.

Table 7. Experiences of COVID-19 among vulnerable populations (BAME participants and those 65 years and over) compared to other populations (White participants and those under 65 years).

Experiences of COVID-19	White participants (n=876)		BAME[a] participants (n=115)		Participants under 65 years (n=874)		Participants 65 years and over (n=127)	
	n (%)	95% CI	n (%)	95% CI	n (%)	95% CI	n (%)	95% CI
Any close experience of COVID-19	371 (42.3)	39.2-45.4	61 (53.0)	49.9-56.1	407 (46.6)	43.5-49.7	28 (22.0)	19.4-24.6
Tested positive	32 (3.7)	2.5-4.9	8 (7.0)	5.4-8.6	39 (4.5)	3.2-5.8	1 (0.9)	0.3-1.5
Member of household tested positive	112 (12.7)	10.6-14.8	29 (25.2)	22.5-27.9	136 (15.6)	13.4-17.8	5 (3.9)	2.7-5.1
Another close person tested positive	241 (27.5)	24.7-30.3	36 (31.3)	28.4-34.2	265 (30.3)	27.5-33.1	14 (11.0)	9.1-12.9
Asked to self-isolate	76 (8.7)	7.0-10.4	7 (7.0)	5.4-8.6	74 (8.5)	6.8-10.2	11 (8.8)	7.0-10.4

[a]BAME: Black, Asian, and minority ethnic.

Table 8. Downloads of the NHS COVID-19 app among vulnerable populations (BAME participants and those 65 years and over) compared to other populations (White participants and those under 65 years).

Download status	White participants (n=876)		BAME[a] participants (n=115)		Participants under 65 years (n=874)		Participants 65 years and over (n=127)	
	n (%)	95% CI	n (%)	95% CI	n (%)	95% CI	n (%)	95% CI
Downloaded	440 (50.2)	47.1-53.3	48 (41.7)	38.6-44.8	424 (48.5)	45.4-51.6	66 (52.0)	48.9-55.1
Downloaded then deleted	65 (7.4)	5.8-9.0	16 (13.9)	11.8-16.0	79 (9.0)	7.2-10.8	2 (1.6)	0.8-2.4
Not downloaded but intend to	105 (12.0)	10.0-14.0	19 (16.5)	14.2-18.8	115 (13.2)	11.1-15.3	13 (10.2)	8.3-12.1
Do not intend to download	236 (26.9)	24.2-29.6	24 (20.9)	18.4-23.4	220 (25.2)	22.5-27.9	44 (34.6)	32.6-38.6

[a]BAME: Black, Asian, and minority ethnic.

Of the 12.7% (127/1001) of participants who were 65 years of age or over, only 22.0% (28/127; 95% CI 19.4%-24.6%) had any close experience of COVID-19 compared to 46.6% (407/874; 95% CI 43.5%-49.7%) of participants under 65 years. Chi-square tests showed that significantly more participants under 65 years than those 65 years and over had a member of their household test positive (χ^2_1=10.4, *P*<.05) or another person close to them (χ^2_1=4.7, *P*<.05), with no significant difference in testing positive for COVID-19 or being asked to self-isolate. Significantly more participants 65 years and over had downloaded the app (χ^2_1=7.0, *P*<.05), fewer participants 65 years and over had deleted it (χ^2_1=8.2, *P*<.05), but significantly more participants 65 years and over, compared to those under 65 years, did not intend to download the app (χ^2_1=13.5, *P*<.05); similar proportions intended to download the app. Participants 65 years and over agreed significantly more that they downloaded the app to help the NHS ($mean_{\geq 65}$ 4.59, SD 0.656; $mean_{<65}$ 4.39, SD 0.764; *U*=16164.0; *P*=.02) and to help protect their friends and family ($mean_{\geq 65}$ 4.50, SD 0.846; $mean_{<65}$ 4.34, SD 0.782; *U*=16220.0; *P*=.02) or broader society ($mean_{\geq 65}$ 4.45, SD 0.786; $mean_{<65}$ 4.16, SD 0.910; *U*=16630.5; *P*=.008); however, they were less likely to agree that it was needed for them to check into venues ($mean_{\geq 65}$ 3.15, SD 1.256; $mean_{<65}$ 3.60, SD 1.157; *U*=11089.5; *P*=.005) or for their job ($mean_{\geq 65}$ 1.94, SD 1.188; $mean_{<65}$ 2.63, SD 1.364; *U*=9874.0; *P*<.001); no other reasons showed a difference.

Of participants who did not have the app, intention to comply with a phone call asking them to self-isolate was the same between both BAME and White participants ($mean_{BAME}$ 3.93, SD 1.201; $mean_{White}$ 4.14, SD 1.192; *U*=13909.0; *P*=.06) and between participants 65 years and over and those under 65 years ($mean_{\geq 65}$ 4.39, SD 1.000; $mean_{<65}$ 4.08, SD 1.207; *U*=10573.0;

P=.11). Of those with the app who had been notified to self-isolate, there was no significant difference in compliance between populations ($mean_{BAME}$ 3.67, SD 1.234; $mean_{White}$ 4.18, SD 1.034; *U*=291.5; *P*=.14; $mean_{\geq 65}$ 4.25, SD 1.500; $mean_{<65}$ 4.05, SD 1.078; *U*=150.0; *P*=.51), but of those who had not been notified, White participants reported a significantly higher intention to comply ($mean_{BAME}$ 4.26, SD 0.855; $mean_{White}$ 4.63, SD 0.700; *U*=4544.5; *P*=.006), as did participants 65 years and over ($mean_{\geq 65}$ 4.90, SD 0.349; $mean_{<65}$ 4.54, SD 0.774; *U*=13838.5; *P*<.001).

BAME participants had a significantly lower understanding of how decisions were made by the app (χ^2_2=9.2, *P*<.05) (Table 9); more participants thought it was either humans only or humans and the app, while far fewer understood that it was only the app. While more participants 65 years and over than those under 65 years also felt that decisions were made entirely by humans, fewer felt decisions were made by both humans and the app, and more of them correctly identified that decisions were made entirely by the app; there was no significant difference between the age groups (Table 9).

Table 9. Beliefs about how decisions are made by the NHS COVID-19 app among vulnerable populations (BAME participants and those 65 years and over) compared to other populations (White participants and those under 65 years).

Beliefs about how decisions are made by the app	White participants (n=876)		BAME[a] participants (n=115)		Participants under 65 years (n=874)		Participants 65 years and over (n=127)	
	n (%)	95% CI	n (%)	95% CI	n (%)	95% CI	n (%)	95% CI
Humans only	156 (17.8)	15.4-20.2	26 (22.6)	20.0-25.2	159 (18.2)	15.8-20.6	25 (19.7)	17.2-22.2
Both humans and the app	324 (37.0)	34.0-40.0	51 (44.3)	41.2-47.4	343 (39.2)	36.2-42.2	36 (28.3)	25.5-31.1
App only	130 (14.8)	12.6-17.0	6 (5.2)	3.8-6.6	116 (13.3)	11.2-15.4	20 (15.7)	13.4-18.0

[a]BAME: Black, Asian, and minority ethnic.

BAME participants were more concerned about how their data would be used ($mean_{White}$ 3.07, SD 1.232; $mean_{BAME}$ 3.58, SD 1.091; *U*=31052.5; *P*=.001), felt more strongly that they had no choice but to download the app ($mean_{White}$ 2.95, SD 1.267; $mean_{BAME}$ 3.59, SD 1.065; *U*=20857.0; *P*<.001), and felt frustrated as a result of a notification from the app ($mean_{White}$ 2.59, SD 1.263; $mean_{BAME}$ 3.42, SD 1.219; *U*=21961.5; *P*<.001). They felt less strongly that the app was easy to use ($mean_{White}$ 4.11, SD 0.871; $mean_{BAME}$ 3.75, SD 1.039; *U*=12844.5; *P*=.004) or that it was useful to wider society ($mean_{White}$ 4.05, SD 0.885; $mean_{BAME}$ 3.75, SD 0.960; *U*=21371.5; *P*=.01). Participants 65 years and over were less concerned about how their data would be used ($mean_{\geq 65}$ 2.72, SD 1.269; $mean_{<65}$ 3.20, SD 1.216; *U*=20663.0; *P*=.009), less likely to feel that they had no choice but to download the app ($mean_{\geq 65}$ 2.68, SD 1.215; $mean_{<65}$ 3.07, SD 1.260; *U*=14010.0; *P*=.01), and less likely to feel frustrated as a result of a notification from the app ($mean_{\geq 65}$ 2.09, SD 1.129; $mean_{<65}$ 2.77, SD 1.283; *U*=11854.5; *P*<.001). It was also less important to them that they could opt in and opt out of contact tracing ($mean_{\geq 65}$ 3.18, SD 1.119; $mean_{<65}$ 3.56, SD 1.101; *U*=19243.0; *P*<.001). There were no other significant differences in attitudes.

BAME participants felt it was less important that they trusted the app ($mean_{White}$ 4.04, SD 0.898; $mean_{BAME}$ 3.79, SD 0.996; *U*=43247.5; *P*=.008). They had more trust in the big tech companies ($mean_{White}$ 3.14, SD 1.106; $mean_{BAME}$ 3.43, SD 1.109; *U*=57731.5; *P*=.008) and private contractors ($mean_{White}$ 2.78, SD 1.099; $mean_{BAME}$ 3.24, SD 1.081; *U*=61597.5; *P*<.001), but less trust in the NHS ($mean_{White}$ 4.12, SD 0.954; $mean_{BAME}$ 3.97, SD 0.912; *U*=44705.5; *P*=.04). Conversely, participants 65 years and over had less trust in the big tech companies ($mean_{\geq 65}$ 2.97, SD 1.133; $mean_{<65}$ 3.21, SD 1.100; *U*=49432.0; *P*=.04) and private contractors ($mean_{\geq 65}$ 2.53, SD 1.133; $mean_{<65}$ 2.88, SD 1.092; *U*=46020.0; *P*=.001), but more trust in the UK government ($mean_{\geq 65}$ 3.24, SD 1.269; $mean_{<65}$ 2.97, SD 0.945; *U*=62275.5; *P*=.02). There were no other significant differences in trust.

Discussion

Principal Findings and Comparison With Prior Work

Just over half of those surveyed had downloaded the app, agreeing with other estimates for the United Kingdom [29]. Reasons for app uptake were predominantly to help the NHS, protect others, and reduce the spread of the virus, broadly agreeing with previous research [2,4]. Older adults had more community-minded and altruistic attitudes, being more likely to download the app to help the NHS, friends, family, and society, but they also had less intention to download the app. However, almost 1 in 9 of those who initially downloaded the app eventually deleted it, especially BAME participants, although the number of these participants who had downloaded and kept the app was similar to that of White participants. However, this increased deletion is a particular concern, as their vulnerability is reflected in the finding that twice as many BAME participants had tested positive or had a household member test positive for COVID-19.

In line with previous studies [4,10,13-15], reasons for not downloading or deleting the app were related to not wanting to be tracked, a feeling that it would be ineffective, and a lack of trust in the people who built the app. People who decided to delete the app were more likely than those who kept it to feel that it was not useful or easy to use, as expected from the TAM2 [32]; they also felt they understood less about how it worked, were more concerned about how their data would be used, and were more likely to have been frustrated by a notification from the app. This shows how important a trustworthy user experience

is for the adoption of contact tracing apps. BAME participants who had the app also had higher levels of concern about their data, felt the app was less easy to use and was less useful to society, and were more likely to have been frustrated by the app; this could lead to them deleting the app in the future. Those involved in the design of contact tracing should pay particular attention to the needs of BAME app users. On the other hand, older adults had less concern about their data and less frustration. Engaging users in the development, implementation, and evaluation of contact tracing can help to maximize engagement and technology acceptance, according to the RRI framework [7,8], helping designers to consider wider social implications of a technology and how real-world use might differ from usage within a trial or with a prototype. Working together with users to anticipate concerns and develop solutions can be an effective mechanism to achieve the adoption of digital solutions.

The feeling that there was a lack of choice in using the app was stronger among BAME participants and lowest among older adults; BAME participants were also more likely to state that they had to download the app for their job. Participants required a level of control over the app, feeling that it was important to get explanations, verify and speak to people about notifications, and be able to opt out of contact tracing; interestingly, the latter was less important for older adults. Most people stated that they would be or had been highly compliant with advice to self-isolate, although intention was significantly higher than actual reported compliance, especially among older adults. Actual reported compliance was similar across the different groups, although our findings suggest that White people tended to overstate their intention to comply. This finding may be impacted by the reduced trust in the government, as previously discussed, as well as a perceived lack of incentives offered for compliance [33]. It is also interesting that when asked how much they complied with "official advice," some participants who had been asked to self-isolate, by the app or otherwise, occasionally answered that they had not received any official advice. This implies that the instruction to self-isolate was perceived as being a suggestion rather than holding any authority, which may, in turn, have led to participants not taking the same precautions as they would have if they felt that the instruction to self-isolate was an official request.

Trust in the app was moderate. Participants who did not download the app had significantly lower trust in the app, especially in whether their data would be deleted and whether the app was reliable. They also had significantly lower trust in other users and in stakeholders surrounding the app, suggesting that trust is an important consideration in the design of contact tracing apps. Trust in the government was particularly low and may be a factor in adoption of app-based contact tracing [2,4,34,35], although it was higher among older adults, who were more concerned about big tech and private contractors. The BAME population had more trust in the big tech companies and private contractors but less trust in the NHS. BAME participants were also significantly less likely to download the app to help the NHS. A recent study on health-related quality of life revealed inequalities within English ethnic minorities [36], including poor primary care experiences, inadequate support from local services, and low patient self-confidence. This indicates that government slogans like "protect the NHS" may not have the intended effect on BAME communities and should be rethought to be more inclusive. However, reasons for a lack of trust on the part of BAME people in the governments and institutions in the United Kingdom are likely linked to persistent issues of structural racism [37] and, thus, are unlikely to be changed through singular measures alone, such as government messaging.

Finally, although participants felt they understood how the app worked, the results show that most people do not know that decisions about notifications are entirely made by the app [27,29], without human involvement. Understanding was particularly low in BAME communities. This lack of understanding may affect uptake and continued use, as it may negatively impact trust and, consequently, the app's popularity, perceived validity, and reliability [38]. At the same time, perceived human intervention may falsely increase trust in the app, as a completely automated system is likely to be recognized as having an unfair impact on the population, limiting freedom without taking into consideration personal circumstances [39].

Limitations and Future Work

While the sample was representative of the UK population in terms of age, gender, region, and ethnicity, some demographics that were not measured, such as income and political leaning, may have affected the results. Additionally, this representativeness means that the sample sizes for the vulnerable populations, although proportionate to population, were small compared to the overall sample. Future work should consider enriching the sample with greater numbers of participants from minority populations in order to capture their views more thoroughly. Additionally, while the sample was drawn from an online panel, this bias toward the online population was considered acceptable for this study, as the focus was on use of a smartphone app, which implied internet access. Recent estimates suggest that 92% of adults in the United Kingdom are recent internet users, including 54% of those over 75 years of age and 81% of disabled adults; only 6.3% of adults had never used the internet [40]. However, this does mean that potential respondents who do not have access to, or the ability to use, the internet—as were individuals who were not part of the online panel—were unable to take part in this study. While out of scope for this paper, this should be examined in future studies, as such issues could disproportionately affect vulnerable communities.

Future work should also consider multivariable analyses to account for the demographics of participants, to aid in explaining the differences found between vulnerable subgroups. For example, it is possible that lower trust in the government or the NHS might be driven by factors such as age, gender, or education. Similarly, further investigation of other groups could be beneficial, for example, to test the effects of having tested positive on their opinions or behaviors. There was a slightly higher proportion of participants who had tested positive than the national proportion (4% vs 3%), and quite a few participants reported they had been otherwise affected by the virus; it is possible that some participants were drawn to the study for this reason. However, this group was not excluded or highlighted

in this paper due to the overall low numbers of self-reported positive cases among respondents.

Finally, as with all self-report studies, in addition to the potential oversubscription of closely affected participants, there is a possibility of other reporting biases in this study; for example, social desirability bias and overreporting of compliance to self-isolation advice. However, as described above, a nontrivial number of participants did report not complying at all or not intending to comply, although it does seem there is a tendency to overstate intentions. It would be interesting to relate this to actual recorded behavior with regard to self-isolation.

Conclusions

This paper adds to the existing evidence surrounding digital contact tracing by reporting on an investigation of acceptance of a live app, which had been available to download for almost 3 months at the time of the study. Based on the TAM with the added factor of trust, an online survey was carried out looking at use of and attitudes toward the United Kingdom's NHS Test and Trace app, NHS COVID-19, among a representative sample of participants, including subgroup analysis of participants 65 years and over and members of the BAME community as potentially vulnerable users. Results indicate that uptake was limited to around 1 in 2 persons. Stated reasons for adoption predominantly surrounded a desire to help the NHS, friends and family, and society, especially among older adults, although BAME respondents agreed significantly less that their reason was to help the NHS. However, of those with the app, only one-fifth understood that the decision to send self-isolation notifications was made by the app without human involvement; in addition, there were a range of significantly more negative views among BAME participants. Respondents without the app reported significantly lower trust and more negative views toward the app. In cohort with other studies, the evidence shows that there are considerable barriers to the uptake of digital contact tracing apps, and these differ across different populations. It is important to consider especially potentially vulnerable groups to ensure that interventions such as these are effective. Potential users must be engaged in the design to enhance uptake and acceptance of test and trace apps, focusing particularly on groups that might be hard to reach or may have different attitudes toward acceptance.

Acknowledgments

This research was supported by the Engineering and Physical Sciences Research Council (grants EP/V00784X/1, EP/M02315X/1, and EP/T022493/1). EPV also acknowledges the resources of the National Institute for Health Research, Nottingham Biomedical Research Centre. We are also grateful for the support of Ipsos MORI in conducting this study. While the questionnaire was designed by the authors, Ipsos MORI was contracted to recruit participants and carry out the survey.

Authors' Contributions

All authors were involved in the conceptualization and the methodology of the study. LD conducted the formal analysis of the data for the study, conducted the visualization of results, and wrote the original draft. LD and JF performed data curation and verified the underlying data. JF and DM supported funding acquisition for the study and conducted project administration. Investigation was carried out by Ipsos MORI, who also verified the underlying data. LD, JF, EPV, VP, MG, and EN contributed to reviewing and editing of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Demographics and additional statistics from the study.
[DOCX File , 51 KB - jmir_v23i9e29085_app1.docx]

References

1. Braithwaite I, Callender T, Bullock M, Aldridge RW. Automated and partly automated contact tracing: A systematic review to inform the control of COVID-19. Lancet Digit Health 2020 Nov;2(11):e607-e621 [FREE Full text] [doi: 10.1016/S2589-7500(20)30184-9] [Medline: 32839755]
2. Milsom L, Abeler J, Altmann S, Toussaert S, Zillessen H, Blasone R. Survey of acceptability of app-based contact tracing in the UK, US, France, Germany, and Italy. OSF. 2020 Mar 26. URL: https://osf.io/7vgq9/ [accessed 2021-07-08]
3. Bachtiger P, Adamson A, Quint JK, Peters NS. Belief of having had unconfirmed Covid-19 infection reduces willingness to participate in app-based contact tracing. NPJ Digit Med 2020 Nov 06;3(1):146 [FREE Full text] [doi: 10.1038/s41746-020-00357-5] [Medline: 33299071]
4. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of app-based contact tracing for COVID-19: Cross-country survey study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]

5. Hinch R, Probert W, Nurtay A, Kendall M, Wymant C, Hall M, et al. Effective Configurations of a Digital Contact Tracing App: A Report to NHSX. 2020 Apr 16. URL: https://cdn.theconversation.com/static_files/files/1009/Report_-_Effective_App_Configurations.pdf?1587531217 [accessed 2021-07-08]
6. Trang S, Trenz M, Weiger WH, Tarafdar M, Cheung CM. One app to trace them all? Examining app specifications for mass acceptance of contact-tracing apps. Eur J Inf Syst 2020 Jul 27;29(4):415-428. [doi: 10.1080/0960085x.2020.1784046]
7. Stilgoe J, Owen R, Macnaghten P. Developing a framework for responsible innovation. Res Policy 2013 Nov;42(9):1568-1580. [doi: 10.1016/j.respol.2013.05.008]
8. von Schomberg R. Towards Responsible Research and Innovation in the Information and Communication Technologies and Security Technologies Fields. Luxembourg: Publications Office of the European Union; 2011. URL: http://op.europa.eu/en/publication-detail/-/publication/60153e8a-0fe9-4911-a7f4-1b530967ef10 [accessed 2021-07-08]
9. Velicia-Martin F, Cabrera-Sanchez J, Gil-Cordero E, Palos-Sanchez P. Researching COVID-19 tracing app acceptance: Incorporating theory from the technological acceptance model. PeerJ Comput Sci 2021;7:e316 [FREE Full text] [doi: 10.7717/peerj-cs.316] [Medline: 33816983]
10. Kaspar K. Motivations for social distancing and app use as complementary measures to combat the COVID-19 pandemic: Quantitative survey study. J Med Internet Res 2020 Aug 27;22(8):e21613 [FREE Full text] [doi: 10.2196/21613] [Medline: 32759100]
11. Bengio Y, Janda R, Yu YW, Ippolito D, Jarvie M, Pilat D, et al. The need for privacy with public digital contact tracing during the COVID-19 pandemic. Lancet Digit Health 2020 Jul;2(7):e342-e344 [FREE Full text] [doi: 10.1016/S2589-7500(20)30133-3] [Medline: 32835192]
12. von Wyl V, Höglinger M, Sieber C, Kaufmann M, Moser A, Serra-Burriel M, et al. Drivers of acceptance of COVID-19 proximity tracing apps in Switzerland: Panel survey analysis. JMIR Public Health Surveill 2021 Jan 06;7(1):e25701 [FREE Full text] [doi: 10.2196/25701] [Medline: 33326411]
13. Jonker M, de Bekker-Grob E, Veldwijk J, Goossens L, Bour S, Rutten-Van Mölken M. COVID-19 contact tracing apps: Predicted uptake in the Netherlands based on a discrete choice experiment. JMIR Mhealth Uhealth 2020 Oct 09;8(10):e20741 [FREE Full text] [doi: 10.2196/20741] [Medline: 32795998]
14. O'Callaghan ME, Buckley J, Fitzgerald B, Johnson K, Laffey J, McNicholas B, et al. A national survey of attitudes to COVID-19 digital contact tracing in the Republic of Ireland. Ir J Med Sci 2021 Aug 16;190(3):863-887 [FREE Full text] [doi: 10.1007/s11845-020-02389-y] [Medline: 33063226]
15. Thomas R, Michaleff ZA, Greenwood H, Abukmail E, Glasziou P. Concerns and misconceptions about the Australian Government's COVIDSafe app: Cross-sectional survey study. JMIR Public Health Surveill 2020 Nov 04;6(4):e23081 [FREE Full text] [doi: 10.2196/23081] [Medline: 33048826]
16. Armitage R, Nellums LB. COVID-19 and the consequences of isolating the elderly. Lancet Public Health 2020 May;5(5):e256 [FREE Full text] [doi: 10.1016/S2468-2667(20)30061-X] [Medline: 32199471]
17. Daoust JF. Elderly people and responses to COVID-19 in 27 countries. PLoS One 2020 Jul 2;15(7):e0235590 [FREE Full text] [doi: 10.1371/journal.pone.0235590] [Medline: 32614889]
18. Perrotta F, Corbi G, Mazzeo G, Boccia M, Aronne L, D'Agnano V, et al. COVID-19 and the elderly: Insights into pathogenesis and clinical decision-making. Aging Clin Exp Res 2020 Aug;32(8):1599-1608 [FREE Full text] [doi: 10.1007/s40520-020-01631-y] [Medline: 32557332]
19. Aldridge RW, Lewer D, Katikireddi SV, Mathur R, Pathak N, Burns R, et al. Black, Asian and minority ethnic groups in England are at increased risk of death from COVID-19: Indirect standardisation of NHS mortality data. Wellcome Open Res 2020;5:88 [FREE Full text] [doi: 10.12688/wellcomeopenres.15922.2] [Medline: 32613083]
20. Ravi K. Ethnic disparities in COVID-19 mortality: Are comorbidities to blame? Lancet 2020 Jul 04;396(10243):22 [FREE Full text] [doi: 10.1016/S0140-6736(20)31423-9] [Medline: 32569579]
21. Khunti K, Singh AK, Pareek M, Hanif W. Is ethnicity linked to incidence or outcomes of COVID-19? BMJ 2020 Apr 20;369:m1548. [doi: 10.1136/bmj.m1548] [Medline: 32312785]
22. Hu Y. Intersecting ethnic and native-migrant inequalities in the economic impact of the COVID-19 pandemic in the UK. Res Soc Stratif Mobil 2020 Aug;68:100528 [FREE Full text] [doi: 10.1016/j.rssm.2020.100528] [Medline: 32834346]
23. Robertson E, Reeve KS, Niedzwiedz CL, Moore J, Blake M, Green M, et al. Predictors of COVID-19 vaccine hesitancy in the UK household longitudinal study. Brain Behav Immun 2021 May;94:41-50 [FREE Full text] [doi: 10.1016/j.bbi.2021.03.008] [Medline: 33713824]
24. Dickerson J, Lockyer B, Moss RH, Endacott C, Kelly B, Bridges S, et al. COVID-19 vaccine hesitancy in an ethnically diverse community: Descriptive findings from the Born in Bradford study [version 1; peer review: 1 approved, 1 approved with reservations]. Wellcome Open Res 2021 Feb 4;6:23 (forthcoming) [FREE Full text] [doi: 10.12688/wellcomeopenres.16576.1]
25. Thorneloe R, Wilcockson H, Lamb M, Jordan C, Arden M. Willingness to receive a COVID-19 vaccine among adults at high-risk of COVID-19: A UK-wide survey. PsyArXiv Preprint posted online on July 20, 2020. [FREE Full text] [doi: 10.31234/osf.io/fs9wk]
26. Cases in United Kingdom. GOV.UK Coronavirus (COVID-19) in the UK. 2021. URL: https://coronavirus.data.gov.uk/details/cases [accessed 2021-05-28]

27. Risk-scoring algorithm: Technical information. NHS COVID-19 App Support. 2020. URL: https://covid19.nhs.uk/risk-scoring-algorithm.html [accessed 2021-09-08]
28. The code behind the NHS COVID-19 app. NHSX. 2020 May 08. URL: https://www.nhsx.nhs.uk/blogs/code-behind-nhs-covid-19-app/ [accessed 2021-07-08]
29. Briers M, Holmes C, Fraser C. Demonstrating the impact of the NHS COVID-19 app. The Alan Turing Institute. 2021 Feb 09. URL: https://www.turing.ac.uk/blog/demonstrating-impact-nhs-covid-19-app [accessed 2021-07-08]
30. Department of Health and Social Care. NHS COVID-19 app alerts 17 million contacts to stop spread of COVID-19. GOV.UK. 2021 Feb 09. URL: https://www.gov.uk/government/news/nhs-covid-19-app-alerts-17-million-contacts-to-stop-spread-of-covid-19 [accessed 2021-07-08]
31. Wymant C, Ferretti L, Tsallis D, Charalambides M, Abeler-Dörner L, Bonsall D, et al. The epidemiological impact of the NHS COVID-19 app. Nature 2021 Jun;594(7863):408-412. [doi: 10.1038/s41586-021-03606-z] [Medline: 33979832]
32. Venkatesh V, Davis FD. A theoretical extension of the Technology Acceptance Model: Four longitudinal field studies. Manage Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
33. Bodas M, Peleg K. Self-isolation compliance in the COVID-19 era influenced by compensation: Findings from a recent survey in Israel. Health Aff (Millwood) 2020 Jun;39(6):936-941. [doi: 10.1377/hlthaff.2020.00382] [Medline: 32271627]
34. Bunker D. Who do you trust? The digital destruction of shared situational awareness and the COVID-19 infodemic. Int J Inf Manage 2020 Dec;55:102201 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102201] [Medline: 32836649]
35. Guillon M, Kergall P. Attitudes and opinions on quarantine and support for a contact-tracing application in France during the COVID-19 outbreak. Public Health 2020 Nov;188:21-31 [FREE Full text] [doi: 10.1016/j.puhe.2020.08.026] [Medline: 33059232]
36. Watkinson RE, Sutton M, Turner AJ. Ethnic inequalities in health-related quality of life among older adults in England: Secondary analysis of a national cross-sectional survey. Lancet Public Health 2021 Mar;6(3):e145-e154 [FREE Full text] [doi: 10.1016/S2468-2667(20)30287-5] [Medline: 33516278]
37. Clementi R. Structural racism remains a primary public health risk amidst COVID and beyond in the United Kingdom. J Public Health (Oxf) 2020 Aug 18;42(3):479. [doi: 10.1093/pubmed/fdaa102] [Medline: 32648928]
38. Vokinger KN, Nittas V, Witt CM, Fabrikant SI, von Wyl V. Digital health and the COVID-19 epidemic: An assessment framework for apps from an epidemiological and legal perspective. Swiss Med Wkly 2020 May 04;150:w20282 [FREE Full text] [doi: 10.4414/smw.2020.20282] [Medline: 32418194]
39. Klenk M, Duijf H. Ethics of digital contact tracing and COVID-19: Who is (not) free to go? Ethics Inf Technol 2020 Aug 24:1-9 [FREE Full text] [doi: 10.1007/s10676-020-09544-0] [Medline: 32863740]
40. Internet users, UK: 2020. Office for National Statistics. 2021 Apr 06. URL: https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2020 [accessed 2021-07-08]

Abbreviations

BAME: Black, Asian, and minority ethnic
NHS: National Health Service
RRI: Responsible Research and Innovation
TAM: Technology Acceptance Model
TAM2: extended Technology Acceptance Model

Edited by C Basch; submitted 25.03.21; peer-reviewed by V von Wyl, D Menges, C Jacob; comments to author 15.04.21; revised version received 10.06.21; accepted 09.07.21; published 17.09.21.

Please cite as:
Dowthwaite L, Fischer J, Perez Vallejos E, Portillo V, Nichele E, Goulden M, McAuley D
Public Adoption of and Trust in the NHS COVID-19 Contact Tracing App in the United Kingdom: Quantitative Online Survey Study
J Med Internet Res 2021;23(9):e29085
URL: https://www.jmir.org/2021/9/e29085
doi:10.2196/29085
PMID:34406960

Original Paper

Information Retrieval in an Infodemic: The Case of COVID-19 Publications

Douglas Teodoro[1,2,3*], PhD; Sohrab Ferdowsi[1,3*], PhD; Nikolay Borissov[4,5], PhD; Elham Kashani[6], MSc; David Vicente Alvarez[1], BSc; Jenny Copara[1,2,3], MSc; Racha Gouareb[1,3], PhD; Nona Naderi[1,2], PhD; Poorya Amini[4,5], PhD

[1]Geneva School of Business Administration, HES-SO University of Applied Arts and Sciences of Western Switzerland, Carouge, Switzerland
[2]Swiss Institute of Bioinformatics, Lausanne, Switzerland
[3]Department of Radiology and Medical Informatics, University of Geneva, Geneva, Switzerland
[4]Risklick AG, Bern, Switzerland
[5]Clinical Trials Unit, Bern, Switzerland
[6]Institute of Pathology, University of Bern, Bern, Switzerland
[*]these authors contributed equally

Corresponding Author:
Douglas Teodoro, PhD
Department of Radiology and Medical Informatics
University of Geneva
Campus Biotech G6-N3 - Chemin des Mines 9
Geneva, 1202
Switzerland
Phone: 41 022 379 0225
Email: douglas.teodoro@unige.ch

Abstract

Background: The COVID-19 global health crisis has led to an exponential surge in published scientific literature. In an attempt to tackle the pandemic, extremely large COVID-19–related corpora are being created, sometimes with inaccurate information, which is no longer at scale of human analyses.

Objective: In the context of searching for scientific evidence in the deluge of COVID-19–related literature, we present an information retrieval methodology for effective identification of relevant sources to answer biomedical queries posed using natural language.

Methods: Our multistage retrieval methodology combines probabilistic weighting models and reranking algorithms based on deep neural architectures to boost the ranking of relevant documents. Similarity of COVID-19 queries is compared to documents, and a series of postprocessing methods is applied to the initial ranking list to improve the match between the query and the biomedical information source and boost the position of relevant documents.

Results: The methodology was evaluated in the context of the TREC-COVID challenge, achieving competitive results with the top-ranking teams participating in the competition. Particularly, the combination of bag-of-words and deep neural language models significantly outperformed an Okapi Best Match 25–based baseline, retrieving on average, 83% of relevant documents in the top 20.

Conclusions: These results indicate that multistage retrieval supported by deep learning could enhance identification of literature for COVID-19–related questions posed using natural language.

(J Med Internet Res 2021;23(9):e30161) doi:10.2196/30161

KEYWORDS

information retrieval; multistage retrieval; neural search; deep learning; COVID-19; coronavirus; infodemic; infodemiology; literature; online information

Introduction

Background

In parallel with its public health crisis with vast social and economic impacts, the COVID-19 pandemic has resulted in an explosive surge of activities within scientific communities and across many disciplines [1]. In turn, it has led to an overabundance of information online and offline — a phenomenon described as an infodemic [2-4] — with often negative impacts on the population [5]. Since early 2020 when the pandemic was officially announced, the number of publications related to COVID-19 has had exponential growth [6]. In addition to the volume and velocity of the generated data, the heterogeneity as a result of the typical variety of concept naming found in the biomedical field, spelling mistakes, and the different source types [7] make searching and discovery of relevant literature within the COVID-19 corpora an important challenge [2].

With the sheer quantity of COVID-19 information continuously produced, researchers, policy makers, journalists, and ordinary citizens, among others, are unable to keep up with the fast-evolving body of knowledge disseminated. As knowledge about the pandemic evolves, study results and conclusions may be improved, contradicted, or even proven wrong [3]. Combined with relentless media coverage and social media interactions, this fast-changing and massive amount of information leads to confusion and desensitization among audiences (eg, as in the case of school opening guidelines and mask-wearing and social distancing recommendations) [5,8]. They also fuel deliberate attempts to create information disorders, such as misinformation, disinformation, malinformation, and fake news [9], reducing the effectiveness of public health measures and endangering countries' ability to stop the pandemic, ultimately having a negative impact on live costs [10,11].

To support states and relevant actors of society to manage the COVID-19 infodemic, the World Health Organization (WHO) has published a framework containing 50 recommendations, of which more than 20% are related to strengthening the scanning, review, and verification of evidence and information [2]. To help actors involved with the pandemic find the most relevant information for their needs, effective information retrieval models for the COVID-19–related corpora became thus a prominent necessity [12]. The information retrieval community, in turn, has responded actively and quickly to this extraordinary situation and has been aiming to address these challenges. To foster research for the scientific communities involved with the pandemic, the COVID-19 Open Research Dataset (CORD-19) [13] collection was built to maintain all the related publications for the family of coronaviruses. This dataset helped research in various directions, and several tasks are built around it, including natural language processing (NLP)–related tasks, like question answering [14] and language model pretraining [15], and information retrieval challenges in Kaggle [16] as well as the TREC-COVID [17,18].

The TREC-COVID [18-20] challenge ran in 5 rounds, each asking for an incremental set of information needs to be retrieved from publications of the CORD-19 collection. In a TREC-COVID round, participants were asked to rank documents of the CORD-19 corpus in decreasing order of likelihood of containing answers to a set of query topics. At the end of the round, experts provided relevance judgments for the top-ranking documents submitted by different participants using a pooling strategy [21]. Although limited to the first several top submissions of the participating teams, these relevance judgments enable the evaluation of the different models and are valuable examples to train retrieval models for the subsequent rounds of the challenge.

To improve search and discovery of COVID-19 scientific literature, in this work we aimed to investigate an information retrieval model supported by deep language models to enhance findability of relevant documents in fast-evolving corpora.

More than 50 teams participated in the TREC-COVID challenge worldwide, developing new information retrieval and NLP methodologies to tackle this complex task [22-27]. Having participated in the TREC-COVID challenge, in this paper we detail our retrieval methodology, which brought us competitive results with the top-ranking teams. Particularly, we used a multistage retrieval pipeline, combining classic probabilistic weighting models with novel learning to rank approaches made by ensemble of deep masked language models. We present our results and analyze how the different components of the pipeline contribute to providing the best answers to the query topics.

Related Work

Two-Stage Information Retrieval

Currently, 2 main methodologies are used to rank documents in information retrieval systems: (1) the classic query-document probabilistic approaches, such as Okapi Best Match 25 (BM25) [28] and probabilistic language models [29], and (2) the learning-to-rank approaches, which usually postprocess results provided by classic systems to improve the original ranked list [30,31]. When there are sufficient training data (ie, queries with relevance judgments for the case of information retrieval), learning-to-rank models often outperform classic one-stage retrieval systems [30,32]. Nevertheless, empiric results have also shown that the reranking step may degrade the performance of the original rank [33]. Progress on learning-to-rank algorithms has been fostered thanks to the public release of annotated benchmark datasets, such as the LETOR [34] and Microsoft Machine Reading Comprehension (MS MARCO) [35].

Learning-to-rank approaches can be categorized into 3 main classes of algorithms — pointwise, pairwise, and listwise — based on whether they consider 1 document, a pair of documents, or the whole ranking list in the learning loss function, respectively [30-32,36]. Variations of these learning-to-rank algorithms are available based on neural networks [31,36] and other learning algorithms, such as boosting trees [37]. More recently, pointwise methods leveraging the power of neural-based masked language models have attracted great attention [38,39]. These learning-to-rank models use the query and document learning representations provided by the masked language model to classify whether a document in the ranked list is relevant to the query. While these two-stage retrieval methods based on neural rerankers provide interesting

features, such as learned word proximity, in practice, the first stage based on classic probabilistic retrieval algorithms is indispensable, as the algorithmic complexity of the reranking methods makes them often prohibitive to classify the whole collection [32].

Recent advances in text analytics, including question answering, text classification, and information retrieval, have indeed mostly been driven by neural-based masked language models. A seminal effort in this direction is the Bidirectional Encoder Representations from Transformers (BERT) model [38], which shows significant success in a wide range of NLP tasks. BERT uses a bidirectional learning approach based on the transformer architecture [40] and is trained to predict masked words in context. Since the introduction of BERT, several works tried to augment its performance. A successful work in this direction is the robustly optimized BERT approach (RoBERTa) [41], using larger and more diverse corpora for training as well as a different tokenizer. While RoBERTa needs larger computing power, it often improves the performance of BERT across different downstream tasks. Another similar effort is the XLNet model [42], which uses a permutation-based masking, showing also consistent improvement over BERT.

TREC-COVID Retrieval Efforts

Recently, the specific case of retrieval of COVID-related scientific publications has been addressed in several efforts [22-27]. These works follow mostly the aforementioned two-stage retrieval process. Among the first efforts is the SLEDGE system [22], where the authors detailed their solutions for the first round of the TREC-COVID challenge using a BM25-based ranking method followed by a neural reranker. An important difficulty for the first round of the challenge is the absence of labelled data. To overcome this limitation, the authors lightly tuned the hyperparameters of the first-stage ranking model using minimal human judgments on a subset of the topics. As for the second stage, they used the SciBERT model [43], which is pretrained on biomedical texts, and fine-tuned on the general MS MARCO set [35] with a simple cross-entropy loss. CO-Search [24] uses a slightly different approach, wherein they incorporated semantic information, as captured by Sentence-BERT [44], also within the initial retrieval stage. Moreover, they used the citation information of publications in their ranking pipeline. In the work of Covidex [23], the authors provided a full-stack search engine implementing a multistage ranking pipeline, where their first stage is based on the Anserini information retrieval toolkit [45], complemented by different neural reranking strategies. They addressed the issue of length variability among documents with an atomic document representation using, for example, paragraph-level indexing.

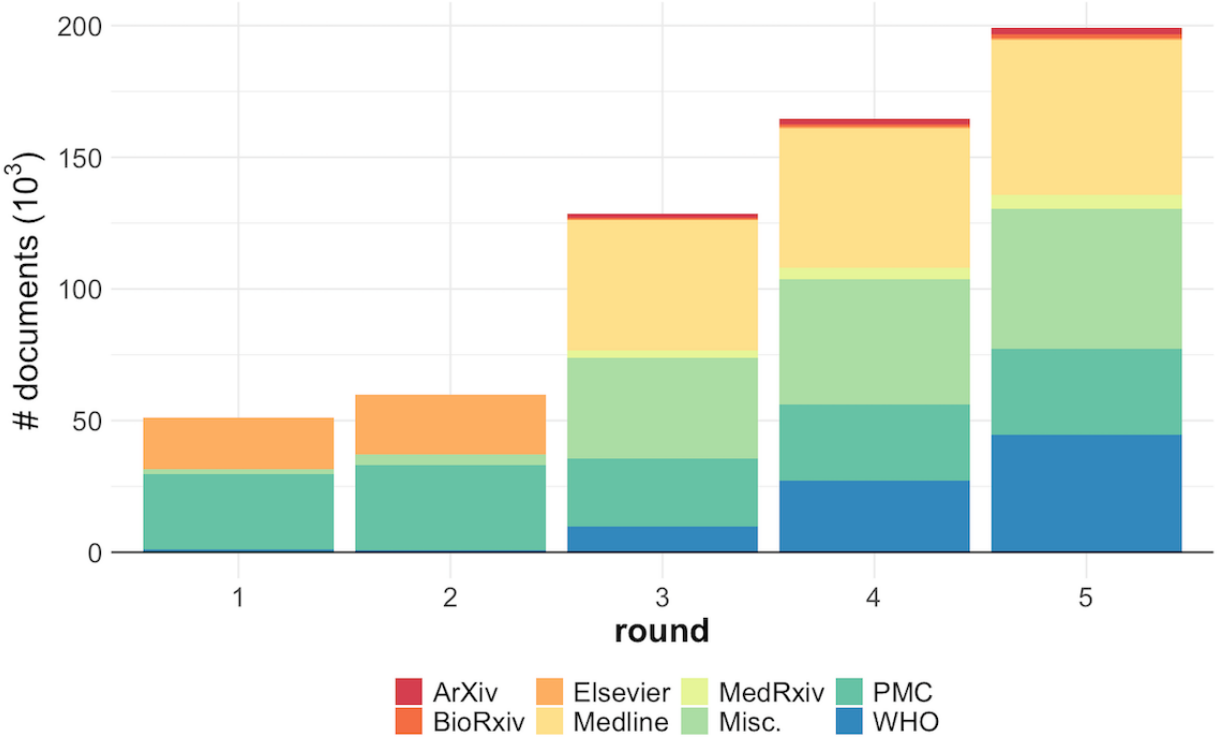
Methods

In this section, we describe the corpus and query set and our methodology for searching COVID-19–related literature in the context of the TREC-COVID challenge. We start by introducing the CORD-19 dataset, which is the corpus used in the competition. We then describe the challenge organization and assessment queries. Then, we detail our searching methodology, based on a multistage retrieval approach. Finally, we present the evaluation criteria used to score the participants' submissions. For further details on the TREC-COVID challenge, see [19,20].

The CORD-19 Dataset

A prominent effort to gather publications, preprints, and reports related to the coronaviruses and acute respiratory syndromes (COVID-19, Middle East respiratory syndrome [MERS], and severe acute respiratory syndrome [SARS]) is the CORD-19 collection of the Allen Institute for Artificial Intelligence (in collaboration with other partners) [13]. Figure 1 describes the size and content origin of the corpus for the different TREC-COVID rounds. As we can see, this is a large and dynamically growing semistructured dataset from various sources like PubMed, PubMed Central (PMC), WHO, and preprint servers like bioRxiv, medRxiv, and arXiv. The dataset contains document metadata, including title, abstract, and authors, among others, but also the full text or link to full-text files when available. A diverse set of related disciplines (eg, from virology and immunology to genetics) are represented in the collection. Throughout the challenge, the dataset was updated daily, and snapshot versions representing its status at a certain time were provided to the participants for each round. In the last round of the TREC-COVID challenge, the corpus contained around 200,000 documents, coming mostly from Medline, PMC, and WHO sources.

Figure 1. Evolution of the CORD-19 corpus across the TREC-COVID rounds stratified by source. PMC: PubMed Central; WHO: World Health Organization.



The TREC-COVID Challenge

To assess the different information retrieval models, the TREC-COVID challenge provided a query set capturing relevant search questions of researchers during the pandemic. These needs are stated in query topics, consisting of 3 free-text fields — query, question, and narrative — with an increasing level of context, as shown in the examples in Table 1. The challenge started with 30 topics in round 1 and added 5 new topics in each new round, thus reaching 50 topics in round 5.

In each round, the participants provided ranked lists of candidate publications of the CORD-19 collection that best answered the query topics. Each list was generated by a different information retrieval model, so called *run*, with up to 5 runs in the first 4 rounds and 7 runs in the last round per team. At the end of the round, domain experts examined the top k candidate publications (where k is defined by the organizers) from the priority runs of the teams and judged them as "highly relevant," "somehow relevant," or "irrelevant." Then, based on the consolidated relevance judgments, the participants were evaluated using standard information retrieval metrics (eg, normalized discounted cumulative gain [NDCG], precision). Judged documents for a specific topic from previous rounds were excluded from the relevance judgment list.

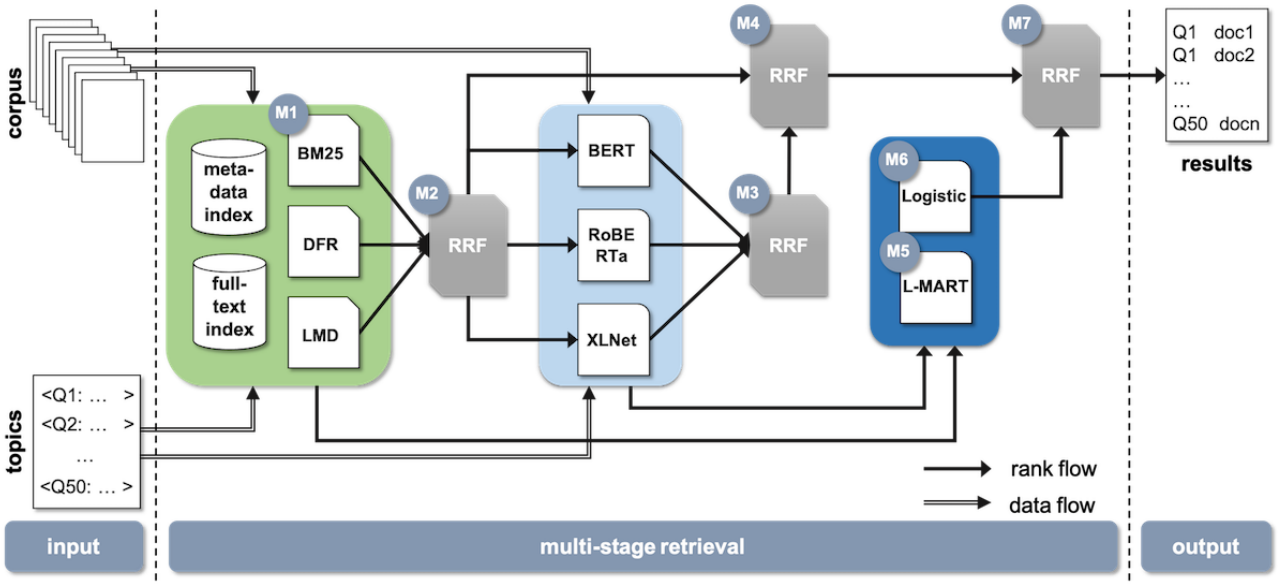
Table 1. Examples of TREC-COVID topics with the fields query, question, and narrative.

Topic	Query	Question	Narrative
1	Coronavirus origin	What is the origin of COVID-19?	Seeking a range of information about the SARS-CoV-2 virus's origin, including its evolution, animal source, and first transmission into humans
25	Coronavirus biomarkers	Which biomarkers predict the severe clinical course of 2019-nCOV infection?	Looking for information on biomarkers that predict disease outcomes in people infected with coronavirus, specifically those that predict severe and fatal outcomes
50	mRNA vaccine coronavirus	What is known about an mRNA vaccine for the SARS-CoV-2 virus?	Looking for studies specifically focusing on mRNA vaccines for COVID-19, including how mRNA vaccines work, why they are promising, and any results from actual clinical studies

Proposed Multistage Retrieval Methodology

Figure 2 shows the different components of our information retrieval pipeline for the COVID-related literature. These components can be divided into 3 main categories: (1) first-stage retrieval using classic probabilistic methods, (2) second-stage (neural) reranking models, and (3) rank fusion algorithms. Given a corpus containing metadata information, such as title and abstract, and full text, when available, documents are stored using directed and inverted indexes. Then, transformer-based and classic learning-to-rank models trained using relevance judgments are used to classify and rerank pairs of query-document answers. The ranked list obtained from the different models are further combined using the reciprocal rank fusion (RRF) algorithm.

Figure 2. Multistage retrieval pipeline, where light green indicates first-stage retrieval, light and dark blue indicate second-stage retrieval, and M1-M7 denote the different models used to create the respective runs 1-7 in round 5. BERT: Bidirectional Encoder Representations from Transformers; BM25: Okapi Best Match 25; DFR: divergence from randomness; L-MART: LambdaMART model; LMD: language model Dirichlet; Logistic: logistic regression model; RoBERTa: robustly optimized BERT approach; RRF: reciprocal rank fusion.



First-Stage Retrieval

For the first-stage retrieval, we assessed 3 variations of the classic query-document probabilistic weighting models: BM25 [46], divergence from randomness (DFR) [47], and language model Dirichlet (LMD) [48].

Our first classical model, Okapi BM25 [28], is based on the popular term frequency-inverse document frequency (tf-idf) framework. In the tf-idf framework, term-weights are calculated using the product of within term-frequency *tf* and the inverse document frequency *idf* statistics. Denote *f(t,d)* as the number of times a term *t* appears in a document *d* within a collection *D*, BM25 calculates the term-weight *w* as:

where $|d|$ is the length of the document, $|D|$ is the size of the collection, avg_l is the average length of the documents in the collection, n_t is the number of documents containing the term *t*, and k_1 and *b* are parameters of the model associated with the term frequency and the document size normalization, respectively.

The second model, DFR, extends the basic tf-idf concept by considering that the more the divergence of the term frequency *tf* from its collection frequency *cf* (*cf* ≈ *df*), the more the information carried by the term in the document [47]. Thus, for a given model of randomness *M*, in the DFR framework, the term-weight is inversely proportional to the probability of term-frequency within the document obtained by *M* for the collection *D*:

$$w(t,d,D) = k \cdot \log p_M (t \quad d|D),$$

where p_M is a probabilistic model, such as binomial or geometric distributions, and *k* is a parameter of the probabilistic model.

The third model, LMD, uses a language model that assigns probabilities to word sequences with a Dirichlet-prior smoothing to measure the similarity between a query and a document [48]. In a retrieval context, a language model specifies the probability that a document is generated by a query, and smoothing is used to avoid zero probabilities to unseen words and improves the overall word probability accuracy. In the LMD algorithm, term-weight is calculated using the following equation:

where $p(t|d)$ denotes the probability of a term in a document, $p(t|D)$ is the probability of a term in the collection, and μ is the Dirichlet parameter to control the amount of smoothing.

In our pipeline, the BM25, DFR, and LMD implementations are based on the Elasticsearch framework. The model parameters were trained using the queries and relevance judgments of round 4 in a 5-fold cross-validation setup.

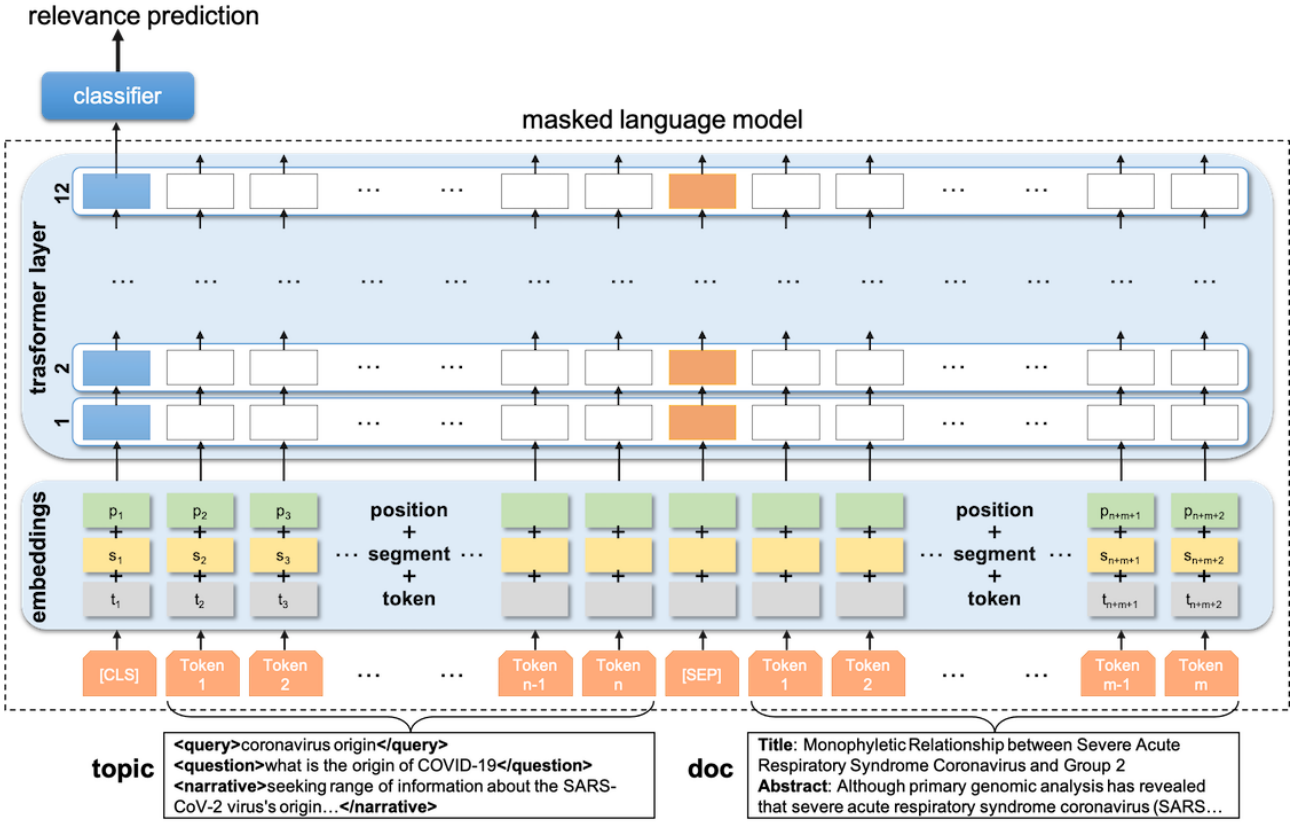
Second-Stage Reranking

The models used in the first-stage ranking were based on the bag-of-words statistics, where essentially we looked at the histogram of query terms and their document and collection statistics but neglected the sequential nature of text and word relations. To mitigate these limitations and improve the initial rankings, after the first-stage retrieval, we used neural masked language models trained on the relevance judgments from previous rounds so that syntactic and semantic relations can be better captured [49,50]. As shown in Figure 2, we assessed 3 masked language models based on the transformer architecture: BERT, RoBERTa, and XLNet.

Figure 3 shows the general idea of how we used the BERT language model to match documents to a query topic. Given a topic and a document associated with it as input and a relevance judgment as the label for the query-document association (relevant or not), the model was trained or fine-tuned in the

BERTology parlance, as it had been previously pretrained on a large corpus, to predict whether the document is relevant to the query. In the input layer of the pretrained model, the topic and candidate publication were tokenized and separated by the language model *[SEP]* token (stands for sentence separation). Moreover, to enforce the sequential structure of text, positional embedding as well as sentence embeddings were added to the main embeddings for each token. These embeddings were then fed to the transformer layers of BERT, which were updated during the fine-tuning step. Finally, the output of the special *[CLS]* token (stands for classification) was used to determine the relevance of the candidate publication to the queried information topic.

Figure 3. Neural masked language model for document relevance classification. As inputs to the pre-trained masked language model, the topics and candidate publications are separated by the [SEP] tag. Inputs are tokenized using subword tokenization methods (grey boxes); segment embeddings (yellow boxes) represent the difference between a topic and a document input; position embeddings (green boxes) enforce the sequential structure of text; the transformer and classification layers are updated in the training phase using the relevance judgments; and the output of the special [CLS] token is finally used to determine the relevance of the candidate publication to the queried information topic.



Using the query topics from a preceding round (round 4 for the results presented here) and their respective list of relevance judgments, we fine-tuned the BERT model to rescore the initial association of the query-document pair between (not relevant) and (very relevant). For this, we used the score associated with the *[CLS]* token position. We limited the input size of the query and document to 512 tokens (or subwords). Then, at the second-stage re-ranking step, we classified the top *k* publications retrieved by the first stage models using the fine-tuned BERT model (we set *k*=5000 in our experiments).

Identical training strategies were used for the RoBERTa and XLNet language models. The main difference for the RoBERTa model is that it was originally pretrained on a corpus with an order of magnitude bigger than that of BERT (160 GB vs 16 GB). Moreover, it uses dynamic masking during the training process, that is, at each training epoch, the model sees different versions of the same sentence with masks on different positions, compared to a static mask algorithm for BERT. Last, RoBERTa uses a byte-level Byte-Pair-Encoding tokenizer compared to BERT's WordPiece. As BERT and its variants (eg, RoBERTa) neglect the dependency between the masked positions and suffer from a pretrain-finetune discrepancy, XLNet adopts a permutation language model instead of masked language model to solve the discrepancy problem. For downstream tasks, the fine-tuning procedure of XLNet is similar to that of BERT and RoBERTa.

We used the BERT (base - 12 layers), RoBERTa, and XLNet model implementations available from the Hugging Face framework. The models were trained using the Adam optimizer [51] with an initial learning rate of 1.5e–5, weight decay of 0.01, and early stopping with a patience of 5 epochs.

Combining Model Results

We used the RRF algorithm [52] to combine the results of different retrieval runs. RRF is a simple, yet powerful technique to rescore a retrieval list based on the scores of multiple retrieval lists. Given a set of documents D to be sorted and a set of ranking files $R = \{r_1 \ldots r_n\}$, each with a permutation on $1 \ldots |D|$, RRF computes the aggregated score using the following equation:

where $r(q,d)$ is the rank of document d for the query q in the ranking file r_i and k is a threshold parameter, which was tuned to k=60 using data from previous rounds.

Second-Step Learning-to-Rank

To exploit the features (relevance score) created by the different bag-of-words and masked language models, we added a second-step learning-to-rank pass to our pipeline. Using the similarity scores s computed by the BM25, DFR, LMD, BERT, RoBERTa, and XLNet as input features and the relevance judgments of previous rounds as labels, we trained 2 learning-to-rank models: LambdaMART and a logistic regression classifier. While the language models exploit the sequential nature of text, they completely neglect the ranking provided by the bag-of-words models. Thus, we investigated the use of the LambdaMART [31] algorithm, which uses a pairwise loss that compares pairs of documents and tells which document is better in the given pair. Moreover, we trained a simple pointwise logistic regression to consider the similarity measures computed by the first- and second-stage retrieval models. We used the pyltr and scikit-learn implementations for the LambdaMART and logistic regression, respectively. For the LambdaMART model, we trained the learning rate and the number of estimators, and for the logistic regression model, we trained the solver and regularization strength parameters.

First-Stage Retrieval: Preprocessing, Querying Strategies, and Parameter Tuning

In the first-stage retrieval step, we applied a classical NLP preprocessing pipeline to the publications (indexing phase) and topics (search phase): lower-casing, removal of nonalphanumerical characters (apart from "-"), and Porter stemming. Additionally, a minimal set of COVID-related synonyms, such as "covid-19" and "sars-cov-2," were created and used for query expansion.

The queries were then submitted to the index in a combinatorial way using the different topic fields and document sections. This means that, for each *query*, *question*, and *narrative* field of a topic, we submitted a query against the index for each of the *title* and *abstract* sections of the publications (abstract + full text in case of the full-text index). Additionally, the whole topic (query + question + narrative) was queried against the whole document. This querying strategy led to 7 queries for each topic, and the final score was computed by summing up the individual scores. Moreover, as the first public announcement of a coronavirus-related pneumonia was made in January 2020, we filtered out all publications before December 2019.

We defined the best query strategy and fine-tuned the basic parameters of the bag-of-words models using the relevance judgments of the previous round in a 5-fold cross-validation approach. As an example, to tune the b and k parameters of the BM25 model in round 5, we took the topics and relevance judgment from round 4 and submitted them to the index in round 5, optimizing the P@10 metric. For round 1, we used default parameter values.

Evaluation Criteria

We used the official metrics of the TREC-COVID challenge to report our results: precision at K documents (P@K), NDCG at K documents (NDCG@K), mean average precision (MAP), and binary preference (Bpref) [19]. For all these metrics, the closest the result is to 1, the better the retrieval model. They were obtained using the *trec_eval* information retrieval evaluation toolkit.

Results

We used 7 models from our pipeline to create the 7 runs submitted for the official evaluation of the TREC-COVID challenge (labels M1 to M7 in Figure 2). Our first model — *bm25* — based on the BM25 weighting model against the metadata index provided the baseline run. Our second model — *bow + rrf* — was a fusion of the BM25, DFR, and LMD weighting models computed against the metadata and full-text indices and combined using the RRF algorithm. Model 3 — *mlm + rrf* — used the RRF combination of the BERT, RoBERTa, and XLNet models applied to the top 5000 documents retrieved by model 2. Model 4 — *bow + mlm + rrf* — combined the results of models 2 and 3 using the RRF algorithm. Then, model 5 — *bow + mlm + lm* — reranked the results of runs 2 and 3 using the LambdaMART algorithm trained using the similarity scores of the individual models 2 and 3. Similarly, model 6 — *bow + mlm + lr* — was based on a logistic regression classifier that uses as features the similarity scores of runs 2 and 3 to classify the relevance of the query-document pairs. Finally, model 7 — *bow + mlm + lr + rrf* — combined runs 2, 3, and 6 using the RRF algorithm. For all RRF combinations, the parameter k was set to 60. All models and parameters were trained using round 4 relevance judgments. Table 2 summarizes the submitted runs.

Table 2. Summary of the submitted runs. Refer to Figure 2 for a pictorial description.

Run	Name	Description
1	bm25	Run based on the baseline BM25[a] model using the metadata index
2	bow + rrf	An RRF[b] combination of BM25, DFR[c], and LMD[d] models computed against the metadata and full-text indices
3	mlm + rrf	An RRF combination of BERT, RoBERTa[e], and XLNet models applied to run 2
4	bow + mlm + rrf	An RRF combination of runs 2 and 3
5	bow + mlm + lm	A LambdaMART-based model using features from the individual models used to create runs 2 and 3
6	bow + mlm + lr	A logistic regression model using features from the individual models used to create runs 2 and 3
7	bow + mlm + lr + rrf	An RRF combination of runs 2, 3, and 6

[a]BM25: Okapi Best Match 25.

[b]RRF: reciprocal rank fusion.

[c]DFR: divergence from randomness.

[d]LMD: language model Dirichlet.

[e]RoBERTa: robustly optimized BERT approach.

Official Evaluation Results

Table 3 shows the official results of the TREC-COVID challenge for the 7 submitted runs. As we can see, the best results are provided by model 7 (bow + mlm + lr + rrf), apart from the metric Bpref, which is the highest for model 5 (bow + mlm + lm). Comparing the NDCG@20 metric, model 7 improved 16.4 percentage points against the baseline model (26.0% relative improvement). On average, almost 17 of the top 20 documents retrieved by model 7 were pertinent to the query. Model 3 was able to retrieve 6.6% more relevant documents compared to the baseline model (6963 vs 6533 of a total of 10,910 documents judged relevant for the 50 queries). On the other hand, it showed a relative improvement in precision of 22.1% for the top 20 documents. Therefore, it not only improved the recall but also brought relevant documents higher in the ranking list. These results show that the use of the masked language models had a significant positive impact in the ranking.

Table 3. Performance of our models in round 5 of the TREC-COVID challenge.

Model	NDCG@20[a]	P@20[b]	Bpref[c]	MAP[d]	# rel[e]
bm25	0.6320	0.6440	0.5021	0.2707	6533
bow + rrf	0.6475	0.6650	0.5174	0.2778	6695
mlm + rrf	0.7716	0.7880	0.5680	0.3468	6963
bow + mlm + rrf	0.7826	0.8050	0.5616	0.3719	7006
bow + mlm + lm	0.7297	0.7460	0.5759	0.3068	6834
bow + mlm + lr	0.7375	0.7450	0.5719	0.3439	6976
bow + mlm + lr + rrf	0.7961	0.8260	0.5659	0.3789	6939

[a]NDCG@20: normalized discounted cumulative gain at 20 documents.

[b]P@20: precision at 20 documents.

[c]Bpref: binary preference.

[d]MAP: mean average precision.

[e]# ref: total number of relevant documents retrieved by the model for the 50 queries.

Table 4 shows the official best results for the different metrics for the top 10 teams participating in round 5 of TREC-COVID (NDCG@20 metric taken as reference). Comparing the NDCG@20 metric, the best model submitted by our team (risklick) was ranked 4 of the 28 teams participating in round 5, 5.4 percentage points below the top-performing team (Unique-ptr). For reference, the best-performing model in the challenge retrieves on average 17.5 relevant documents per query in the top 20 retrieved documents compared to 16.5 for our model. If we consider a reference baseline made by the median of the participating teams' best values, our pipeline outperforms the baseline by 11.7%, 14.6%, 16.7%, and 25.0% for the MAP, P@20, NDCG@20, and Bpref metrics, respectively. All data and results of the TREC-COVID challenge can be found here: [17].

Table 4. Official leaderboard of the top 10 teams in the final round of the TREC-COVID challenge.

Team	NDCG@20[a]	P@20[b]	Bpref[c]	MAP[d]
unique_ptr	0.8496	0.8760	0.6378	0.4731
covidex	0.8311	0.8460	0.5330	0.3922
Elhuyar_NLP_team	0.8100	0.8340	0.6284	0.4169
risklick (ours)	0.7961	0.8260	0.5759	0.3789
udel_fang	0.7930	0.8270	0.5555	0.3682
CIR	0.7921	0.8320	0.5735	0.3983
uogTr	0.7921	0.8420	0.5709	0.3901
UCD_CS	0.7859	0.8440	0.4488	0.3348
sabir	0.7789	0.8210	0.6078	0.4061
mpiid5	0.7759	0.8110	0.5873	0.3903

[a]NDCG@20: normalized discounted cumulative gain at 20 documents.

[b]P@20: precision at 20 documents.

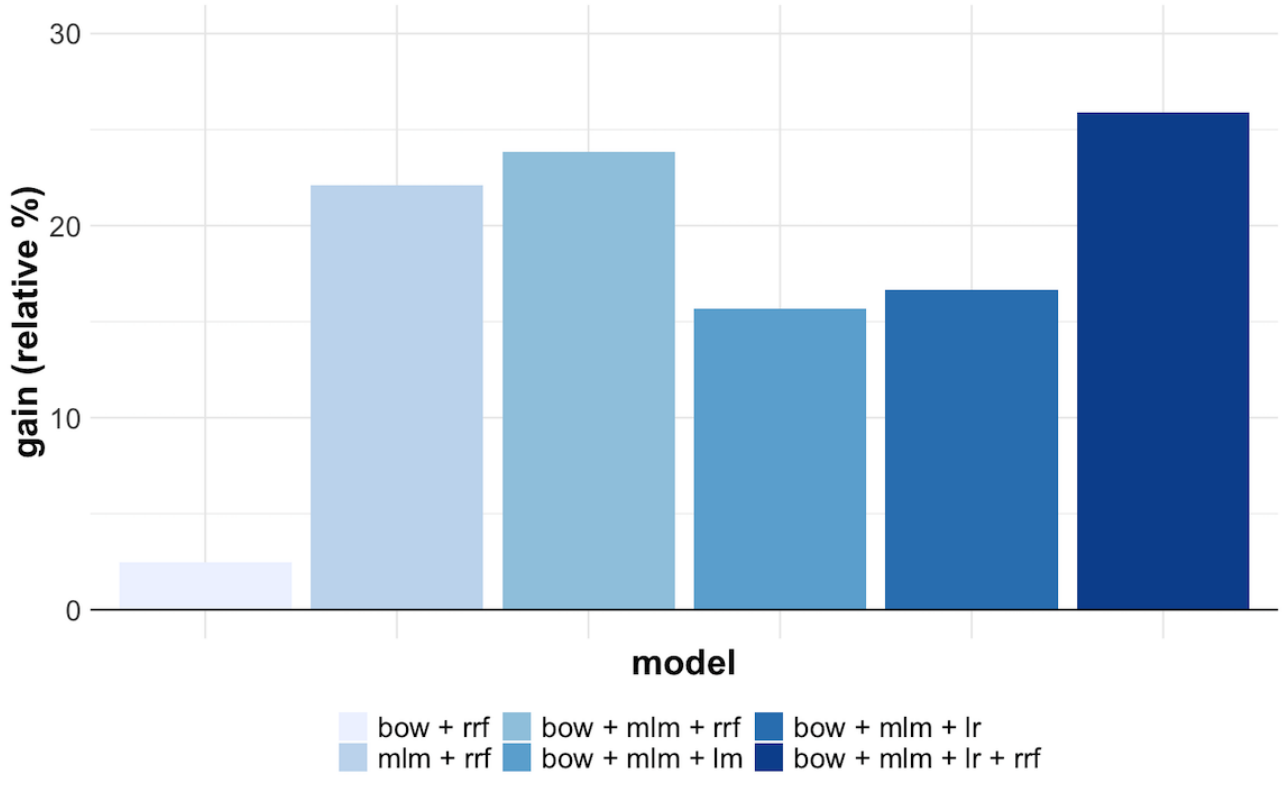
[c]Bpref: binary preference.

[d]MAP: mean average precision.

Model Performance Analyses

Figure 4 shows the relative improvement of the different models in the pipeline in relation to the baseline (model 1 – bm25) according to the NDCG@20 metric. The most significant contribution to the final performance came from the inclusion of the masked language models in the pipeline — model 3: mlm + rrf and model 4: bow + mlm + rrf — adding a relative performance gain to the results of 22.1% and 23.8%, respectively. The classic learning-to-rank models — model 5 and model 6 — actually jeopardized the performance when compared to their previous model in the pipeline (model 4). However, when model 6 was combined with model 4, a 2.1 percentage point gain was achieved on top of model 4, leading to the best model (model 7: bow + mlm + lr + rrf). Indeed, it is important to notice the consistent benefit of combining models using the RRF algorithm. Interestingly, the effect of LambdaMART seemed to be significantly detrimental for P@20, NDCG@20, and MAP, but marginally beneficial for Bpref, for which it is the best model.

Figure 4. Relative contribution of each model for the normalized discounted cumulative gain at document 20 (NDCG@20) metric compared to the baseline model bm25.



The performance of the individual masked language models is shown in Table 5. Surprisingly, they are similar to the baseline model, with small performance reductions for BERT and RoBERTa models and a small performance gain for the XLNet model. However, when combined, they provide the significant performance improvement shown in Figure 4. Our assumption is that they retrieve different documents as relevant and their combination using RRF ends up aligning these documents in the top rank. Indeed, looking at the top 3 documents for query 1 retrieved by these models, for example, there is no overlap between the documents, with 8 relevant and 1 unjudged (out of the 9 documents). This result clearly shows the beneficial effect of using an ensemble of masked language models, as well as the success of RRF in fusing their retrievals.

Table 5. Performance of the individual masked language models and their combination using reciprocal rank fusion (RRF).

Model	NDCG@20[a]	P@20[b]	Bpref[c]	MAP[d]	# rel[e]
BERT[f]	0.6209	0.6430	0.5588	0.2897	6879
RoBERTa[g]	0.6261	0.6440	0.5530	0.2946	6945
XLNet	0.6436	0.6570	0.5644	0.3064	6926
mlm + rrf	0.7716	0.7880	0.5680	0.3468	6963

[a]NDCG@20: normalized discounted cumulative gain at 20 documents.

[b]P@20: precision at 20 documents.

[c]Bpref: binary preference.

[d]MAP: mean average precision.

[e]# ref: total number of relevant documents retrieved by the model for the 50 queries.

[f]BERT: Bidirectional Encoder Representations from Transformers.

[g]RoBERTa: robustly optimized BERT approach.

Topic Performance Analyses

The performance analyses for the individual topics shows that our best model had a median value of 0.9000 for the P@20 metric (max=1.0000, min=0.3000), which demonstrates successful overall performance. However, as shown in Figure 5, for some topics, notably 11, 12, 19, 33, and 50, less than 50% of documents in the top 20 retrieved are relevant. For topics 11, 12, and 19, which searched for “coronavirus hospital rationing,” “coronavirus quarantine,” and “what alcohol sanitizer kills coronavirus” information, respectively, all our models have poor performance, and indeed, the combination of the different models in the pipeline managed to boost the results. On the other hand, for topics 33 and 50, which searched for “coronavirus vaccine candidates” and “mRNA vaccine coronavirus” information, respectively, it was the combination with the logistic regression model that lowered the performance (notice in Figure 5 that model 4: bow + mlm + rrf has a significantly better performance compared to model 7 for those queries).

The difference in performance per topic between our best model and the median of the submitted runs in round 5 for all teams for the P@20 metric is shown in Figure 6. Indeed, topics 11, 12, and 19 seemed hard for all the models participating in the TREC-COVID challenge to retrieve the correct documents. Even if our best model had poor performance for those topics, it still outperformed most of the runs submitted to the official evaluation. In particular, topic 19 had only 9 relevant or somewhat relevant documents in the official relevance judgments, which means that its max performance can be at most around 50% for the P@20 metric. For our worst performing topics compared to the other participants — topics 33 and 50 — better tuning between the ranking weights of the bag-of-words, masked language, and logistic regression models could have boosted the results.

Figure 5. Per topic precision at rank 20 (P@20) in round 5 of TREC-COVID per each run. The baseline run1 and the best-performing run7, which benefits from neural language models, are highlighted with dashed lines. Note that for most topics, the transformer-based runs have significantly improved performance.

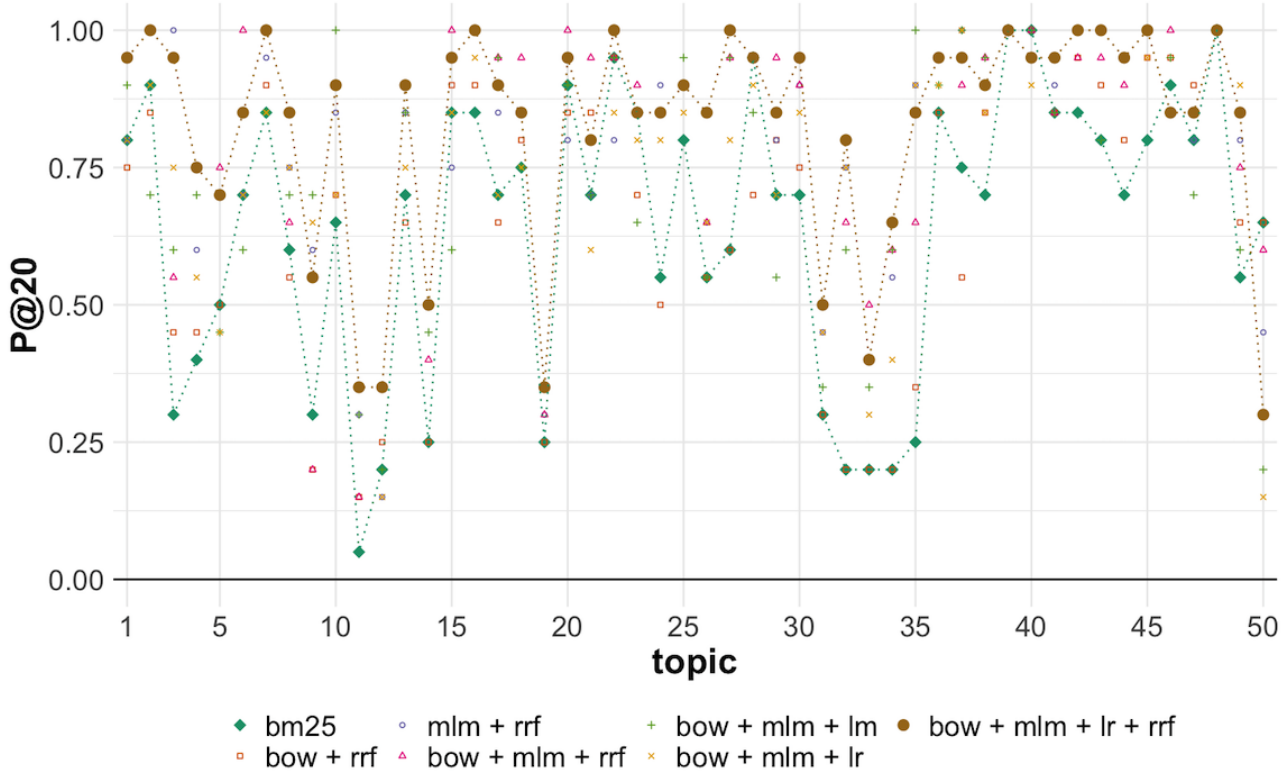
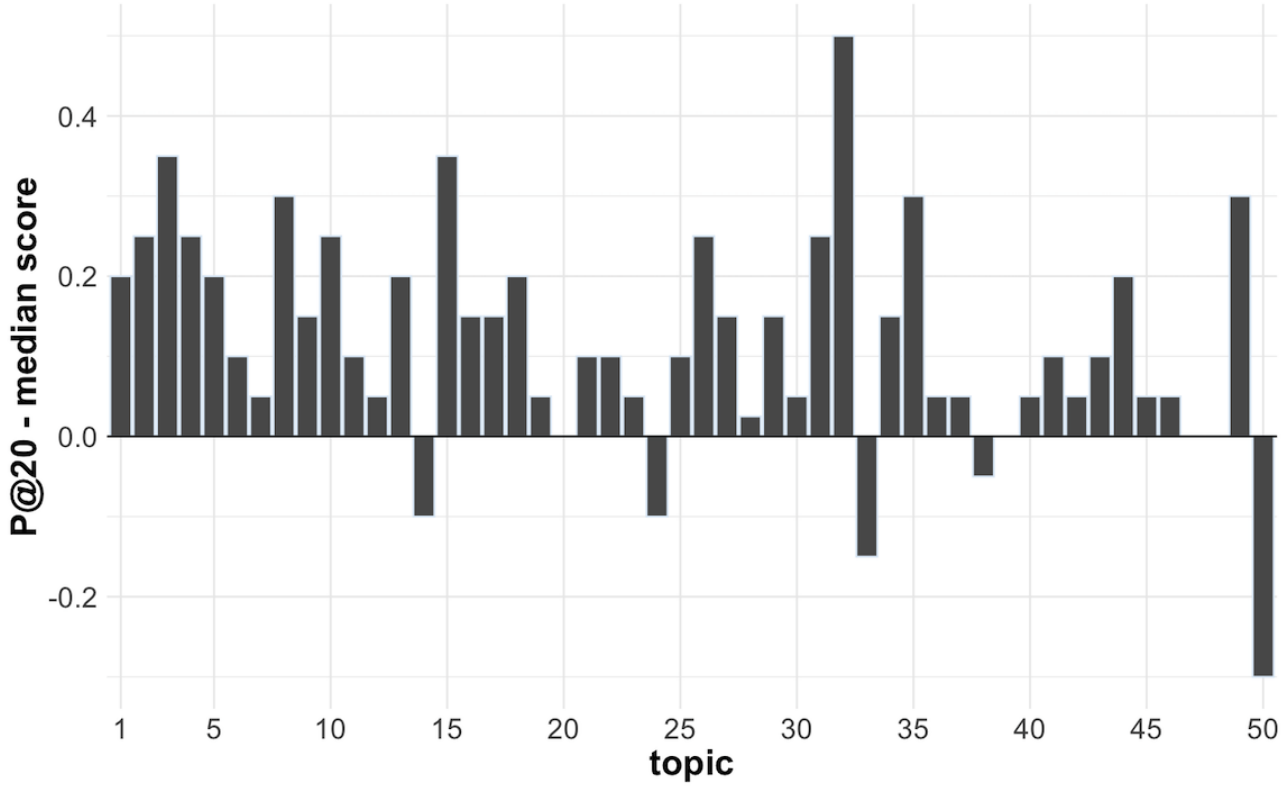


Figure 6. Per topic performance difference between our best model (model 7) and the median of all official submissions for the precision at document 20 (P@20) metric in round 5.

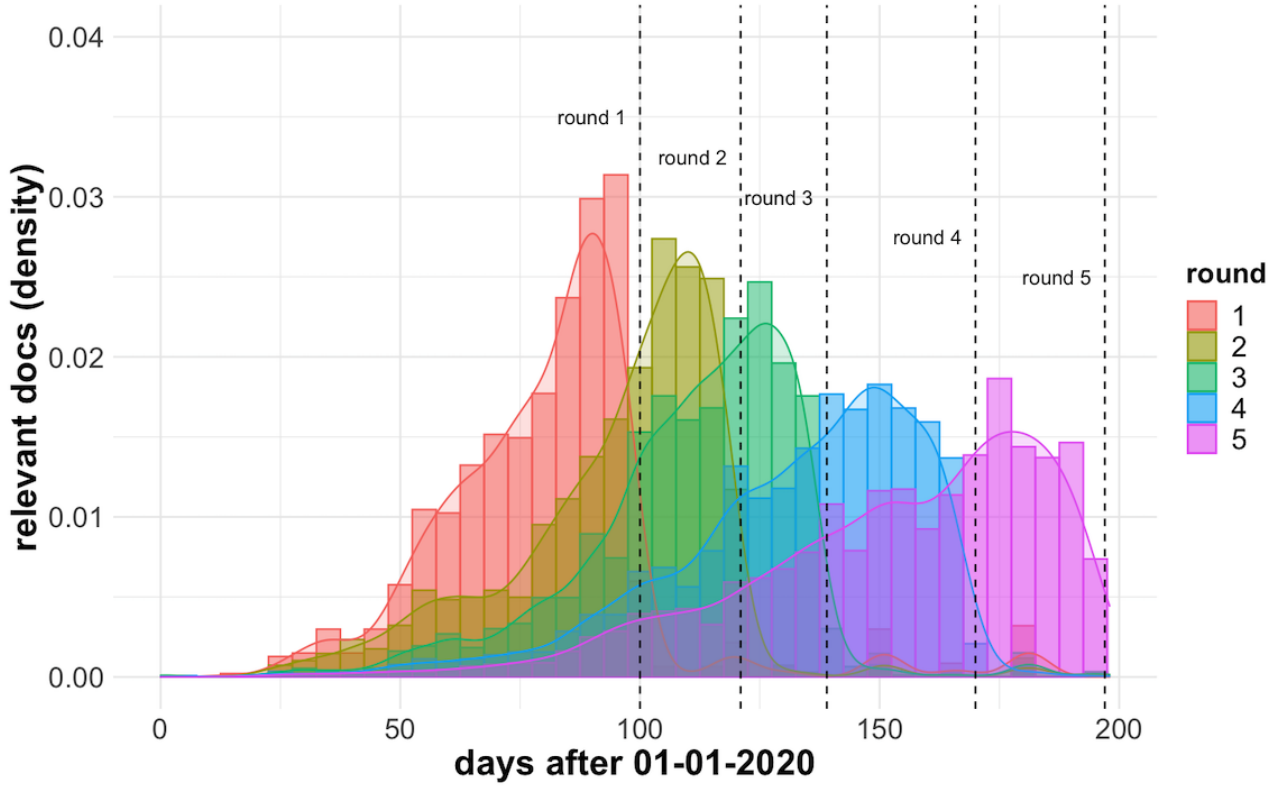


Time-Dependent Relevance Analyses

Given the dynamics of the COVID-19 pandemic, with a relatively well-defined starting period, a particularly effective technique to remove noise from the results, also adopted by some other participating teams [22], is filtering documents based on their publication dates. For our first-stage retrieval models, we filtered out publications before December 2019 when the outbreak was first detected. This led to a small negative impact on recall but highly improved the precision of our models.

To better understand how the document relevance varied over time, we analyzed the publication date of the official relevance judgments for the 5 rounds of TREC-COVID. As we can see in Figure 7, there is a clear exponential decay pattern in the number of relevant articles over time for all the rounds, with a faster decay in the first rounds and a longer tail for the later ones. We noticed that more recent publications closer to the round start, when the snapshot of the collection was created and queries were submitted, tended to have a higher probability of being relevant to the information need, with a half-life of around 20 days for round 1. This is somehow expected. First, as the documents found in previous query rounds were explored and are no longer relevant, only the most recent data are interesting, particularly in the gap between rounds. A second explanation is that in the case of a pandemic, new evidence arrives at an explosive rate, possibly refuting older knowledge.

Figure 7. Distribution of the publication dates of the "highly relevant" articles for each of the TREC-COVID rounds.



Discussion

To support effective search and discovery of COVID-19–related relevant literature in the COVID-19 infodemic, we explored the use of a multistage retrieval pipeline supported by bag-of-words models, masked language models, and classic learning-to-rank methods. The proposed methodology was evaluated in the context of the TREC-COVID challenge and achieved competitive results, being ranked in the top 4 of 126 runs among 28 teams participating in the challenge. The use of the multistage retrieval approach significantly improved the search results of COVID-related literature, leading to a gain in performance of 25.9% in terms of the NDCG@20 metric compared to a bag-of-words baseline. Particularly, the ensemble of masked language models brought the highest performance gain to the search pipeline. Indeed, ensembles of language models have proved to be a robust methodology to improve predictive performance [53-55].

The COVID-19 pandemic has led to a huge amount of literature being published in the most diverse sources, including scientific journals, grey repositories, and white reports, among others. As the pandemic continues, the number of scientific publications grows at an unprecedented rate causing an infodemic within many of the different disciplines involved [3]. Finding the most relevant information sources to answer different information needs within the huge volume of data created had become of utmost necessity [2]. By enabling the discovery of relevant information sources to complex user queries, effective retrieval models as proposed in this work may help to tackle the spread of misinformation. Such models empower experts with a minimal cost to search and discover information sources within a massive and fast-evolving corpus. Indeed, our model provides relevant information sources for more than 8 documents in the top-10 rank. Thus, continuous active search methods could be put in place to monitor and discover sources of evidence to certain query topics of relevant public health interest (eg, "coronavirus origin") in a timely manner. This in turn, would

enable experts to analyze, identify, and curate both sources of the best evidence at the time and sources of misinformation. The former would foster the creation among others of living systematic reviews [56,57], which is one of the recommendations of the WHO to tackle the COVID infodemic [2]. On the other hand, the latter could help fight, for example, the spread of scientific fake news by early retraction of misinforming articles, particularly in preprint servers, and thus limiting their exposition.

Looking at the boost in performance of model 3 (mlm + rrf) alone, one could be tempted to argue that masked language models could be the main component in a retrieval system. However, 2 issues may arise: algorithmic complexity and search effectiveness. The former is related to the high complexity of masked language models $(O(n^2 \cdot h)$, where n is the sentence length and h is the number of attention heads), which makes it prohibitive to classify a whole collection, often containing millions of documents, for every given query. The latter is related to the effectiveness of the individual models themselves. As shown in Table 5, individually, the performance of the language models is not significantly different from the baseline BM25 model. Thus, we believe it is the combination of models with different properties that can provide a successful search strategy in complex corpora, such as the one that originated from the COVID-19 infodemic.

In terms of practical implications, by effectively processing natural language, the methodology proposed can help biomedical researchers and clinicians to find the COVID-19 papers that they need. The efficient literature discovery process fostered by our methods may lead to faster publication cycles when required, for example reducing from weeks to days the drafting time of COVID-19 reviews [58], but also to less costly creation of curated living evidence portals, which will inform clinicians and public health officers with the best available evidence [59]. Indeed, as shown in [27,60], these methodologies outperform commercially available tools for searching and discovering COVID-19–related literature. Moreover, as they are data-driven, it is expected that they can be extrapolated to other types of corpora, such as clinical trial protocols and biomedical metadata datasets [60,61], enabling thus a more comprehensive identification of scientific evidence. Equally important, as the COVID-19 infodemic is not the first and unlikely the last [62,63], our methodology and findings could be extended to help tackling future epi-, pan-, and infodemics by supporting relevant actors to scan large and fast-changing collections to create timely reviews and curated evidence and apply localized infodemic management approaches.

With the rapid surge of published information and the variety of topics and sources related to COVID-19, it became hard for professionals dealing with the pandemic to find the correct information for their needs. While the automation discussed in this work can support more effective search and discovery, some high-level topics are still challenging. Indeed, some topics assessed in the TREC-COVID challenge were shown to be particularly hard for the retrieval models. For example, for topic 11, which searched for documents providing information on "guidelines for triaging patients infected with coronavirus," our best model prioritized documents providing information about indicators for diagnosis (eg, "early recognition of coronavirus," "RT-PCR testing of SARS-CoV-2 for hospitalized patients clinically diagnosed"). On the other hand, it missed documents including passages such as "telephone triage of patients with respiratory complaints." Similarly, for topic 12, which searched information about the "best practices in hospitals and at home in maintaining quarantine," our model prioritized documents providing information about "hospital preparedness" (eg, "improving preparedness for," "preparedness among hospitals") and missed documents containing information about "home-based exercise note in Covid-19 quarantine situation."

The methodology proposed has some limitations. First, it fails to explore transfer learning of learning-to-rank datasets. While the top-ranked teams all used multistage retrieval approaches, confirming the value of such methodology in modern retrieval models [18,23], the reranking strategy within the different pipelines varied slightly among the participants. For example, the top-ranked team used transfer learning from the MS MARCO learning-to-rank dataset and from a zero-shot learning approach. Other teams in the top 3 used transfer learning from a silver collection, based on the known item search technique [64]. Second, while we explored the combination of different topic items to build our queries, we failed to work on the document indexing unit, leaving all the normalization work to the probabilistic weighting models. As the COVID-19 literature comes from heterogeneous collections, containing sometimes only title and sometimes large full text, even with good finetuning of the model parameters, such variation in size and content poses a challenge to the first-stage retrieval model. Indeed, some strategies that explored decomposing the indexing unit into small structures, such as sentences and paragraphs, have achieved more competitive results [23].

Another limitation of our work was the ability to explore the freshness of the corpus. The TREC-COVID challenge dynamics, running throughout a sequence of rounds with new incremental search topics added on each round, provides an interesting setting for evaluating retrieval models in an infodemic context. It simulates typical search and discovery workflows, in which evolving queries are posed against an evolving body of knowledge over time, and already discovered documents in previous searches are no longer relevant [65,66]. A successful strategy in this case is to filter out results according to a cut-off date, thus reducing noise in the retrieval set. However, in retrospect, we noticed that another useful technique, which is very natural to an infodemic case, could be to decay the score of publications by their distance to the present time or explore their recency or freshness [67,68], as highlighted in Figure 7, rather than a hard cut-off (ie, December 2019 in our case) for all the rounds. We leave exploring such a strategy as future work.

To conclude, we believe our information retrieval pipeline can provide a potential solution to help researchers, decision makers, and medical doctors, among others, search and find the correct information in the unique situation caused by the COVID-19 pandemic. We detailed the different components of this pipeline, including the traditional index-based information retrieval methods and the modern NLP-based neural network models,

as well as insights and practical recipes to increase the quality of information retrieval of scientific publications targeted to the case of an infodemic. We grounded our results in the TREC-COVID challenge, where around 50 different teams participated in the 5 rounds of the competition. We showed very competitive results as judged by the official leaderboard of the challenge. Apart from the COVID-19 case, we believe our solutions can also be useful for other potential future infodemics.

Acknowledgments

The study received funding from Innosuisse project funding number 41013.1 IP-ICT. CINECA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825775.

Authors' Contributions

DT conceived the experiments. DT and SF conducted the experiments. DT, SF, EK, and PA analyzed the results. NB, EK, DVA, and PA prepared the data. SF, DT, JC, RG, and NN drafted the manuscript. All authors reviewed the manuscript.

Conflicts of Interest

PA and NB work for Risklick AG. The other authors declare no competing interests.

References

1. Haghani M, Bliemer MCJ, Goerlandt F, Li J. The scientific literature on Coronaviruses, COVID-19 and its associated safety-related research dimensions: A scientometric analysis and scoping review. Saf Sci 2020 Sep;129:104806 [FREE Full text] [doi: 10.1016/j.ssci.2020.104806] [Medline: 32382213]
2. Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia-Saiso S, et al. Framework for Managing the COVID-19 Infodemic: Methods and Results of an Online, Crowdsourced WHO Technical Consultation. J Med Internet Res 2020 Jun 26;22(6):e19659 [FREE Full text] [doi: 10.2196/19659] [Medline: 32558655]
3. Eysenbach G. How to Fight an Infodemic: The Four Pillars of Infodemic Management. J Med Internet Res 2020 Jun 29;22(6):e21820 [FREE Full text] [doi: 10.2196/21820] [Medline: 32589589]
4. Eysenbach G. Infodemiology: The epidemiology of (mis)information. Am J Med 2002 Dec 15;113(9):763-765. [doi: 10.1016/s0002-9343(02)01473-0] [Medline: 12517369]
5. Kristensen K, Lorenz E, May J, Strauss R. Exploring the use of web searches for risk communication during COVID-19 in Germany. Sci Rep 2021 Mar 19;11(1):6419 [FREE Full text] [doi: 10.1038/s41598-021-85873-4] [Medline: 33742054]
6. Palayew A, Norgaard O, Safreed-Harmon K, Andersen TH, Rasmussen LN, Lazarus JV. Pandemic publishing poses a new COVID-19 challenge. Nat Hum Behav 2020 Jul;4(7):666-669. [doi: 10.1038/s41562-020-0911-0] [Medline: 32576981]
7. Cheerkoot-Jalim S, Khedo KK. A systematic review of text mining approaches applied to various application areas in the biomedical domain. JKM 2020 Dec 21;25(3):642-668. [doi: 10.1108/jkm-09-2019-0524]
8. Massey D, Huang C, Lu Y, Cohen A, Oren Y, Moed T, et al. Engagement With COVID-19 Public Health Measures in the United States: A Cross-sectional Social Media Analysis from June to November 2020. J Med Internet Res 2021 Jun 21;23(6):e26655 [FREE Full text] [doi: 10.2196/26655] [Medline: 34086593]
9. Wardle C, Derakhshan H. Thinking about 'information disorder': formats of misinformation, disinformation, and mal-information. Journalism, 'Fake News' & Disinformation. 2018. URL: https://en.unesco.org/sites/default/files/f._jfnd_handbook_module_2.pdf [accessed 2021-08-26]
10. Barua Z, Barua S, Aktar S, Kabir N, Li M. Effects of misinformation on COVID-19 individual responses and recommendations for resilience of disastrous consequences of misinformation. Prog Disaster Sci 2020 Dec;8:100119 [FREE Full text] [doi: 10.1016/j.pdisas.2020.100119] [Medline: 34173443]
11. West JD, Bergstrom CT. Misinformation in and about science. Proc Natl Acad Sci U S A 2021 Apr 13;118(15):1 [FREE Full text] [doi: 10.1073/pnas.1912444117] [Medline: 33837146]
12. Zarocostas J. How to fight an infodemic. Lancet 2020 Feb 29;395(10225):676 [FREE Full text] [doi: 10.1016/S0140-6736(20)30461-X] [Medline: 32113495]
13. Wang LL, Lo K, Chandrasekhar Y, Reas R, Yang J, Burdick D, et al. CORD-19: The COVID-19 Open Research Dataset. Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020 2020:1 [FREE Full text]
14. Lee J, Yi S, Jeong M, Sung M, Yoon W, Choi Y, et al. Answering Questions on COVID-19 in Real-Time. Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020 2020:1 [FREE Full text] [doi: 10.18653/v1/2020.nlpcovid19-2.1]
15. deepset / covid_bert_base. Hugging Face. URL: https://huggingface.co/deepset/covid_bert_base [accessed 2021-08-26]
16. COVID-19 Open Research Dataset Challenge (CORD-19). Kaggle. URL: https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge [accessed 2021-08-26]
17. TREC-COVID. National Institutes of Standards and Technology. URL: https://ir.nist.gov/covidSubmit/index.html [accessed 2021-08-26]

18. Roberts K, Alam T, Bedrick S, Demner-Fushman D, Lo K, Soboroff I, et al. Searching for scientific evidence in a pandemic: An overview of TREC-COVID. J Biomed Inform 2021 Jul 08;121:103865 [FREE Full text] [doi: 10.1016/j.jbi.2021.103865] [Medline: 34245913]
19. Roberts K, Alam T, Bedrick S, Demner-Fushman D, Lo K, Soboroff I, et al. TREC-COVID: rationale and structure of an information retrieval shared task for COVID-19. J Am Med Inform Assoc 2020 Jul 01;27(9):1431-1436 [FREE Full text] [doi: 10.1093/jamia/ocaa091] [Medline: 32365190]
20. Voorhees E, Alam T, Bedrick S, Demner-Fushman D, Hersh WR, Lo K, et al. TREC-COVID. SIGIR Forum 2020 Jun;54(1):1-12. [doi: 10.1145/3451964.3451965]
21. Tonon A, Demartini G, Cudré-Mauroux P. Pooling-based continuous evaluation of information retrieval systems. Inf Retrieval J 2015 Sep 8;18(5):445-472. [doi: 10.1007/s10791-015-9266-y]
22. MacAvaney S, Cohan A, Goharian N. SLEDGE: A Simple Yet Effective Baseline for COVID-19 Scientific Knowledge Search. Cornell University. 2020 May 05. URL: http://arxiv.org/abs/2005.02365 [accessed 2021-08-26]
23. Zhang E, Gupta N, Tang R, Han X, Pradeep R, Lu K, et al. Covidex: Neural Ranking Models and Keyword Search Infrastructure for the COVID-19 Open Research Dataset. Proceedings of the First Workshop on Scholarly Document Processing 2020:31-41 [FREE Full text] [doi: 10.18653/v1/2020.sdp-1.5]
24. Esteva A, Kale A, Paulus R, Hashimoto K, Yin W, Radev D, et al. COVID-19 information retrieval with deep-learning based semantic search, question answering, and abstractive summarization. NPJ Digit Med 2021 Apr 12;4(1):68 [FREE Full text] [doi: 10.1038/s41746-021-00437-0] [Medline: 33846532]
25. Nguyen V, Rybinski M, Karimi S, Xing Z. Searching Scientific Literature for Answers on COVID-19 Questions. Cornell University. 2020. URL: http://arxiv.org/abs/2007.02492 [accessed 2021-08-26]
26. Li C, Yates A, MacAvaney S, He B, Sun Y. PARADE: Passage Representation Aggregation for Document Reranking. Cornell University. 2021. URL: http://arxiv.org/abs/2008.09093 [accessed 2021-08-26]
27. Soni S, Roberts K. An evaluation of two commercial deep learning-based information retrieval systems for COVID-19 literature. J Am Med Inform Assoc 2021 Jan 15;28(1):132-137 [FREE Full text] [doi: 10.1093/jamia/ocaa271] [Medline: 33197268]
28. Robertson S, Zaragoza H. The Probabilistic Relevance Framework: BM25 and Beyond. FNT in Information Retrieval 2009;3(4):333-389. [doi: 10.1561/1500000019]
29. Lafferty J, Zhai C. Document language models, query models, and risk minimization for information retrieval. 2001 Presented at: 24th annual international ACM SIGIR conference on Research and development in information retrieval; 2001; New Orleans, LA. [doi: 10.1145/383952.383970]
30. Joachims T, Li H, Liu T, Zhai C. Learning to rank for information retrieval (LR4IR 2007). SIGIR Forum 2007 Dec;41(2):58-62. [doi: 10.1145/1328964.1328974]
31. Burges CJC. From RankNet to LambdaRank to LambdaMART: An Overview. Microsoft. 2010. URL: https://www.microsoft.com/en-us/research/uploads/prod/2016/02/MSR-TR-2010-82.pdf [accessed 2021-08-26]
32. Craswell N, Mitra B, Yilmaz E, Campos D, Voorhees EM. Overview of the TREC 2019 deep learning track. Cornell University. 2020. URL: https://arxiv.org/abs/2003.07820 [accessed 2021-08-26]
33. Faessler E, Oleynik M, Hahn U. TREC. 2019. URL: https://trec.nist.gov/pubs/trec28/papers/julie-mug.PM.pdf [accessed 2021-08-26]
34. Qin T, Liu T, Xu J, Li H. LETOR: A benchmark collection for research on learning to rank for information retrieval. Inf Retrieval 2010 Jan 1;13(4):346-374. [doi: 10.1007/s10791-009-9123-y]
35. Nguyen T, Rosenberg M, Song X, Gao J, Tiwary S, Majumder R, et al. MS MARCO: A human generated machine reading comprehension dataset. Microsoft. 2016 Nov. URL: https://www.microsoft.com/en-us/research/publication/ms-marco-human-generated-machine-reading-comprehension-dataset/ [accessed 2021-08-26]
36. Cao Z, Qin T, Liu TY, Tsai MF, Li H. Learning to rank: from pairwise approach to listwise approach. 2007 Presented at: 24th international conference on Machine learning; June 20-24, 2007; Corvalis, OR. [doi: 10.1145/1273496.1273513]
37. Li P, Burges CJ, Wu Q. Mcrank: Learning to rank using multiple classification and gradient boosting. NeurIPS Proceedings. 2007. URL: https://papers.nips.cc/paper/2007/file/b86e8d03fe992d1b0e19656875ee557c-Paper.pdf [accessed 2021-08-26]
38. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) 2019:4171-4186. [doi: 10.18653/v1/N19-1423]
39. Yilmaz ZA, Wang S, Yang W, Zhang H, Lin J. Applying BERT to Document Retrieval with Birch. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations 2019:19-24 [FREE Full text] [doi: 10.18653/v1/d19-3004]
40. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Cornell University. 2017. URL: https://arxiv.org/abs/1706.03762 [accessed 2021-08-26]
41. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: A robustly optimized BERT pretraining approach. Cornell University. 2019. URL: https://arxiv.org/abs/1907.11692 [accessed 2021-08-26]

42. Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov R, Le QV. XLNet: Generalized autoregressive pretraining for language understanding. Cornell University. 2019. URL: https://arxiv.org/abs/1906.08237 [accessed 2021-08-26]
43. Beltagy I, Lo K, Cohan A. SciBERT: A Pretrained Language Model for Scientific Text. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) 2019:3615-3620 [FREE Full text] [doi: 10.18653/v1/d19-1371]
44. Reimers N, Gurevych I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) 2019:3982-3992 [FREE Full text] [doi: 10.18653/v1/d19-1410]
45. Yang P, Fang H, Lin J. Anserini: Enabling the Use of Lucene for Information Retrieval Research. SIGIR '17: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval 2017:1253-1256 [FREE Full text] [doi: 10.1145/3077136.3080721]
46. Robertson SE, Jones KS. Relevance weighting of search terms. J. Am. Soc. Inf. Sci 1976 May;27(3):129-146. [doi: 10.1002/asi.4630270302]
47. Amati G, Van Rijsbergen CJ. Probabilistic models of information retrieval based on measuring the divergence from randomness. ACM Trans. Inf. Syst 2002 Oct;20(4):357-389. [doi: 10.1145/582415.582416]
48. Zhai C, Lafferty J. A Study of Smoothing Methods for Language Models Applied to Ad Hoc Information Retrieval. SIGIR Forum 2017 Aug 02;51(2):268-276. [doi: 10.1145/3130348.3130377]
49. Hewitt J, Manning CD. A Structural Probe for Finding Syntax in Word Representations. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) 2019:4129-4138 [FREE Full text] [doi: 10.18653/v1/N19-1419]
50. Yenicelik D, Schmidt F, Kilcher Y. How does BERT capture semantics? A closer look at polysemous words. Proceedings of the Third BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP 2020:156-162 [FREE Full text] [doi: 10.18653/v1/2020.blackboxnlp-1.15]
51. Kingma DP, Ba J. Adam: A method for stochastic gradient descent. Cornell University. 2014. URL: https://arxiv.org/abs/1412.6980 [accessed 2021-08-26]
52. Cormack GV, Clarke CLA, Buettcher S. Reciprocal rank fusion outperforms condorcet and individual rank learning methods. SIGIR '09: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval 2009:758-759 [FREE Full text] [doi: 10.1145/1571941.1572114]
53. Knafou J, Naderi N, Copara J, Teodoro D, Ruch P. BiTeM at WNUT 2020 Shared Task-1: Named Entity Recognition over Wet Lab Protocols using an Ensemble of Contextual Language Models. Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020) 2020:305-313 [FREE Full text] [doi: 10.18653/v1/2020.wnut-1.40]
54. Copara J, Knafou J, Naderi N, Moro C, Ruch P, Teodoro D. Contextualized French Language Models for Biomedical Named Entity Recognition. Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Atelier DÉfi Fouille de Textes 2020:36-48 [FREE Full text] [doi: 10.18653/v1/2020.wnut-1.40]
55. Copara J, Naderi N, Knafou J, Ruch P, Teodoro D. Named Entity Recognition in Chemical Patents using Ensemble of Contextual Language Models. 2020 Presented at: Work Notes CLEF 2020 - Conf Labs Eval Forum; September 2020; Thessaloniki, Greece p. 22-25. [doi: 10.18653/v1/2020.wnut-1.40]
56. Siemieniuk RA, Bartoszko JJ, Ge L, Zeraatkar D, Izcovich A, Kum E, et al. Drug treatments for covid-19: living systematic review and network meta-analysis. BMJ 2020 Jul 30;370:m2980 [FREE Full text] [doi: 10.1136/bmj.m2980] [Medline: 32732190]
57. Buitrago-Garcia D, Egli-Gany D, Counotte MJ, Hossmann S, Imeri H, Ipekci AM, et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. PLoS Med 2020 Sep;17(9):e1003346 [FREE Full text] [doi: 10.1371/journal.pmed.1003346] [Medline: 32960881]
58. Hutson M. Artificial-intelligence tools aim to tame the coronavirus literature. Nature 2020 Jun 09:1. [doi: 10.1038/d41586-020-01733-7] [Medline: 34103725]
59. Vogel JP, Tendal B, Giles M, Whitehead C, Burton W, Chakraborty S, National COVID-19 Clinical Evidence Taskforce. Clinical care of pregnant and postpartum women with COVID-19: Living recommendations from the National COVID-19 Clinical Evidence Taskforce. Aust N Z J Obstet Gynaecol 2020 Dec;60(6):840-851 [FREE Full text] [doi: 10.1111/ajo.13270] [Medline: 33119139]
60. Haas Q, Alvarez DV, Borissov N, Ferdowsi S, von Meyenn L, Trelle S, et al. Utilizing Artificial Intelligence to Manage COVID-19 Scientific Evidence Torrent with Risklick AI: A Critical Tool for Pharmacology and Therapy Development. Pharmacology 2021;106(5-6):244-253 [FREE Full text] [doi: 10.1159/000515908] [Medline: 33910199]
61. Teodoro D, Mottin L, Gobeill J, Gaudinat A, Vachon T, Ruch P. Improving average ranking precision in user searches for biomedical research datasets. Database (Oxford) 2017 Jan 01:2017 [FREE Full text] [doi: 10.1093/database/bax083] [Medline: 29220475]
62. Rothkopf DJ. When the buzz bites back. Washington Post. 2003 May 11. URL: http://www1.udel.edu/globalagenda/2004/student/readings/infodemic.html [accessed 2021-08-26]

63. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
64. Ogilvie P, Callan J. Combining document representations for known-item search. SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval 2003:143-150 [FREE Full text] [doi: 10.1145/860435.860463]
65. Li X, Croft WB. Time-based language models. CIKM '03: Proceedings of the twelfth international conference on Information and knowledge management 2003:469-475 [FREE Full text] [doi: 10.1145/956863.956951]
66. Moulahi B, Tamine L, Yahia S. When time meets information retrieval: Past proposals, current plans and future trends. Journal of Information Science 2016 Jul 11;42(6):725-747. [doi: 10.1177/0165551515607277]
67. Dong A, Zhang R, Kolari P, Bai J, Diaz F, Chang Y, et al. Time is of the essence: improving recency ranking using Twitter data. WWW '10: Proceedings of the 19th international conference on World wide web 2010:331-340 [FREE Full text] [doi: 10.1145/1772690.1772725]
68. Amati G, Amodeo G, Gaibisso C. Survival analysis for freshness in microblogging search. CIKM '12: Proceedings of the 21st ACM international conference on Information and knowledge management 2012:2483-2486 [FREE Full text] [doi: 10.1145/2396761.2398672]

Abbreviations

BERT: Bidirectional Encoder Representations from Transformers
BM25: Okapi Best Match 25
Bpref: binary preference
CORD-19: COVID-19 Open Research Dataset
DFR: divergence from randomness
LMD: language model Dirichlet
MAP: mean average precision
MERS: Middle East respiratory syndrome
MS MARCO: Microsoft Machine Reading Comprehension
NDCG: normalized discounted cumulative gain
NLP: natural language processing
PMC: PubMed Central
RoBERTa: robustly optimized BERT approach
RRF: reciprocal rank fusion
SARS: severe acute respiratory syndrome
tf-idf: term frequency-inverse document frequency
WHO: World Health Organization

Edited by C Basch; submitted 03.05.21; peer-reviewed by M Lotto, L Guo; comments to author 01.07.21; revised version received 22.07.21; accepted 05.08.21; published 17.09.21.

Please cite as:
Teodoro D, Ferdowsi S, Borissov N, Kashani E, Vicente Alvarez D, Copara J, Gouareb R, Naderi N, Amini P
Information Retrieval in an Infodemic: The Case of COVID-19 Publications
J Med Internet Res 2021;23(9):e30161
URL: https://www.jmir.org/2021/9/e30161
doi:10.2196/30161
PMID:34375298

Original Paper

Severe Fatigue in Long COVID: Web-Based Quantitative Follow-up Study in Members of Online Long COVID Support Groups

Maarten Van Herck[1,2,3,4*], MSc; Yvonne M J Goërtz[2,3,4*], MSc; Sarah Houben-Wilke[2], PhD; Felipe V C Machado[2,3,4], MSc; Roy Meys[2,3,4], MSc; Jeannet M Delbressine[2], MSc; Anouk W Vaes[2], PhD; Chris Burtin[1], PhD; Rein Posthuma[2,3,4], MD; Frits M E Franssen[2,3,4], MD, PhD; Bita Hajian[2], MD, PhD; Herman Vijlbrief[5], MSc; Yvonne Spies[5], MSc; Alex J van 't Hul[6], PhD; Daisy J A Janssen[2,7], MD, PhD; Martijn A Spruit[2,3,4], Prof Dr

[1]REVAL Rehabilitation Research Center, BIOMED Research Institute, Faculty of Rehabilitation Sciences, Hasselt University, Diepenbeek, Belgium

[2]Department of Research and Development, Ciro, Horn, Netherlands

[3]Nutrim School of Nutrition and Translational Research in Metabolism, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands

[4]Department of Respiratory Medicine, Maastricht University Medical Centre, Maastricht, Netherlands

[5]Lung Foundation Netherlands, Amersfoort, Netherlands

[6]Department of Pulmonary Disease, Radboud University Medical Center, Nijmegen, Netherlands

[7]Department of Health Services Research, Care and Public Health Research Institute, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands

[*]these authors contributed equally

Corresponding Author:
Maarten Van Herck, MSc
REVAL Rehabilitation Research Center
BIOMED Research Institute
Faculty of Rehabilitation Sciences, Hasselt University
Agoralaan gebouw A
Diepenbeek, 3590
Belgium
Phone: 32 494758248
Email: maarten.vanherck@uhasselt.be

Abstract

Background: Fatigue is the most commonly reported symptom in patients with persistent complaints following COVID-19 (ie, long COVID). Longitudinal studies examining the intensity of fatigue and differentiating between physical and mental fatigue are lacking.

Objective: The objectives of this study were to (1) assess the severity of fatigue over time in members of online long COVID peer support groups, and (2) assess whether members of these groups experienced mental fatigue, physical fatigue, or both.

Methods: A 2-wave web-based follow-up study was conducted in members of online long COVID peer support groups with a confirmed diagnosis approximately 3 and 6 months after the onset of infectious symptoms. Demographics, COVID-19 diagnosis, received health care (from medical professionals or allied health care professionals), fatigue (Checklist Individual Strength–subscale subjective fatigue [CIS-Fatigue]; 8-56 points), and physical and mental fatigue (self-constructed questions; 3-21 points) were assessed. Higher scores indicated more severe fatigue. A CIS-Fatigue score ≥36 points was used to qualify patients as having severe fatigue.

Results: A total of 239 patients with polymerase chain reaction/computed tomography–confirmed COVID-19 completed the survey 10 weeks (SD 2) and 23 weeks (SD 2) after onset of infectious symptoms, respectively (T1 and T2). Of these 239 patients, 198 (82.8%) were women; 142 (59.4%) had no self-reported pre-existing comorbidities; 208 (87%) self-reported being in good health before contracting COVID-19; and 62 (25.9%) were hospitalized during acute infection. The median age was 50 years (IQR 39-56). The vast majority of patients had severe fatigue at T1 and T2 (n=204, 85.4%, and n=188, 78.7%, respectively). No significant differences were found in the prevalence of normal, mild, and severe fatigue between T1 and T2 (*P*=.12). The median

CIS-Fatigue score was 48 points (IQR 42-53) at T1, and it decreased from T1 to T2 (median change: –2 points, IQR –7 to 3; *P*<.001). At T1, a median physical fatigue score of 19 points (IQR 16-20) and a median mental fatigue score of 15 points (IQR 10-17) were reported; these scores were lower at T2 for physical but not for mental fatigue (median change for physical fatigue –1 point, IQR –3 to 0, *P*<.001; median change for mental fatigue 0 points, IQR –3 to 3, *P*=.52). At the time of completing the follow-up survey, 194/239 (81.2%) and 164/239 (68.6%) of all patients had received care from at least one medical professional and one allied health care professional, respectively.

Conclusions: Fatigue in members of online long COVID support groups with a confirmed COVID-19 diagnosis decreases from 10 to 23 weeks after onset of symptoms. Despite this, severe fatigue remains highly prevalent. Both physical and mental fatigue are present. It remains unclear whether and to what extent fatigue will resolve spontaneously in the longer term.

Trial Registration: Netherlands Trial Register NTR8705; https://www.trialregister.nl/trial/8705.

(J Med Internet Res 2021;23(9):e30274) doi:10.2196/30274

KEYWORDS

COVID-19; SARS-CoV-2; long COVID; post-COVID-19 syndrome; post-acute sequelae of COVID-19; fatigue; post-viral fatigue; pandemic; online health; mental health; online support

Introduction

As the current COVID-19 pandemic continues to evolve, its impact becomes apparent. Clinical studies of hospitalized, laboratory-confirmed patients have shown that the acute phase of COVID-19 is characterized by a large array of respiratory and non-respiratory symptoms [1]. Over time, it has become clear that not all previously hospitalized patients fully recover from these symptoms in the months after the infection [2,3]. In addition, nonhospitalized patients can present persistent complaints months after the onset of infection-related symptoms [4]. These long-lasting symptoms after COVID-19 are referred to as *long COVID* [5], "a condition whereby affected individuals do not recover for several weeks or months following the onset of symptoms suggestive of COVID-19" [6], and they have a major impact on patients' quality of life (QoL) [7,8], care dependency [9], work participation [10,11], day-to-day activities, and physical functioning [12-14].

Fatigue, defined as "a subjective, unpleasant symptom which incorporates total body feelings ranging from tiredness to exhaustion creating an unrelenting overall condition which interferes with individuals' ability to function to their normal capacity" [15], is the most commonly reported symptom in patients with long COVID [2-4,16]. Similarly, other infections, such as severe acute respiratory syndrome (SARS) [17,18], Middle East respiratory syndrome (MERS) [19], and Q fever [20] have previously been linked to long-term fatigue, often referred to as postviral fatigue syndrome. Existing literature suggests that fatigue has several clinical presentations. A common distinction is made between physical fatigue (ie, difficulty performing physical activities) and mental fatigue (ie, difficulties concentrating and performing cognitive tasks) [21].

To date, longitudinal studies that examine fatigue intensity in patients with long COVID are lacking. Moreover, it is not known whether patients experience mostly mental or physical fatigue during and after the infection. Therefore, the objectives of this study were to (1) assess the severity of fatigue over time in members of online long COVID peer support groups; and (2) assess whether members of online long COVID peer support groups experience mental fatigue, physical fatigue, or both. We hypothesized that fatigue would be common and persistent and that both physical and mental fatigue would be present in patients with long COVID.

Methods

Study Design and Participants

This study is a prospective web-based survey of members of two Facebook peer support groups for patients with long COVID in the Netherlands (approximately 11,000 members; [22]) and Flanders (Belgium, approximately 1200 members; [23]), and a panel of approximately 1200 people who registered at a website of the Netherlands Lung Foundation (*coronaplein* [24]), an online platform providing additional information, advice, and peer support. Note that these totals represent the number of members of each group at the period of data collection. Between June 4 and June 11, 2020 (the time point of completing the first survey [T1]), members were invited to complete a web-based survey. Participants who completed the first survey [4,8,9] and who agreed to be contacted for a follow-up study received a second survey between August 31 and September 8, 2020 (ie, approximately three months after the first survey; the time point of completing the second survey [T2]). Ethical approval for this study was waived by the medical ethics committee of Maastricht University because the Medical Research Involving Human Subjects Act (WMO) does not apply to this study (METC2020-1978 and METC2020-2254). The medical ethics committee of Hasselt University formally judged and approved the study (MEC2020/041). Digital informed consent was obtained twice from all respondents (at the start of each survey). Exclusion criteria were intensive care unit (ICU) admission during the acute phase of infection, an onset of symptoms before January 1, 2020, being in the acute phase of COVID-19 when answering the first survey (ie, onset of infectious symptoms less than 3 weeks before filling out the first survey [25]), or an incomplete survey. Cross-sectional and follow-up data from this study on persistent symptoms, QoL, care dependency, construct-validity of the post–COVID-19 functional status scale, and information and care needs of members of online long COVID peer support groups have been published before [4,8,9,11,26,27]. This 2-wave web-based follow-up study was registered at the Netherlands Trial Registry (NTR8705). The

Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist was used to guide reporting [28]. Of note: the current study focusses on patients with a confirmed COVID-19 diagnosis (ie, test-diagnosed cases). The results for the patients without a confirmed COVID-19 diagnosis are presented in Multimedia Appendix 1.

Assessment via Web-Based Surveys

The survey was developed in close collaboration with scientists, methodologists, health care professionals and COVID-19 patients from the long COVID peer support groups (the Netherlands and Flanders). It was digitalized by ASolutions [29] and was made available via their online platform. The survey consisted of general questions regarding demographics, clinical characteristics, and standardized questionnaires, including a fatigue questionnaire.

Demographic and Clinical Characteristics

Respondents received questions regarding demographical aspects such as gender, age, weight, height, educational level (low/medium/high; classification according to the International Standard Classification of Education 2011 [30]), and married/living with a partner (yes/no). In addition, the following clinical characteristics were assessed via self-report: pre-existing comorbidities (see Multimedia Appendix 2), health status (good/moderate/poor) during the infection and at the moment of completing the surveys, date of symptom onset, symptoms during acute phase of COVID-19 and at the moment of completing the surveys (see Multimedia Appendix 2), COVID-19–related hospitalization, and COVID-19 diagnosis (based on reverse transcription polymerase chain reaction (PCR) test or computed tomography (CT) scan of the thorax/symptom-based medical diagnosis by a physician/no formal test or diagnosis). Based upon the latter, patients were classified as either "test-diagnosed" COVID-19 (PCR or CT) or "presumed" COVID-19 (physician-diagnosed or no formal diagnosis/testing).

Received Health Care

Information regarding received health care (yes/no) by a medical professional (eg, medical specialist; general practitioner [GP]; nurse) or an allied health care professional (AHP; eg, physiotherapist [PT]; psychologist; occupational therapist [OT]; dietician; speech and language therapist) was recorded.

Standardized Fatigue Questionnaire

Fatigue, the primary outcome measure, was measured using a subscale of the Checklist Individual Strength (CIS). The Checklist Individual Strength–subscale subjective fatigue (CIS-Fatigue) is a standardized questionnaire [31,32] with high internal consistency and test-retest reliability; good discriminant, concurrent and criterion validity; and ability to detect change in subjective fatigue [33-37]. The questionnaire consists of 8 items scored on a 7-point Likert scale. Scores range from 8 to 56 points, and a higher score indicates more clinical symptoms of general fatigue (see Multimedia Appendix 3 for the CIS-Fatigue questionnaire) [31,32]. Based upon validated cutoff values, individuals can be classified as having normal (≤26 points), mild (27-35 points), and severe (≥36 points) fatigue [31-33].

Self-constructed Physical and Mental Fatigue Questions

A total of 3 self-constructed questions (all part of the CIS-Fatigue subscale) were used to evaluate physical fatigue ("*Physically* I feel exhausted," "*Physically* I feel I am in a bad condition," and "*Physically* I feel in a good shape"). In addition, to differentiate between physical and mental fatigue, 3 questions were constructed in which the word "physically" was replaced by the word "mentally" ("*Mentally* I feel exhausted," "*Mentally* I feel I am in a bad condition," and "*Mentally* I feel in a good shape," respectively) to estimate mental fatigue. The physical and mental fatigue questions were scored on a 7-point Likert scale, with scores ranging from 3 to 21 points. A higher score indicates worse physical and mental fatigue, respectively. These self-constructed physical and mental fatigue questions and explanations of the scoring are reported in Multimedia Appendix 4.

Statistical Analyses

Data are presented as means and standard deviations, medians, and interquartile ranges or as frequencies and proportions, where appropriate. Differences over time were analyzed by a paired *t*-test (or Wilcoxon signed-rank test) in continuous data and a McNemar test (or McNemar-Bowker test) in categorical data. If significant, a post hoc comparison of the McNemar-Bowker test was performed, and significant Bonferroni-adjusted *P* values were generated as corrections for multiple comparison. Statistical analyses were conducted using SPSS 25.0 (IBM Corporation). Figures were generated via GraphPad Prism 8.3.5 (GraphPad Software) and SankeyMATIC [38]. The level of significance was set at .01 for all statistical tests (two-tailed).

Results

Participants' Inclusion

In total, 2159 members of online long COVID peer support groups filled out the first survey, of which 220 were excluded for being in the acute phase of COVID-19 (n=14), ICU admission during the acute phase of COVID-19 (n=15), onset of symptoms before January 1, 2020 (n=8), and an incomplete first survey (n=183). From the 1939 patients who were included, 1556 consented to be approached for follow-up research, of which 1005 (64.6%) completed the second survey. Patients who did not respond to the second survey were younger and more often had a presumed COVID-19 diagnosis. Further details can be found in a previously published paper [11]. The 1005 patients completed the surveys on average 11.3 weeks (SD 2.2) and 23.5 weeks (SD 2.2) after onset of symptoms (T1 and T2, respectively). Overall, 239 test-diagnosed (hospitalized, n=62, and nonhospitalized, n=177) and 766 presumed (physician-diagnosed, n=454, and patients with no formal diagnosis/testing, n=312) patients with COVID-19 participated in this 2-wave web-based survey (see Multimedia Appendix 5 for the flowchart).

Demographical and Clinical Characteristics

Patients with confirmed COVID-19 were mostly middle-aged women (median age 50.0 years, IQR 39.0-56.0; 198/239 women, 82.8%) with a BMI indicating slight overweight (median BMI 26.0 kg/m^2, IQR 23.4-30.5), and they completed the first (T1)

and second (T2) survey on average 10.4 weeks (SD 2.4) and 22.6 weeks (SD 2.4) after onset of symptoms. Approximately 1 out of 4 patients (62/239, 25.9%) was hospitalized during the acute phase of COVID-19. The majority of respondents had no self-reported comorbidities (142/239, 59.4%) and good self-reported health status before the infection (208/239, 87%). Moreover, at T1 and T2, a minority of respondents self-reported good health (22/239, 9.2%, and 40/239, 16.7%, respectively [11]). Furthermore, patients retrospectively reported a median of 15 symptoms (IQR 11-18) during the acute phase of COVID-19, and 6 symptoms (IQR 4-9) and 6 symptoms (IQR 3-8) at T1 and T2, respectively. All details regarding patient characteristics can be found in Table 1.

Table 1. Characteristics of patients with confirmed COVID-19 (n=239).

Characteristic	Value
Women, n (%)	198 (82.8)
Age (years), median (IQR)	50.0 (39.0-56.0)
BMI (kg/m^2), median (IQR)	26.0 (23.4-30.5)
Time between onset of symptoms and T1[a] survey (weeks), mean (SD)	10.4 (2.4)
Time between onset of symptoms and T2[b] survey (weeks), mean (SD)	22.6 (2.4)
Married/living with partner, n (%)	173 (72.4)
Educational level, n (%)	
Low	6 (2.5)
Medium	126 (52.7)
High	107 (44.8)
Pre-existing comorbidities, n (%)	
None	142 (59.4)
1	62 (25.9)
≥2	35 (14.6)
Health status before infection, n (%)	
Good	208 (87)
Moderate	28 (11.7)
Poor	3 (1.3)
Health status at T1, n (%)	
Good	22 (9.2)
Moderate	156 (65.3)
Poor	61 (25.5)
Number of symptoms, median (IQR)	
During acute infection	15 (11-18)
At T1	6 (4-9)
At T2	6 (3-8)
Hospitalized during acute infection, n (%)	62 (25.9)

[a]T1: time of completing the first survey.

[b]T2: time of completing the second survey.

Received Health Care

During the first 10 weeks after the onset of symptoms, 2 out of 3 patients (157/239, 65.7%) received or sought care from at least one medical professional (GP: 139/239, 58.2%; medical specialist: 73/239, 30.5%; nurse: 18/239, 7.5%), whereas 1 out of 3 (90/239, 37.7%) received or sought care from at least one allied health care professional (PT: 76/239, 31.8%; psychologist: 27/239, 11.3%; OT: 7/239, 2.9%, dietician: 25/239, 10.5%; and speech and language therapist: 6/239, 2.5%). The cumulative proportion of patients who received care from a medical professional and allied health care professional at T2 respectively increased significantly to 81.2% (194/239; GP: 170/239, 71.1%; medical specialist: 131/239, 54.8%; nurse: 32/239, 13.4%; all P<.001) and 68.6% (164/239; PT: 157/239, 65.7%; psychologist: 55/239, 23%; OT: 27/239, 11.3%; dietician, 51/239, 21.3%; speech and language therapist, 21/239, 8.8%; all P<.001). Furthermore, the cumulative proportion of

patients who participated in an interdisciplinary rehabilitation program (in- or outpatient) increased significantly from T1 to T2 (10/239, 4.2%, to 32/239, 13.4%, respectively; *P*<.001). Details regarding received health care can be found in Table 2.

Table 2. Fatigue-related measures and received health care in patients with confirmed COVID-19 at on average 10 weeks (T1) and 23 weeks (T2) after onset of symptoms (n=239).

	Value		
	T1[a]	T2[b]	*P* value
Fatigue-related measures			
General fatigue (points on CIS-Fatigue[c] scale), median (IQR)	48 (42-53)	46 (37-50)	<.001
Severe fatigue, n (%)	204 (85.4)	188 (78.7)	.03
Mental fatigue (points on self-constructed questions), median (IQR)	15 (10-17)	14 (10-17)	.52
Physical fatigue (points on self-constructed questions), median (IQR)	19 (16-20)	18 (14-19)	<.001
Received health care from a medical professional, n (%)			
Received care from ≥1 medical professionals	157 (65.7)	194 (81.2)	<.001
General practitioner	139 (58.2)	170 (71.1)	<.001
Medical specialist	73 (30.5)	131 (54.8)	<.001
Nurse	18 (7.5)	32 (13.4)	<.001
Received health care from an allied health care provider, n (%)			
Received care from ≥1 allied health care providers	90 (37.7)	164 (68.6)	<.001
Physiotherapist	76 (31.8)	157 (65.7)	<.001
Psychologist	27 (11.3)	55 (23)	<.001
Occupational therapist	7 (2.9)	27 (11.3)	<.001
Dietician	25 (10.5)	51 (21.3)	<.001
Speech and language therapist	6 (2.5)	21 (8.8)	<.001
Rehabilitation (in- or outpatient), n (%)	10 (4.2)	32 (13.4)	<.001

[a]T1: time of completing the first survey.

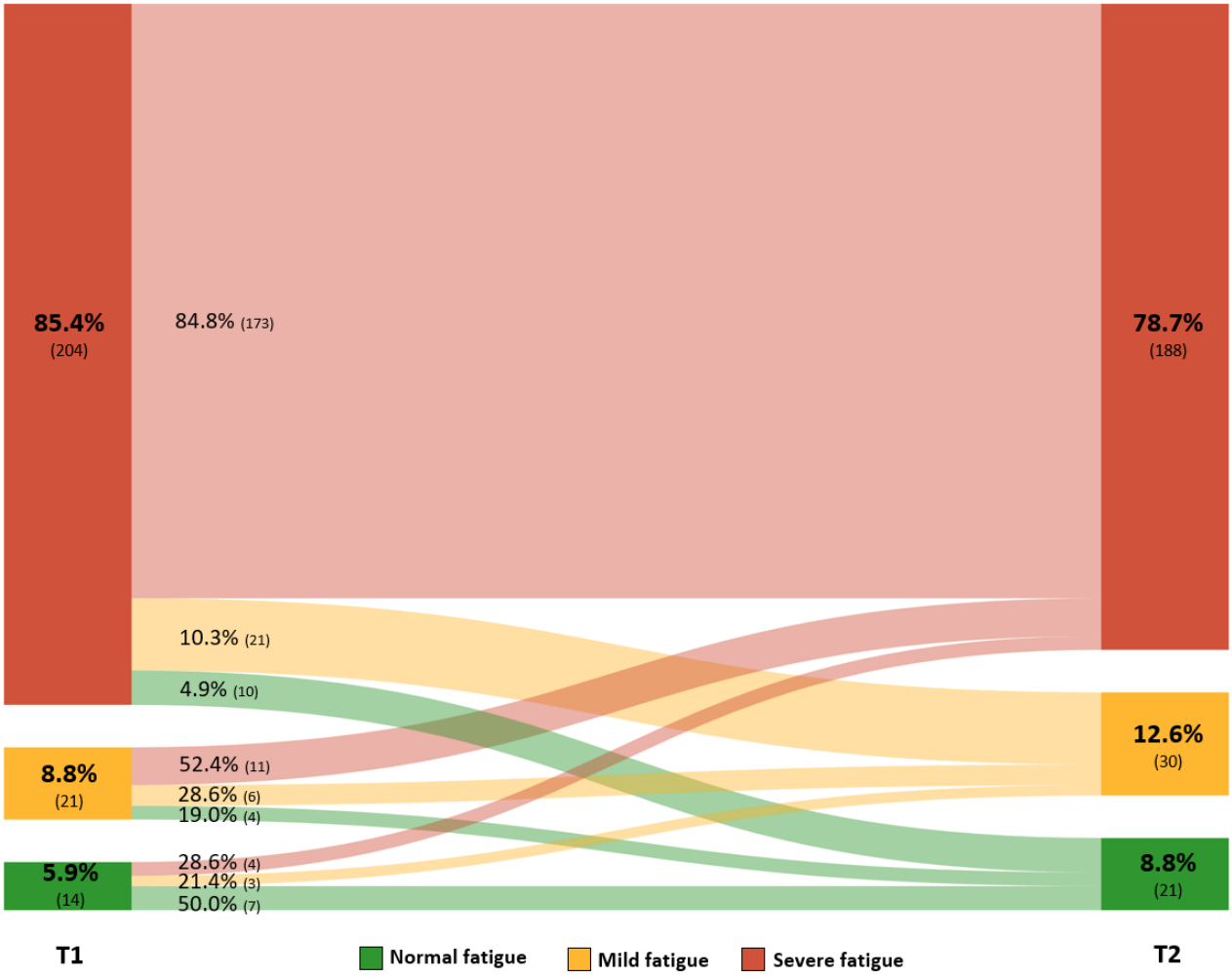
[b]T2: time of completing the second survey.

[c]CIS-Fatigue: Checklist Individual Strength–subscale subjective fatigue.

Standardized Fatigue Questionnaire

Patients with confirmed COVID-19 reported a median CIS-Fatigue score of 48 points (IQR 42-53) at T1. The majority (204/239, 85.4%) reported severe fatigue at approximately 3 months after the onset of COVID-19 symptoms. The median CIS-Fatigue score improved significantly between T1 and T2 (median change –2 points, IQR –7 to 3; *P*<.001) (Table 2), whereas no significant differences were found in the proportions of normal, mild, or severe fatigue (*P*=.12). An overview of the proportions of patients with normal, mild, and severe fatigue at T1 and T2, the proportional flow, and the direction of change can be found in Figure 1. In addition, Multimedia Appendix 6 shows the proportion, flow, and direction of the change of fatigue stratified for the type of diagnosis (ie, hospitalized and nonhospitalized test-diagnosed patients, physician-diagnosed patients, and patients without a formal diagnosis/test).

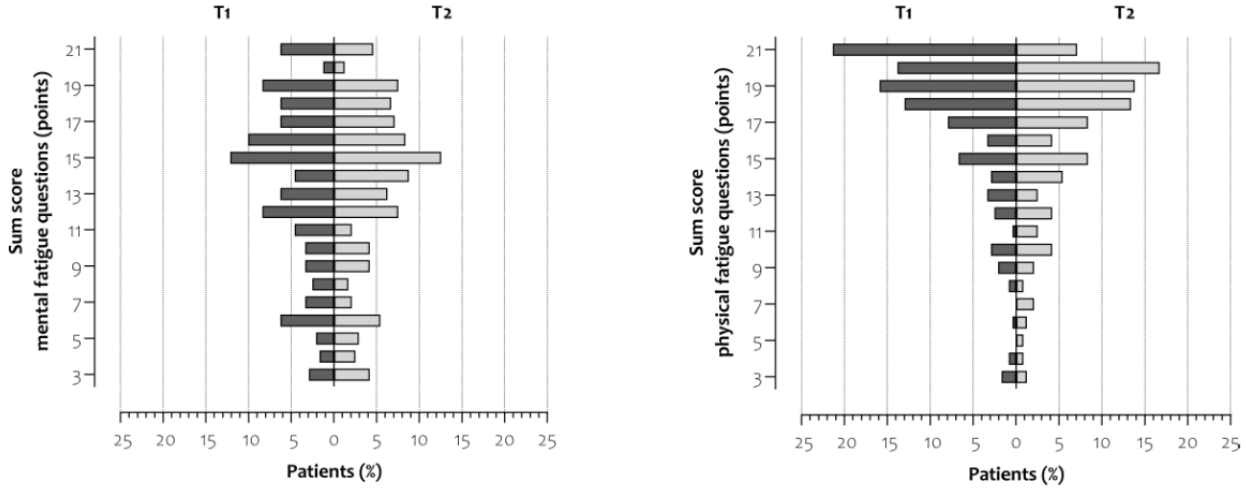
Figure 1. Prevalence and change in fatigue in patients with long COVID who have confirmed COVID-19, measured using the CIS-Fatigue scale at on average 10 (T1) and 23 (T2) weeks after onset of symptoms (n=239). The width of the lines is proportional to the flow rate. No significant change in the prevalence of normal (≤26 points), mild (27-35 points), or severe (≥36 points) fatigue was found between T1 and T2 (McNemar-Bowker test, *P*=.12). CIS-Fatigue: Checklist Individual Strength–subscale subjective fatigue; T1: time of completing the first survey; T2: time of completing the second survey.



Self-constructed Physical and Mental Fatigue Questions

Patients with confirmed COVID-19 reported median physical and mental fatigue scores of 19 points (IQR 16-20) and 15 points (IQR 10-17) at T1. Between T1 and T2, a significant decrease was found in physical fatigue score (median change –1 point, IQR –3 to 0; *P*<.001), but not in mental fatigue score (median change 0 points, IQR –3 to 3; *P*=.52) (Table 2). Figure 2 shows the distributions of patients across the spectrum of physical and mental fatigue at T1 and T2.

Figure 2. The distribution of patients with long COVID who have confirmed COVID-19 across the spectrum of self-constructed mental (left) and physical (right) fatigue at on average 10 (T1) and 23 (T2) weeks after onset of symptoms (n=239). Scores range from 3 to 21 points, and higher scores indicate higher levels of fatigue. T1: time of completing the first survey; T2: time of completing the second survey.



Discussion

Principal Findings

To the best of our knowledge, this is the first study to measure fatigue over time in members from online long COVID peer support groups with confirmed COVID-19 using a validated and standardized measurement with generic cutoff values to determine normal, mild, and severe fatigue. Our study indicates that severe fatigue is highly prevalent in patients with long COVID at approximately 3 and 6 months after the infection. Furthermore, our longitudinal follow-up data suggest that fatigue does not resolve over time in all patients, even if they receive health care. In addition, patients experience both physical and mental fatigue.

Fatigue is the most prominent symptom in patients with long COVID [2,4], irrespective of the severity of the initial infection [14]. Nevertheless, most studies are cross-sectional and use a binary question (eg, fatigued/not fatigued) to assess the prevalence of fatigue [2,16,39]. Therefore, little is known about the change in fatigue intensity over time [14,40]. Our study used a validated and standardized questionnaire to assess fatigue and was able to quantify fatigue intensity. Indeed, fatigue is highly prevalent in our sample. Moreover, fatigue was reported to be generally high. The median fatigue scores found in our sample are equal to or higher than those of other chronic diseases that are characterized by fatigue, such as chronic obstructive pulmonary disease [41], asthma [42], Q fever [20], multiple sclerosis [43], rheumatoid arthritis [44], or systemic sclerosis [45]. These findings are remarkable for such a young population with few self-reported comorbidities and good self-reported health status before the infection. Previously, other viral and nonviral infections have been linked to prolonged and debilitating fatigue [20,46-50]. For example, Lam and colleagues investigated long-term complaints in SARS survivors and found that approximately one-third of SARS survivors met the modified 1994 US Centers for Disease Control and Prevention criteria for chronic fatigue syndrome more than 3 years after having SARS [17]. Moreover, MERS survivors often experience chronic fatigue [19]. For patients with long COVID, it remains unclear whether fatigue will resolve spontaneously in the longer term. Our follow-up data show little to no improvement in the proportion of patients with severe fatigue between 3 and 6 months, despite receiving medical and allied health care. Consequently, almost two-thirds of the patients in our sample are progressing toward chronic fatigue (ie, severe fatigue that persists longer than six months [51]). The fact that some patients may experience debilitating chronic fatigue is worrisome and could have a major long-term impact upon these individuals as well as on the health care system and society as a whole [10,11,52]. Indeed, fatigue is strongly related to health-related QoL and aspects of day-to-day life [14,25,53,54], and it often involves sick leave, increased health care consumption, and more hidden costs, such as informal care by friends or family members [55-57].

Fatigue is a complex and challenging symptom, as multiple factors can play a role in the initiation and maintenance of fatigue, as seen in other chronic diseases [58]. It can present itself as mental fatigue, physical fatigue, or both [40]. Therefore, a patient-tailored treatment based upon a holistic and comprehensive assessment of systemic, physical, psychological, and behavioral factors is proposed to alleviate the fatigue symptom burden [59]. To date, it remains unknown which treatment strategies are effective to improve fatigue in patients with long COVID. Several treatment strategies for fatigue are proposed based upon knowledge from the fast-growing evidence regarding COVID-19 and other pathologies, such as multidisciplinary rehabilitation, energy conservation techniques, pacing, cognitive behavioral therapy, graded exercise therapy, or physical training [25,54,60-65]. Future research needs to provide evidence regarding underlying pathways, evaluate the effectiveness of existing treatment strategies, and identify susceptible candidates, as it is expected that not everyone will benefit from the same treatment strategy due to the multifactorial nature of fatigue. Moreover, anecdotal evidence shows that patients report having within-day and between-day variations in their daily symptoms, including fatigue [54,66,67]; these cannot be captured in detail by completing a questionnaire once or twice over a longer period of time. In this, the use of an

ecological momentary assessment may be valuable, as this approach involves repeated measurements of the participant's symptoms, behavior, and context in vivo and in real time [68]. More insights in diurnal variation in fatigue and its association with other symptoms may be useful in the development of more tailored treatment strategies for fatigue in patients with long COVID.

Methodological Considerations and Limitations

The current study has several limitations. First, the survey was only made available to members of online long COVID peer support groups. This probably caused selection bias, as it is reasonable to assume that patients with high symptom burden are more likely to become members of online long COVID peer support groups. Second, all results were collected using a web-based survey. Therefore, besides the self-reported symptoms, the patients' height, body weight, and medical status before and during the infection were also based on self-report, which may have affected the internal validity of the current findings to some extent. Recently, the National Institute for Health and Care Excellence [69] proposed a case definition of long COVID whereby alternative diagnosis should be excluded when identifying patients with long COVID. Due to the nature and timing of this study (ie, early phase of the pandemic), this was not possible in the current study. Third, approximately 1 out of 3 participants who consented to be approached for follow-up research did not respond to the second wave of the survey. The authors have no information about the possible reasons for not responding to the second wave of the survey, although a between-group comparison was made to find possible differences [11]. Fourth, the majority of our sample were women, which limits our external validity. Nevertheless, evidence is growing that women are more prone to develop long COVID [70]. Fifth, self-constructed questions were used to quantify mental fatigue, although validated questionnaires (such as the Chalder fatigue index) to assess mental (and physical) fatigue are available [71]. Therefore, no definite conclusions on the burden of mental fatigue in online long COVID peer support groups can be drawn based on the current study. Nevertheless, the current results indicate that COVID-19 can impact both physical and mental fatigue in the long term. Furthermore, this study was conducted in adults, although evidence regarding long COVID in children and adolescents is starting to emerge [72,73].

Conclusions

Severe fatigue is highly prevalent in members of online long COVID peer support groups both at approximately 3 and 6 months after onset of symptoms. As not enough time has passed since the start of the COVID-19 pandemic, it is unclear whether this fatigue will resolve spontaneously in the longer term. Future research needs to focus on the prognosis, possible causes, and treatment strategies for physical and mental fatigue in patients with long COVID.

Acknowledgments

The research team acknowledges the valuable input from the patient representatives to develop the survey, and the technical support by ASolutions' Martijn Briejers and Oscar Wagemakers. The scientific work of YMJG is financially supported by Lung Foundation Netherlands grant 4.1.16.085, FVCM is financially supported by ZonMW (ERACoSysMed grant 90030355), and RM is financially supported by Lung Foundation Netherlands grant 5.1.18.232.

Conflicts of Interest

MAS reports grants from Netherlands Lung Foundation, AstraZeneca, Boehringer Ingelheim, and Stichting Astma Bestrijding, all outside the submitted work. FMEF reports grants and personal fees from AstraZeneca, personal fees from Boehringer Ingelheim, personal fees from Chiesi, personal fees from GlaxoSmithKline, grants and personal fees from Novartis, and personal fees from TEVA, outside the submitted work. RP reports personal fees from MEDtalk and Health Investment. DJAJ has received lecture fees from Chiesi and Boehringer Ingelheim in the last 3 years, which are unrelated to this paper. All other authors declare that they have no conflicts of interest. No financial support was received for the preparation of this manuscript.

Multimedia Appendix 1
Members of online long COVID peer support groups with presumed COVID-19 (n=766): characteristics, received health care, and fatigue-related measures.
[DOCX File , 19 KB - jmir_v23i9e30274_app1.docx]

Multimedia Appendix 2
Additional information regarding self-reported pre-existing comorbidities and symptoms during the acute phase of COVID-19 and at the moment of completing the surveys.
[DOCX File , 15 KB - jmir_v23i9e30274_app2.docx]

Multimedia Appendix 3
The Checklist Individual Strength–subscale subjective fatigue.
[DOCX File , 24 KB - jmir_v23i9e30274_app3.docx]

Multimedia Appendix 4
Self-constructed physical and mental fatigue based upon 3 items of the Checklist Individual Strength–subscale subjective fatigue.
[DOCX File , 15 KB - jmir_v23i9e30274_app4.docx]

Multimedia Appendix 5
Flowchart of participants' inclusion.
[DOCX File , 77 KB - jmir_v23i9e30274_app5.docx]

Multimedia Appendix 6
Prevalence of normal, mild, and severe fatigue using the Checklist Individual Strength–subscale subjective fatigue at T1 and T2, the proportional flow, and the direction of change of fatigue stratified for type of diagnosis.
[DOCX File , 176 KB - jmir_v23i9e30274_app6.docx]

References

1. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, ISARIC4C investigators. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. BMJ 2020 May 22;369:m1985 [FREE Full text] [doi: 10.1136/bmj.m1985] [Medline: 32444460]
2. Carfì A, Bernabei R, Landi F. Persistent symptoms in patients after acute COVID-19. JAMA 2020 Aug 11;324(6):603. [doi: 10.1001/jama.2020.12603]
3. Huang C, Huang L, Wang Y, Li X, Ren L, Gu X, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. Lancet 2021 Jan 16;397(10270):220-232 [FREE Full text] [doi: 10.1016/S0140-6736(20)32656-8] [Medline: 33428867]
4. Goërtz YMJ, Van Herck M, Delbressine JM, Vaes AW, Meys R, Machado FVC, et al. Persistent symptoms 3 months after a SARS-CoV-2 infection: the post-COVID-19 syndrome? ERJ Open Res 2020 Oct;6(4):00542-2020 [FREE Full text] [doi: 10.1183/23120541.00542-2020] [Medline: 33257910]
5. Mahase E. Covid-19: What do we know about "long covid"? BMJ 2020 Jul 14;370:m2815. [doi: 10.1136/bmj.m2815] [Medline: 32665317]
6. Nabavi N. Long covid: How to define it and how to manage it. BMJ 2020 Sep 07;370:m3489. [doi: 10.1136/bmj.m3489] [Medline: 32895219]
7. Wong AW, Shah AS, Johnston JC, Carlsten C, Ryerson CJ. Patient-reported outcome measures after COVID-19: a prospective cohort study. Eur Respir J 2020 Nov;56(5):2003276 [FREE Full text] [doi: 10.1183/13993003.03276-2020] [Medline: 33008936]
8. Meys R, Delbressine JM, Goërtz YMJ, Vaes AW, Machado FVC, Van Herck M, et al. Generic and respiratory-specific quality of life in non-hospitalized patients with COVID-19. J Clin Med 2020 Dec 09;9(12):3993 [FREE Full text] [doi: 10.3390/jcm9123993] [Medline: 33317214]
9. Vaes AW, Machado FVC, Meys R, Delbressine JM, Goertz YMJ, Van Herck M, et al. Care dependency in non-hospitalized patients with COVID-19. J Clin Med 2020 Sep 12;9(9):2946 [FREE Full text] [doi: 10.3390/jcm9092946] [Medline: 32932582]
10. Tenforde MW, Kim SS, Lindsell CJ, Billig Rose E, Shapiro NI, Files DC, CDC COVID-19 Response Team, et al. Symptom duration and risk factors for delayed return to usual health among outpatients with COVID-19 in a multistate health care systems network - United States, March-June 2020. MMWR Morb Mortal Wkly Rep 2020 Jul 31;69(30):993-998 [FREE Full text] [doi: 10.15585/mmwr.mm6930e1] [Medline: 32730238]
11. Vaes AW, Goërtz YMJ, Van Herck M, Machado FVC, Meys R, Delbressine JM, et al. Recovery from COVID-19: a sprint or marathon? 6-month follow-up data from online long COVID-19 support group members. ERJ Open Res 2021 Apr 26;7(2):00141-2021 [FREE Full text] [doi: 10.1183/23120541.00141-2021] [Medline: 34041295]
12. Belli S, Balbi B, Prince I, Cattaneo D, Masocco F, Zaccaria S, et al. Low physical functioning and impaired performance of activities of daily life in COVID-19 patients who survived hospitalisation. Eur Respir J 2020 Oct;56(4):2002096 [FREE Full text] [doi: 10.1183/13993003.02096-2020] [Medline: 32764112]
13. Writing Committee for the COMEBAC Study Group, Morin L, Savale L, Pham T, Colle R, Figueiredo S, et al. Four-month clinical status of a cohort of patients after hospitalization for COVID-19. JAMA 2021 Apr 20;325(15):1525-1534. [doi: 10.1001/jama.2021.3331] [Medline: 33729425]
14. Townsend L, Dyer AH, Jones K, Dunne J, Mooney A, Gaffney F, et al. Persistent fatigue following SARS-CoV-2 infection is common and independent of severity of initial infection. PLoS One 2020;15(11):e0240784 [FREE Full text] [doi: 10.1371/journal.pone.0240784] [Medline: 33166287]
15. Ream E, Richardson A. Fatigue: a concept analysis. Int J Nurs Stud 1996 Oct;33(5):519-529. [doi: 10.1016/0020-7489(96)00004-1] [Medline: 8886902]

16. Garrigues E, Janvier P, Kherabi Y, Le Bot A, Hamon A, Gouze H, et al. Post-discharge persistent symptoms and health-related quality of life after hospitalization for COVID-19. J Infect 2020 Dec;81(6):e4-e6 [FREE Full text] [doi: 10.1016/j.jinf.2020.08.029] [Medline: 32853602]
17. Lam MH, Wing Y, Yu MW, Leung C, Ma RCW, Kong APS, et al. Mental morbidities and chronic fatigue in severe acute respiratory syndrome survivors: long-term follow-up. Arch Intern Med 2009 Dec 14;169(22):2142-2147. [doi: 10.1001/archinternmed.2009.384] [Medline: 20008700]
18. Moldofsky H, Patcai J. Chronic widespread musculoskeletal pain, fatigue, depression and disordered sleep in chronic post-SARS syndrome; a case-controlled study. BMC Neurol 2011 Mar 24;11(1):37 [FREE Full text] [doi: 10.1186/1471-2377-11-37] [Medline: 21435231]
19. Lee SH, Shin H, Park HY, Kim JL, Lee JJ, Lee H, et al. Depression as a mediator of chronic fatigue and post-traumatic stress symptoms in Middle East respiratory syndrome survivors. Psychiatry Investig 2019 Jan;16(1):59-64 [FREE Full text] [doi: 10.30773/pi.2018.10.22.3] [Medline: 30605995]
20. Bronner MB, Haagsma JA, Dontje ML, Barmentloo L, Kouwenberg RMCEJ, Olde Loohuis AGM, et al. Long-term impact of a Q-fever outbreak: an evaluation of health symptoms, health-related quality of life, participation and health care satisfaction after ten years. J Psychosom Res 2020 Dec;139:110258 [FREE Full text] [doi: 10.1016/j.jpsychores.2020.110258] [Medline: 33069049]
21. Karshikoff B, Sundelin T, Lasselin J. Role of inflammation in human fatigue: relevance of multidimensional assessments and potential neuronal mechanisms. Front Immunol 2017;8:21 [FREE Full text] [doi: 10.3389/fimmu.2017.00021] [Medline: 28163706]
22. Corona patiënten met langdurige klachten (Nederland). Facebook. URL: https://www.facebook.com/groups/236723204035929.9 [accessed 2020-06-04]
23. Corona patiënten (met langdurige klachten) (Vlaanderen). Facebook. URL: https://www.facebook.com/groups/241043323639334/ [accessed 2020-06-04]
24. Coronaplein. Website in Dutch. Long Fonds. URL: https://coronaplein.nu/ [accessed 2021-09-09]
25. Greenhalgh T, Knight M, A'Court C, Buxton M, Husain L. Management of post-acute covid-19 in primary care. BMJ 2020 Aug 11;370:m3026. [doi: 10.1136/bmj.m3026] [Medline: 32784198]
26. Machado FVC, Meys R, Delbressine JM, Vaes AW, Goërtz YMJ, van Herck M, et al. Construct validity of the Post-COVID-19 Functional Status Scale in adult subjects with COVID-19. Health Qual Life Outcomes 2021 Feb 03;19(1):40 [FREE Full text] [doi: 10.1186/s12955-021-01691-2] [Medline: 33536042]
27. Houben-Wilke S, Delbressine JM, Vaes AW, Goërtz YM, Meys R, Machado FV, et al. Understanding and being understood: information and care needs of 2113 patients with confirmed or suspected COVID-19. J Patient Exp 2021 Mar 08;8:2374373521997222 [FREE Full text] [doi: 10.1177/2374373521997222] [Medline: 34179387]
28. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Int J Surg 2014 Dec;12(12):1500-1524 [FREE Full text] [doi: 10.1016/j.ijsu.2014.07.014] [Medline: 25046751]
29. ASolutions. Webpage in Dutch. URL: https://www.asolutions.nl/ [accessed 2021-09-09]
30. International Standard Classification of Education ISCED 2011. UNESCO Institute for Statistics. URL: http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf [accessed 2021-09-09]
31. Vercoulen JH, Swanink CM, Fennis JF, Galama JM, van der Meer JW, Bleijenberg G. Dimensional assessment of chronic fatigue syndrome. J Psychosom Res 1994 Jul;38(5):383-392. [doi: 10.1016/0022-3999(94)90099-x] [Medline: 7965927]
32. Vercoulen J, Alberts M, Bleijenberg G. De Checklist Individual Strength (CIS). Gedragstherapie 1999;32:131-136. [doi: 10.1037/t11003-000]
33. Worm-Smeitink M, Gielissen M, Bloot L, van Laarhoven HWM, van Engelen BGM, van Riel P, et al. The assessment of fatigue: psychometric qualities and norms for the Checklist individual strength. J Psychosom Res 2017 Jul;98:40-46. [doi: 10.1016/j.jpsychores.2017.05.007] [Medline: 28554371]
34. Beurskens AJ, Bültmann U, Kant I, Vercoulen JH, Bleijenberg G, Swaen GM. Fatigue among working people: validity of a questionnaire measure. Occup Environ Med 2000 May;57(5):353-357 [FREE Full text] [doi: 10.1136/oem.57.5.353] [Medline: 10769302]
35. Panitz S, Kornhuber M, Hanisch F. The checklist individual strength (CIS20-R) in patients with amyotrophic lateral sclerosis - a longitudinal study. Acta Neurol Scand 2015 Jun;131(6):372-380. [doi: 10.1111/ane.12349] [Medline: 25639482]
36. van Koulil S, Kraaimaat FW, van Lankveld W, van Riel PLCM, Evers AWM. A patient's perspective on multidisciplinary treatment gain for fibromyalgia: an indicator for pre-post treatment effects? Arthritis Rheum 2009 Dec 15;61(12):1626-1632 [FREE Full text] [doi: 10.1002/art.24792] [Medline: 19950305]
37. Evers AWM, Kraaimaat FW, van Riel PLCM, de Jong AJL. Tailored cognitive-behavioral therapy in early rheumatoid arthritis for patients at risk: a randomized controlled trial. Pain 2002 Nov;100(1-2):141-153. [doi: 10.1016/s0304-3959(02)00274-9] [Medline: 12435467]
38. SankeyMATIC: a Sankey diagram builder for everyone. SankeyMATIC. URL: http://sankeymatic.com/ [accessed 2021-04-08]

39. Daher A, Balfanz P, Cornelissen C, Müller A, Bergs I, Marx N, et al. Follow up of patients with severe coronavirus disease 2019 (COVID-19): Pulmonary and extrapulmonary disease sequelae. Respir Med 2020;174:106197 [FREE Full text] [doi: 10.1016/j.rmed.2020.106197] [Medline: 33120193]
40. Stavem K, Ghanima W, Olsen MK, Gilboe HM, Einvik G. Prevalence and determinants of fatigue after COVID-19 in non-hospitalized subjects: a population-based study. Int J Environ Res Public Health 2021 Feb 19;18(4):2030 [FREE Full text] [doi: 10.3390/ijerph18042030] [Medline: 33669714]
41. Goërtz YMJ, Spruit MA, Van 't Hul AJ, Peters JB, Van Herck M, Nakken N, et al. Fatigue is highly prevalent in patients with COPD and correlates poorly with the degree of airflow limitation. Ther Adv Respir Dis 2019;13:1753466619878128 [FREE Full text] [doi: 10.1177/1753466619878128] [Medline: 31558115]
42. Van Herck M, Spruit MA, Burtin C, Djamin R, Antons J, Goërtz YMJ, et al. Fatigue is highly prevalent in patients with asthma and contributes to the burden of disease. J Clin Med 2018 Nov 23;7(12):471 [FREE Full text] [doi: 10.3390/jcm7120471] [Medline: 30477110]
43. Van Heugten C, Köhler S, Francke M, Bol Y. The association between executive functioning, coping styles and depressive symptoms in patients with Multiple Sclerosis. Mult Scler Relat Disord 2019 Nov;36:101392. [doi: 10.1016/j.msard.2019.101392] [Medline: 31526983]
44. van Hoogmoed D, Fransen J, Bleijenberg G, van Riel P. Physical and psychosocial correlates of severe fatigue in rheumatoid arthritis. Rheumatology (Oxford) 2010 Jul;49(7):1294-1302. [doi: 10.1093/rheumatology/keq043] [Medline: 20353956]
45. Kwakkenbos L, van Lankveld WGJM, Vonk MC, Becker ES, van den Hoogen FHJ, van den Ende CHM. Disease-related and psychosocial factors associated with depressive symptoms in patients with systemic sclerosis, including fear of progression and appearance self-esteem. J Psychosom Res 2012 Mar;72(3):199-204 [FREE Full text] [doi: 10.1016/j.jpsychores.2011.12.005] [Medline: 22325699]
46. Sigal LH. Summary of the first 100 patients seen at a Lyme disease referral center. Am J Med 1990 Jun;88(6):577-581. [doi: 10.1016/0002-9343(90)90520-n] [Medline: 2346158]
47. Straus SE, Tosato G, Armstrong G, Lawley T, Preble OT, Henle W, et al. Persisting illness and fatigue in adults with evidence of Epstein-Barr virus infection. Ann Intern Med 1985 Jan;102(1):7-16. [doi: 10.7326/0003-4819-102-1-7] [Medline: 2578268]
48. Voss JG. Predictors and correlates of fatigue in HIV/AIDS. J Pain Symptom Manage 2005 Feb;29(2):173-184 [FREE Full text] [doi: 10.1016/j.jpainsymman.2004.05.006] [Medline: 15733809]
49. Schanke AK, Stanghelle JK. Fatigue in polio survivors. Spinal Cord 2001 May;39(5):243-251. [doi: 10.1038/sj.sc.3101147] [Medline: 11438839]
50. Epstein L, Wong KK, Kallen AJ, Uyeki TM. Post-Ebola signs and symptoms in U.S. survivors. N Engl J Med 2015 Dec 17;373(25):2484-2486. [doi: 10.1056/NEJMc1506576] [Medline: 26672870]
51. Korenromp IHE, Meeus M, Bleijenberg G. Dutch language area definition of chronic fatigue. Article in Dutch. Ned Tijdschr Geneeskd 2012;156(16):A4403. [Medline: 22510416]
52. Del Rio C, Collins LF, Malani P. Long-term health consequences of COVID-19. JAMA 2020 Nov 03;324(17):1723-1724. [doi: 10.1001/jama.2020.19719] [Medline: 33031513]
53. Garner P. The BMJ opinion: COVID-19 and fatigue – a game of snakes and ladders. BMJ Opinion. 2020. URL: https://blogs.bmj.com/bmj/2020/05/19/paul-garner-covid-19-and-fatigue-a-game-of-snakes-and-ladders/
54. Humphreys H, Kilby L, Kudiersky N, Copeland R. Long COVID and the role of physical activity: a qualitative study. BMJ Open 2021 Mar 10;11(3):e047632 [FREE Full text] [doi: 10.1136/bmjopen-2020-047632] [Medline: 33692189]
55. Janssen N, Kant IJ, Swaen GMH, Janssen PPM, Schröer CAP. Fatigue as a predictor of sickness absence: results from the Maastricht cohort study on fatigue at work. Occup Environ Med 2003 Jun;60(Suppl 1):i71-i76 [FREE Full text] [doi: 10.1136/oem.60.suppl_1.i71] [Medline: 12782750]
56. McCrone P, Darbishire L, Ridsdale L, Seed P. The economic cost of chronic fatigue and chronic fatigue syndrome in UK primary care. Psychol Med 2003 Feb;33(2):253-261. [doi: 10.1017/s0033291702006980] [Medline: 12622304]
57. Sabes-Figuera R, McCrone P, Hurley M, King M, Donaldson AN, Ridsdale L. The hidden cost of chronic fatigue to patients and their families. BMC Health Serv Res 2010 Mar 04;10:56 [FREE Full text] [doi: 10.1186/1472-6963-10-56] [Medline: 20202216]
58. Menting J, Tack CJ, Bleijenberg G, Donders R, Droogleever Fortuyn HA, Fransen J, et al. Is fatigue a disease-specific or generic symptom in chronic medical conditions? Health Psychol 2018 Jun;37(6):530-543. [doi: 10.1037/hea0000598] [Medline: 29781654]
59. Spruit MA, Vercoulen JH, Sprangers MAG, Wouters EFM, FAntasTIGUE consortium. Fatigue in COPD: an important yet ignored symptom. Lancet Respir Med 2017 Jul;5(7):542-544. [doi: 10.1016/S2213-2600(17)30158-3] [Medline: 28438501]
60. Ferraro F, Calafiore D, Dambruoso F, Guidarini S, de Sire A. COVID-19 related fatigue: which role for rehabilitation in post-COVID-19 patients? A case series. J Med Virol 2021 Apr;93(4):1896-1899. [doi: 10.1002/jmv.26717] [Medline: 33295637]

61. Vink M, Vink-Niese A. Could cognitive behavioural therapy be an effective treatment for long COVID and post COVID-19 fatigue syndrome? Lessons from the Qure Study for Q-Fever Fatigue Syndrome. Healthcare (Basel) 2020 Dec 11;8(4):552 [FREE Full text] [doi: 10.3390/healthcare8040552] [Medline: 33322316]
62. Van Herck M, Antons J, Vercoulen JH, Goërtz YMJ, Ebadi Z, Burtin C, et al. Pulmonary rehabilitation reduces subjective fatigue in COPD: a responder analysis. J Clin Med 2019 Aug 20;8(8):1264 [FREE Full text] [doi: 10.3390/jcm8081264] [Medline: 31434343]
63. Rimes KA, Chalder T. Treatments for chronic fatigue syndrome. Occup Med (Lond) 2005 Jan;55(1):32-39. [doi: 10.1093/occmed/kqi015] [Medline: 15699088]
64. Gloeckl R, Leitl D, Jarosch I, Schneeberger T, Nell C, Stenzel N, et al. Benefits of pulmonary rehabilitation in COVID-19: a prospective observational cohort study. ERJ Open Res 2021 Apr 11;7(2):00108-2021 [FREE Full text] [doi: 10.1183/23120541.00108-2021] [Medline: 34095290]
65. Daynes E, Gerlis C, Chaplin E, Gardiner N, Singh SJ. Early experiences of rehabilitation for individuals post-COVID to improve fatigue, breathlessness exercise capacity and cognition - A cohort study. Chron Respir Dis 2021;18:14799731211015691 [FREE Full text] [doi: 10.1177/14799731211015691] [Medline: 33957805]
66. Davis HE, Assaf GS, McCorkell L, Wei H, Low RJ, Re'em Y, et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. EClinicalMedicine 2021 Aug;38:101019 [FREE Full text] [doi: 10.1016/j.eclinm.2021.101019] [Medline: 34308300]
67. Ladds E, Rushforth A, Wieringa S, Taylor S, Rayner C, Husain L, et al. Persistent symptoms after Covid-19: qualitative study of 114 "long Covid" patients and draft quality principles for services. BMC Health Serv Res 2020 Dec 20;20(1):1144 [FREE Full text] [doi: 10.1186/s12913-020-06001-y] [Medline: 33342437]
68. Maes IHL, Delespaul PAEG, Peters ML, White MP, van Horn Y, Schruers K, et al. Measuring health-related quality of life by experiences: the experience sampling method. Value Health 2015 Jan;18(1):44-51 [FREE Full text] [doi: 10.1016/j.jval.2014.10.003] [Medline: 25595233]
69. COVID-19 rapid guideline: managing the long-term effects of COVID-19. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/ng188 [accessed 2021-01-21]
70. Sudre CH, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, et al. Attributes and predictors of long COVID. Nat Med 2021 Apr;27(4):626-631. [doi: 10.1038/s41591-021-01292-y] [Medline: 33692530]
71. Chalder T, Berelowitz G, Pawlikowska T, Watts L, Wessely S, Wright D, et al. Development of a fatigue scale. J Psychosom Res 1993;37(2):147-153. [doi: 10.1016/0022-3999(93)90081-p] [Medline: 8463991]
72. Hertting O. More research is needed on the long-term effects of COVID-19 on children and adolescents. Acta Paediatr 2021 Mar;110(3):744-745 [FREE Full text] [doi: 10.1111/apa.15731] [Medline: 33395729]
73. Magnusson K, Skyrud K, Suren P, Greve-Isdahl M, Størdal K, Kristoffersen D. Health care use up to 6 months after COVID-19 in 700.000 children and adolescents: a pre-post study. medRxiv Preprint posted online on June 05, 2021. [doi: 10.1101/2021.06.02.21258211]

Abbreviations

CIS-Fatigue: Checklist Individual Strength–subscale subjective fatigue
CT: computed tomography
GP: general practitioner
ICU: intensive care unit
MERS: Middle East respiratory syndrome
OT: occupational therapist
PCR: polymerase chain reaction
PT: physiotherapist
QoL: quality of life
SARS: severe acute respiratory syndrome
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
T1: time point of completing the first survey
T2: time point of completing the second survey

Edited by C Basch; submitted 08.05.21; peer-reviewed by B Chaudhry, C Roldan; comments to author 28.06.21; revised version received 31.07.21; accepted 29.08.21; published 21.09.21.

Please cite as:
Van Herck M, Goërtz YMJ, Houben-Wilke S, Machado FVC, Meys R, Delbressine JM, Vaes AW, Burtin C, Posthuma R, Franssen FME, Hajian B, Vijlbrief H, Spies Y, van 't Hul AJ, Janssen DJA, Spruit MA
Severe Fatigue in Long COVID: Web-Based Quantitative Follow-up Study in Members of Online Long COVID Support Groups
J Med Internet Res 2021;23(9):e30274
URL: https://www.jmir.org/2021/9/e30274
doi:10.2196/30274
PMID:34494964

Original Paper

Content and Communication of Inpatient Family Visitation Policies During the COVID-19 Pandemic: Sequential Mixed Methods Study

Joanna Hart[1,2,3,4], MSc, MD; Amy Summer[3], BA, MSW; Kuldeep N Yadav[2,3], BA; Summer Peace[3]; David Hong[3], BA; Michael Konu[3]; Justin T Clapp[2,4,5], DPhil

[1]Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[2]Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA, United States

[3]Palliative and Advanced Illness Research Center, Department of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[4]Department of Medical Ethics and Health Policy, University of Pennsylvania, Philadelphia, PA, United States

[5]Department of Anesthesia and Critical Care, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Joanna Hart, MSc, MD
Division of Pulmonary, Allergy, and Critical Care Medicine
Department of Medicine
University of Pennsylvania
423 Guardian Drive
Blockley Hall 305
Philadelphia, PA, 19104
United States
Phone: 1 215 573 3976
Email: joanna.hart@pennmedicine.upenn.edu

Abstract

Background: Inpatient health care facilities restricted inpatient visitation due to the COVID-19 pandemic. There is no existing evidence of how they communicated these policies to the public nor the impact of their communication choices on public perception.

Objective: This study aims to describe patterns of inpatient visitation policies during the initial peak of the COVID-19 pandemic in the United States and the communication of these policies to the general public, as well as to identify communication strategies that maximize positive impressions of the facility despite visitation restrictions.

Methods: We conducted a sequential, exploratory, mixed methods study including a qualitative analysis of COVID-19 era visitation policies published on Pennsylvania-based facility websites, as captured between April 30 and May 20, 2020 (ie, during the first peak of the COVID-19 pandemic in the United States). We also conducted a factorial survey-based experiment to test how key elements of hospitals' visitation policy communication are associated with individuals' willingness to seek care in October 2020. For analysis of the policies, we included all inpatient facilities in Pennsylvania. For the factorial experiment, US adults were drawn from internet research panels. The factorial survey-based experiment presented composite policies that varied in their justification for restricted visitation, the degree to which the facility expressed ownership of the policy, and the inclusion of family-centered care support plans. Our primary outcome was participants' willingness to recommend the hypothetical facility using a 5-point Likert scale.

Results: We identified 104 unique policies on inpatient visitation from 363 facilities' websites. The mean Flesch-Kincaid Grade Level for the policies was 14.2. Most policies prohibited family presence (99/104, 95.2%). Facilities justified the restricted visitation policies on the basis of community protection (59/104, 56.7%), authorities' guidance or regulations (34/104, 32.7%), or scientific rationale (23/104, 22.1%). A minority (38/104, 36.5%) addressed how restrictive visitation may impair family-centered care. Most of the policies analyzed used passive voice to communicate restrictions. A total of 1321 participants completed the web-based survey. Visitation policy elements significantly associated with willingness to recommend the facility included justifications based on community protection (OR 1.44, 95% CI 1.24-1.68) or scientific rationale (OR 1.30, 95% CI 1.12-1.51), rather than those based on a governing authority. The facility expressed a high degree of ownership over the decision (OR 1.16, 95% CI 1.04-1.29), rather than a low degree of ownership; and inclusion of family-centered care support plans (OR 2.80, 95% CI 2.51-3.12), rather than no such support.

Conclusions: Health systems can immediately improve public receptiveness of restrictive visitation policies by emphasizing community protection, ownership over the facility's policy, and promoting family-centered care.

(J Med Internet Res 2021;23(9):e28897) doi:10.2196/28897

KEYWORDS

family-centered care; hospital policy; public health; health communication; ethics; health systems; public perceptions; COVID-19; pandemic

Introduction

During the global COVID-19 pandemic, health systems rapidly changed procedures and policies to minimize viral transmission and accommodate increased patient volumes and illness acuity [1]. One example of such a policy change is the restriction or elimination of family presence in inpatient settings [2-7]. The Centers for Disease Control and Prevention (CDC), the Center for Medicare and Medicaid Services, and national professional organizations recommended that health systems limit the presence of patients' family members to reduce viral spread [2,8,9]. This recommendation reverses decades of policy and cultural evolution emphasizing the integral role of family members during inpatient care, based on research demonstrating that family presence during hospitalization improves patients' and family members' outcomes [2,3,10,11]. The language health systems use to communicate such policy changes likely shapes public perceptions of inpatient health care facilities at a time when there is an increased need for urgent medical care [12,13].

In the last two decades, the internet has become a critical vehicle in communicating information during public health emergencies [14]. During past crises, the public reported relying on health websites and physicians as trustworthy information sources. Prior work reveals that when the public lacks trust in a health care system, individuals are less willing to cooperate with health recommendations and seek the health services offered [15]. Communication during crises shapes individuals' trust in health care systems through modifying their perceptions of the institutions' competency, responsibility, and ability to fulfill its obligations to the public [16]. Therefore, effective communication of novel, crisis-era policies by health systems is critical to promoting public health and trust.

Crisis communication during a public health emergency is most useful when leveraged to improve community health and outcomes, rather than aimed to promote the reputation or image of an organization [16-18]. For example, the Crisis and Emergency Risk Communication (CERC) framework [19], which was developed by the CDC, merges both risk communication (ie, stakeholder-tailored communication focused on promoting behavior change) and crisis communication (ie, communication to reduce the negative effects of a crisis on stakeholders). CERC emphasizes that communication is critical to reducing health and psychological stressors during a crisis, in part through reestablishing a sense of personal control and encouraging self-efficacy. In this way, the current understanding of crisis communication is that its ultimate goal is not to inform the public but rather to develop high-quality social relationships [14,16,20].

Inpatient visitation policies are an example of crisis communication provided by health systems during the COVID-19 pandemic, and these frequently appear on health systems' public websites. Despite widespread attention in the lay press [21-24], there is no published evidence systematically summarizing how health systems implemented national recommendations for restricted inpatient visitation, how they communicated these policies to the general public, and the impact of their communication choices on the public's willingness to engage with the health care system. Therefore, we sought to characterize external communication made by inpatient health care facilities regarding their visitation policies during the COVID-19 pandemic. We aimed to identify key strategies for communicating such policies to improve their public reception [10]. To do so, we aimed to (1) describe the visitation policies of inpatient health care facilities during the COVID-19 pandemic, (2) classify patterns of how facilities communicated these policies to the public, and (3) identify associations between these communication patterns and public perception of the facility.

Methods

Overview

We conducted a sequential exploratory mixed-methods study [25]: first, we conducted a qualitative content analysis of visitation policies published on the internet by inpatient health care facilities during the COVID-19 era, followed by a factorial survey-based experiment to identify associations between elements of visitation policy communication and individuals' perceptions of the facilities.

Qualitative Analysis of Visitation Policy Communication

We identified inpatient health care facilities using the Pennsylvania Department of Health directory. We collected text and screenshots of each facility's inpatient visitation policy from their websites between April 30 and May 20, 2020, during the first peak of COVID-19 cases in the United States and Pennsylvania (Multimedia Appendix 1) [26]. We eliminated duplicates of identical policies published by facilities within the same health system. We used Rural-Urban Commuting Area (RUCA) Codes to classify each facility as urban or rural (ie, RUCA >2) [27,28]. We assessed the readability of each policy using the Flesch-Kincaid Grade Level calculator [29]. This readability tool calculates an overall complexity of text using the mean sentence and word length, providing an overall score of educational school-grade levels 1 to 12 [30]. Scores >12 indicate a college-level education would be required to read the text.

Five investigators conducted qualitative analysis of the visitation policy statements. We first used qualitative content analysis, which sorts open-ended data using categories derived from close reading of texts; it is not associated with any particular theory or paradigm, but it is a generic means of categorizing the "informational content" of data [31]. We annotated the policy statements to identify the various functions they accomplished (eg, stating restrictions, justifying restrictions, expressing a facility's values; Table 1). Within each of these function categories, we derived more specific categories to describe their content (eg, what the specific restrictions were, what information a facility presented to justify them, what specific values a facility expressed). Second, we performed discourse analysis—a method that examines the contextual use of language [32,33], to allow us to generate hypotheses about the effects of the specific language used to convey policies on readers. Through annotation, we observed that policies varied in how they positioned facilities as agentive or nonagentive in enacting restrictions through use of voice (active vs passive), sentence structure, and modality. Annotation was done independently and then discussed as a team. Based on this discussion, a formal codebook was generated to capture patterns in function, content, and discourse. Each policy was then coded using NVivo 12 qualitative analysis software (QSR International) by at least 2 team members (AS, SP, DH, and MK), with discrepancies resolved and codebook refinements made by consensus and consultation with supervising investigators (JH and JC). Finally, we performed focused coding, prioritizing the codes that we believed were most important to how visitation policies were communicated, identifying relationships between these codes, and synthesizing related codes [34-38]. Focused coding was done through group discussion and captured via analytic memos.

Factorial Experiment of Visitation Policy Communication

We used the results of the qualitative analysis to develop a factorial survey-based experiment that tested the impact of variation in hospitals' public visitation policy communication. We selected three elements to vary based on our qualitative findings by generating a list of elements that varied and eliminating those well-established to influence public opinion and understanding (eg, reading level) or those that health systems may be unable or unwilling to change (eg, the visitation rules themselves). Our final selections included (1) how the inpatient facility justified restricting visitation (eg, community protection, regulations from a governing authority, or scientific rationale); (2) the degree to which the facility expressed ownership over the policy decisions (eg, use of active voice to indicate responsibility for decision-making); and (3) the inclusion or absence of family-centered care support plans (eg, virtual visits or team communication plans and expressions of concern for well-being). Although actual policies justifying the restricted visitation on the basis of governing authorities more frequently used passive voice, in our factorial experiment, these varied independently.

Our final 3×2×2 factorial experiment structure generated 12 hypothetical policies (Multimedia Appendix 1). For our visitation policy statements presented to participants in the survey, we used the most common visitation rules and language drawn from actual facility policies in a composite manner to represent our selected elements. We used a Flesch-Kincaid Grade Level range of 9.0 to 10.8 for all policies. This was lower than the average policy in our sample, but it remains above the recommended reading level for patient education materials, as we used facilities' actual word choices whenever possible for constructing the composite policies [39,40]. We reviewed our element selections and the survey with groups of clinical and nonclinical research staff and pilot tested the survey with nonclinical individuals prior to launch.

We recruited participants in October 2020, during the start of the second and largest peak of COVID-19 cases in the United States (Multimedia Appendix 1). We used Cloud Research's Prime Panels, an internet study recruitment platform that enables and ensures high-quality survey-based data collection from the over 50 million diverse enrolled participants who more closely represent the US population than other similar platforms, such as MTurk [41]. The platform also allows for recruitment based on targeted eligibility criteria. Potential participants were eligible to voluntarily choose to complete this study if they resided in the United States, were ≥18 years old, and were fluent in written English. We conducted stratified recruitment to ensure that a minimum of 44% of participants were women and 13% were Black or African American to reflect US demographics more closely [41]. We asked all participants to complete a survey that included reviewing the 12 hypothetical policies and responding after each using a 5-point Likert scale (ie, ranging from "1=not at all" to "5=very much"), indicating the participant's willingness to recommend the hospital on this basis alone. They then responded to survey items collecting their personal perspectives on COVID-19 risk and mitigation strategies, the 9-item Revised Healthcare System Distrust Scale (HSDS) [42], and a demographic questionnaire. We included three attention checks and participants who failed two of these checks were no longer eligible to be included in the analysis [41,43]. Participants received a nominal compensation for successful survey completion. All participants provided informed consent after reviewing written information prior to beginning the survey. All study procedures were approved by the University of Pennsylvania's Institutional Review Board.

Using the n.multiway function from the R package easypower, we calculated sample size estimates of our 3×2×2 factorial design based on the number of factors (ie, policy elements), the number of levels of each factor, and various effect sizes for a main effect of a factor or for the interaction between two factors on the willingness to recommend the hospital (Table S1 in Multimedia Appendix 1). We estimated that recruiting 1272 participants (ie, 106 for each of 12 hypothetical policies) would be sufficient to detect an odds ratio (OR) of 1.437 for the main effect or interaction with 90% power and an α=.05.

We examined the relationship among the three policy elements and willingness to recommend the hospital using mixed effects logistic regression. The participant was entered as a random effect to account for the clustering of responses by each participant [44]. The regression model was fit using maximum likelihood estimation and was implemented using the glmer function from the R package lme4 [45]. Prior to model building, we determined that continuous variables were appropriate to

enter as linear terms without transformation. The unadjusted associations between all predictor variables and the primary outcome were examined using *t* tests and chi-square tests for continuous and categorical variables, respectively. A priori, we decided to include all three tested elements in the baseline model. Other variables were selected for inclusion using backward variable selection in a stepwise manner from a model with all possible covariates. Those that were not significant (eg, age and gender of the participant) were not included in the final model given we had no prespecified hypotheses as to the direction of effect. The final model only included predictor variables with *P* values <.05 for the association between the variable and the primary outcome in the baseline model, including the three exposure variables [46]. We explored potential interactions among all three visitation policy elements by adding two-way interaction terms one at a time to the final model, but none were significant. Two-sided *P* values <.05 were considered statistically significant. All statistical analyses were conducted in R Studio.

Results

Qualitative Analysis of Visitation Policy Communication

We identified 513 inpatient health care facilities within Pennsylvania. Using RUCA classification, 80.3% (n=412) were urban and 20.7% (n=101) were rural. We identified inpatient visitation policies for 363 facilities, resulting in 104 unique policies after eliminating duplicates used for multiple facilities within a health system (Figure S1 in Multimedia Appendix 1). The mean Flesch-Kincaid Grade Level for the policies was 14.2 (ie, requiring college-level reading skills). In addition to English, the policy was available in Spanish for 2 of the 104 (1.9%) unique policies, representing 1.1% (4/363) of all facilities' policies.

Nearly all policies restricted inpatient visitation. A minority (29/104, 27.9%) prohibited all visitors, whereas most (70/104, 67.3%) identified exceptions permitting some visitation. Few (5/104, 4.8%) policies specified no restrictions. Of the 70 policies allowing exceptions, 37 (52.9%) offered standardized exceptions (eg, end-of-life care, childbirth, and pediatric patients), 16 (22.9%) offered case-by-case exceptions, 13 (18.6%) included both, and 4 (5.7%) included no information about exceptions. Other common rules were restrictions on visitors' movement through the facility and mandatory screening (eg, symptom checks) prior to entry.

There were 8 major functions of phrases within the published policies (Table 1), including communication of rules and exceptions, justification of the restrictive policy, indication that the policy was a deviation from normal procedures, statements of the facilities' values, and a recognition of the hardship that these policies cause for patients and their families. Most (85/104, 81.7%) policies included a justification, most commonly the protection of patients, staff, or community (59/104, 56.7%); guidance or regulations from governing or scientific authorities (34/104, 32.7%); or a scientific rationale (23/104, 22.1%). A minority (38/104, 36.5%) included a statement recognizing that these policies may distress families and patients or that the delivery of family-centered care may be impaired. Governmental agencies at the state or federal level (eg, the Centers for Disease Control and Prevention or the Pennsylvania Department of Health) were named as responsible for the restrictive policies. Policies rarely included an explanation of how restricting visitation relates to viral transmission.

Most policy statements delineating the visitation rules or indicating that the restrictions were a deviation from typical procedures used the passive voice (Table 2). Statements addressing resulting hardship, stating the facilities' values, or emphasizing the temporary nature of the policy used the active voice. Policies that justified restricted visitation on the basis of protection generally used the active voice, whereas those that justified restrictions due to authorities used passive voice.

Table 1. Function of statements within policies.

Function of statement	Example
Indicate that the policy is temporary	"[FACILITY] has put into place a temporary hospital and outpatient visitation policy"
Acknowledge that the policy differs from standard hospital policy	"This policy replaces our traditional open visitation policy"
State the rules themselves	"In end-of-life situations [FACILITY] will allow loved ones to visit with patients"
Justify implementation of the policy	"We are now increasing these restrictions due to federal and state recommendations related to COVID-19"
Recognize the hardship that the policy brings for patients and families	"We know that the no visitor policy may be upsetting for our patients, residents, and their families"
Emphasize the values driving the facility or health system	"The health and well-being of the children and adults we serve remains our top priority in every decision we make"
Identify where patients and potential visitors can obtain additional information	"Please call the [FACILITY] COVID-19 Information Line (XXX-XXX-XXXX) for clarification and additional details"

Table 2. Degree of ownership displayed within and across policy statements.

Statements demonstrating high or low ownership over policy decisions and implementation	Examples
High ownership	
Indicate that the policy is temporary	"We will regularly re-evaluate these visitor restrictions as we receive updates about COVID-19"
Recognize hardship that policy brings for patients and families	"We understand the importance of the support of friends and family to the healing process"
Emphasize values driving the facility or health system	"[FACILITY] is committed to the health, safety, and well-being of the communities we serve"
Offer alternatives to in-person family presence	"We strongly encourage the use of electronic methods to stay connected with loved ones including telemedicine, zoom, and extended phone time"
Justify policy implementation based on protection and safety	"Be assured that we are making these decisions in your best interest so that we can ensure the safety of you, your baby, and our staff"
Low ownership	
Acknowledge that the policy differs from standard hospital policy	"Beginning Tuesday, March 24, inpatient and outpatient visitation guidelines will shift from limited to restricted, outlined below"
State the policy itself	"All in-person visitation has been suspended (with limited critical exceptions)"
Justify the policy based on authority guidance and regulations	"This restriction has been implemented in compliance with updated corporate and state regulations to further reduce the risks associated with COVID-19"

Factorial Experiment of Visitation Policy Communication

In total, 6272 individuals in the Cloud Research Prime Panel system may have viewed the study description, and 1602 individuals may have opened the study link. Of these, 1356 participants completed the instrument for a conservatively estimated response rate of 25.5%. We excluded 34 responses due to failed attention checks. The 1321 participants in the final analysis had a median completion time of 9.2 minutes (IQR 6.4-14.2 minutes). Participants' characteristics are summarized in Table 3. A minority of individuals self-reported high personal risk of COVID-19 or having close contact with an individual at high risk of COVID-19. The majority were insured and demonstrated moderate to low levels of health care system distrust [42].

Results of the multivariable model are displayed in Figure 1 and Table S2 of Multimedia Appendix 1. Visitation policy elements significantly associated with a willingness to recommend the hospital included justifications for restricted visitation based on community protection (OR 1.44, 95% CI 1.24-1.68) or scientific rationale (OR 1.30, 95% CI 1.12-1.51), in contrast to regulations from a governing authority. The hospital displayed a high degree of expressed ownership over the policy (OR 1.16, 95% CI 1.04-1.29 vs low degree of ownership) and inclusion of family-centered care support plans (OR 2.80, 95% CI 2.51-3.12 vs absence of such support). Justification of visitation restrictions to support community protection was significantly more likely to result in willingness to recommend the hospital when compared with scientific rationale for restrictions in the pairwise comparisons from the final model (Figure 1). In the multivariable model, participants living in urban settings, reporting more trust in the health care system, and identifying as Democrats were significantly more willing to recommend the hypothetical hospitals after adjusting for visitation policy elements (Figure 1).

Table 3. Characteristics of survey participants.

Characteristic	Participants (N=1321)[a]
Age (years), median (IQR)	39 (30-55)
Gender, n (%)	
Women	710 (53.7)
Men	601 (45.5)
Transgender, nonbinary, other	10 (0.8)
Race, n (%)	
White, Caucasian American	967 (73.2)
Black, African American	248 (18.8)
Other[b]	106 (8)
Ethnicity, n (%)	
Not Hispanic, Latinx	1195 (90.5)
Hispanic, Latinx	115 (8.7)
Not reported	11 (0.8)
Educational attainment[c], n (%)	
Less than high school, GED[d]	269 (20.4)
Some college	383 (29)
Bachelor's degree	385 (29.1)
Graduate degree	284 (21.5)
Urban or rural residence, n (%)	
Urban	535 (40.5)
Suburban	514 (38.9)
Rural	261 (19.8)
Not reported	11 (0.8)
Political party, n (%)	
Democrat	566 (42.8)
Republican	396 (30)
Other[e]	359 (27.2)
Insurance status, n (%)	
Insured	1207 (91.4)
Uninsured	114 (8.6)
At high risk for COVID-19, n (%)	
No	775 (58.7)
Yes	325 (24.6)
Unsure, not reported	221 (16.7)
Close contact with someone at high risk for COVID-19, n (%)	
No	856 (64.8)
Yes	358 (27.1)
Unsure, not reported	107 (8.1)
Revised HSDS[f] score	
Mean (SD)	22.4 (7.2)
Median (IQR)	24 (17-28)

[a]Percentages may not add up to 100% due to rounding. For gender and HSDS (n=1320), 1 participant did not provide responses.

[b]For race, "Other" includes Asian, Asian American, Pacific Islander, Native American, Hispanic/Latino, Mestizo, Middle Eastern, Romanian, and multiracial.

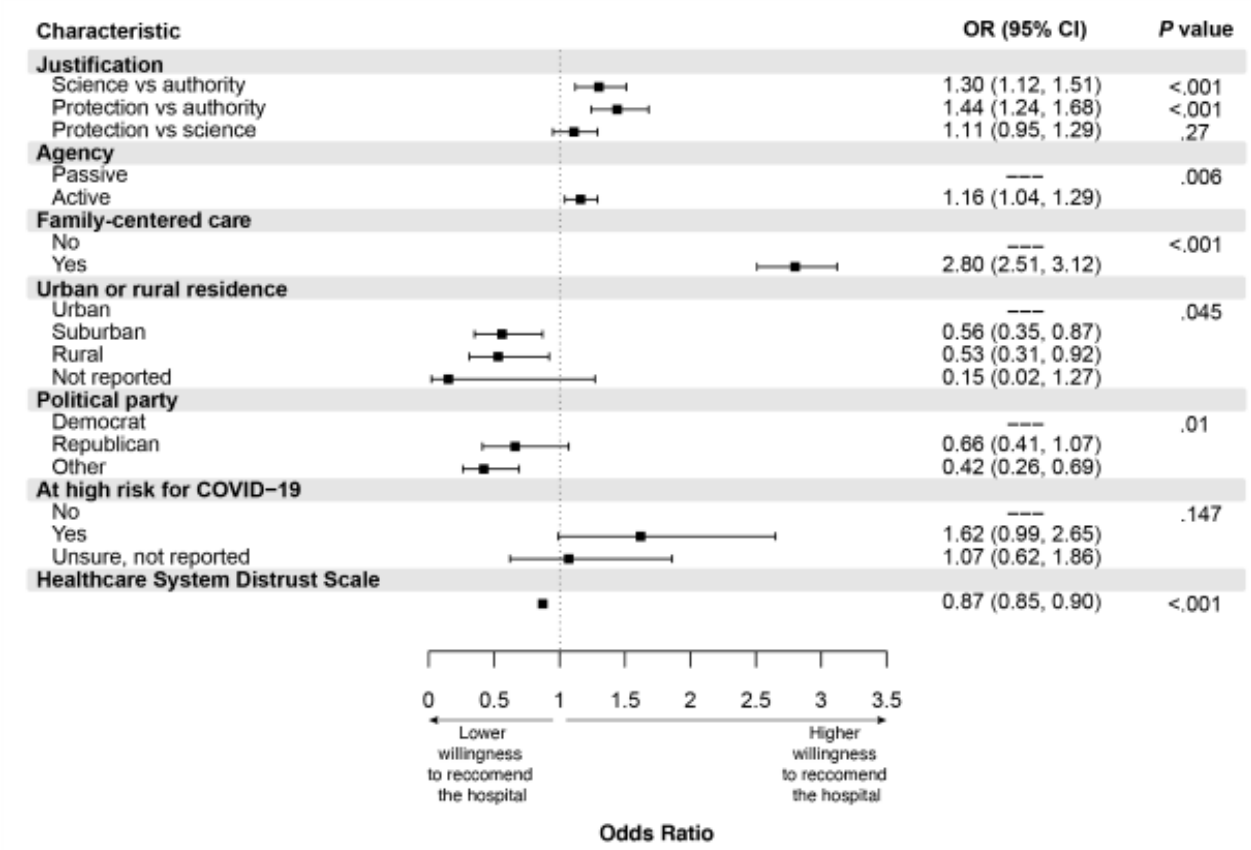
[c]For education attainment, "Some college" includes associate's or professional certificate; "Graduate degree" includes master's, doctoral, or professional degree.

[d]GED: general educational development.

[e]For political party, "Other" includes independent, democratic socialist, libertarian, and no political affiliation.

[f]HSDS: Healthcare System Distrust Scale. For the revised HSDS (9-item scale; range: 5-45), higher total scores indicate more distrust of the health care system.

Figure 1. Association of policy elements and respondent characteristics with willingness to recommend the hospital. OR: odds ratio.



Discussion

Principal Findings

Although health care systems across the globe have restricted inpatient family visitation during the COVID-19 pandemic [3-5], there is a lack of empirical research presenting what visitation rules health systems enacted, how health systems communicated these policies to the lay public, and the impact of those communication choices on public opinion. In this study, we present a mixed methods investigation including qualitative analysis of publicly available COVID-19 era inpatient visitation policies and a subsequent factorial experiment examining the impact of three key policy statement elements on lay individuals' willingness to recommend the hypothetical hospital to a family member. Most of the website-published policies prohibited family presence and justified these restricted visitations on the basis of community protection over authorities' guidance, regulations, or scientific rationale. A few policies addressed how restrictive visitation may impair family-centered care.

Our findings confirm the importance of communication choices during periods of crisis [12,13]. Importantly, we identify specific elements of visitation policy statements strongly associated with individuals' willingness to recommend the hospital. Because these are unrelated to the visitation rules themselves, health systems may rapidly implement these findings.

The goals of family-centered care are to respect family members as care partners, foster collaborations between family members and the health care team, and maintain family integrity [2,10]. There is broad recognition among clinicians and the general public that visitation policies restricting family presence for inpatients are undesirable and detrimental to family-centered care [2,3,5,21-24,47]. Our participants valued the inclusion of family-centered care support plans and recognition of resulting distress in hospitals' communication of visitation restrictions. Indeed, family-centered content was more strongly associated with individuals' willingness to recommend a hospital than any other element. However, few inpatient facilities' policy statements provided resources to promote family members' engagement with admitted patients or health care teams or

expressed empathy and compassion. Including empathy in crisis messages is known to result in a more positive public response and enhances the credibility of the messenger [12]. Health systems should prioritize communication and coping resources for family members distanced from inpatients during and after pandemic conditions and clearly communicate the existence of these resources to engender trust among community members [2].

Inpatient facilities' frequent use of the passive voice to communicate restrictive visitation rules suggests that they seek to distance themselves from unfavorable policies. For example, facilities frequently justified restrictive policies by referring to governing authorities while using the passive voice. When facilities used the active voice to express a high degree of ownership in the policies, they did so for favorable components of the policy statements. These patterns suggest that facilities may attempt to deflect responsibility for undesirable visitation restrictions. However, our findings indicate this is likely counterproductive. Individuals were more likely to recommend a facility that expressed ownership of restrictive policies, including providing a justification that did not rely on governing authorities and using the active voice when communicating restrictive rules. Community members may interpret facilities as being connected and invested in the communities they serve if systems express a high degree of ownership over decision-making, rely on community protection as a salient rationale for restrictions, and support family-centered care despite visitation restrictions.

Many of the inpatient facilities' policy statements did not reflect best practices in community and crisis communication [12]. Many failed to engage in two-way risk communication, in which the public is treated as full participants rather than passive information receivers or rule followers [16,48,49]. Such two-way risk communication strategies focus on building dialogue with public citizens who have rights and values. This can lead to improved public understanding of the risks and rationale for policies and recommendations. Furthermore, this approach to communication positions the rule-makers as transparent and worthy of trust. In contrast, facilities rarely explained the mechanisms through which restrictive family presence policies may reduce the transmission of COVID-19. Particularly early in the pandemic, when we collected the policies, community members likely did not have this scientific knowledge, limiting their ability to participate in the dialogue around visitation restrictions [50]. Additionally, policy statements intending to convey key messages to the general public should be written at a reading level accessible to those with low reading levels, given that the average American reads at an 8th grade reading level [39,40,51]. We elected not to test varied reading levels in our factorial design given the large body of work that exists supporting the need for accessible public health information [51], particularly during the COVID-19 crisis as the burden of impact has been disproportionately high in medically underserved communities [40,52-54]. Policy statements that require a college reading level are likely to contribute to misunderstandings and lack of trust [39,40,51,55]. Health systems should also offer their materials in non-English languages that are spoken by community members, yet this was rarely observed in our sample.

Although the communication of these crisis-era policies should be improved upon, an organization's pre-crisis credibility is critical to the public's trust in the organization during a crisis [12]. If a health system or facility has not already established itself as supportive of families, is not engaged with community health, and is not concerned about staff well-being, the organization's statements of these values may be meaningless. Therefore, the approach to communicating policies during COVID-19 and similar crises should also reflect a broader attention to the relationship between health systems and their surrounding communities both during and between acute public health crises.

Limitations

Our findings should be interpreted in the context of several limitations. First, we gathered policy statements from a single state. However, we did seek to capture all publicly available policies to eliminate selection or response bias introduced with other methods, such as surveys. Second, we captured policy statements immediately following the initial peak of COVID-19 cases in Pennsylvania, so our analysis does not reflect longitudinal changes to the policy statements or visitation rules [26]. We cannot comment on longitudinal trends in visitation policies enacted by inpatient facilities nor public perception of these policies in the context of national or local COVID-19 burden or governmental policies intended to control viral spread. Third, we were unable to find policies for all licensed inpatient facilities in Pennsylvania. However, we used widely available search engines, mitigating the risk that the public had access to an internet-published policy we did not locate. Fourth, the public may not read or compare policies available on websites. Nevertheless, the majority of inpatient facilities had internet-published policies, suggesting they believe there is an audience for this information. Fifth, our factorial experiment relied on participants from an internet research panel, which may not reflect the general US population [36]. We took steps to mitigate the risk of bias, including stratified recruitment on the basis of race and gender [49]. Sixth, we selected elements to vary in the policies for the factorial experiment. There may be alternative elements more important than the ones we selected. However, we selected elements based on our preceding content analysis, making our findings valuable even if other elements are considered important. Seventh, we tested composite policy statements without contextual factors that may lead a community member to recommend or discourage a particular hospital, and we did not require participants to choose between hospitals based on the policy statements, which may have limited variation in their responses. Despite this, we identified several significant findings.

Conclusions

Inpatient facilities frequently enacted restrictive inpatient visitation policies during the COVID-19 pandemic. Communication of these policies did not reflect best practices in crisis communication and may fail to represent health systems' commitment to their communities' health and their motives for establishing restrictive visitation policies. To

improve public perception, health systems and facilities should convey ownership over their policy decisions, justify unfavorable visitation rules based on community protection, and include resources supporting family-centered care. Policies should also use language accessible to those with lower reading levels and in languages other than English as locally relevant, provide clear explanations for novel policies, and more frequently incorporate statements of empathy and compassion when communicating crisis-era policies affecting the general public.

Acknowledgments

Funding for this study was provided the University of Pennsylvania's Leonard Davis Institute of Health Economics.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Timeline of local COVID-19 burden during the study period; survey instrument used; sample size estimates; identification of unique inpatient facilities' visitation policies; and multivariable model.
[DOCX File , 105 KB - jmir_v23i9e28897_app1.docx]

References

1. Kerlin MP, Costa DK, Davis BS, Admon AJ, Vranas KC, Kahn JM. Actions taken by us hospitals to prepare for increased demand for intensive care during the first wave of COVID-19: a national survey. Chest 2021 Aug;160(2):519-528 [FREE Full text] [doi: 10.1016/j.chest.2021.03.005] [Medline: 33716038]
2. Hart JL, Turnbull AE, Oppenheim IM, Courtright KR. Family-centered care during the COVID-19 era. J Pain Symptom Manage 2020 Aug;60(2):e93-e97 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.04.017] [Medline: 32333961]
3. Valley TS, Schutz A, Nagle MT, Miles LJ, Lipman K, Ketcham SW, et al. Changes to visitation policies and communication practices in Michigan ICUs during the COVID-19 pandemic. Am J Respir Crit Care Med 2020 Sep 15;202(6):883-885. [doi: 10.1164/rccm.202005-1706le]
4. Voo TC, Senguttuvan M, Tam CC. Family presence for patients and separated relatives during COVID-19: physical, virtual, and surrogate. J Bioeth Inq 2020 Dec;17(4):767-772 [FREE Full text] [doi: 10.1007/s11673-020-10009-8] [Medline: 32840835]
5. Kuntz JG, Kavalieratos D, Esper GJ, Ogbu N, Mitchell J, Ellis CM, et al. Feasibility and acceptability of inpatient palliative care e-family meetings during COVID-19 pandemic. J Pain Symptom Manage 2020 Sep;60(3):e28-e32 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.06.001] [Medline: 32505643]
6. Robert R, Kentish-Barnes N, Boyer A, Laurent A, Azoulay E, Reignier J. Ethical dilemmas due to the Covid-19 pandemic. Ann Intensive Care 2020 Jun 17;10(1):84 [FREE Full text] [doi: 10.1186/s13613-020-00702-7] [Medline: 32556826]
7. Suresh D, Flatley K, McDonough M, Cochran-Caggiano N, Inglis P, Fordyce S, et al. Providing compassionate care: the role of medical students and videoconference technology in the COVID-19 pandemic. J Patient Exp 2020 Dec;7(6):1002-1006 [FREE Full text] [doi: 10.1177/2374373520978873] [Medline: 33457538]
8. Hospital Visitation – Phase II Visitation for Patients who are Covid-19 Negative. Centers for Medicare and Medicaid. 2020 Jun 26. URL: https://www.cms.gov/files/document/covid-hospital-visitation-phase-ii-visitation-covid-negative-patients.pdf [accessed 2021-08-27]
9. Management of Visitors to Healthcare Facilities in the Context of COVID-19: Non-US Healthcare Settings. Centers for Disease Control and Prevention. 2020 Sep 15. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/non-us-settings/hcf-visitors.html [accessed 2021-08-27]
10. Davidson JE, Aslakson RA, Long AC, Puntillo KA, Kross EK, Hart J, et al. Guidelines for family-centered care in the neonatal, pediatric, and adult ICU. Crit Care Med 2017 Jan;45(1):103-128. [doi: 10.1097/CCM.0000000000002169] [Medline: 27984278]
11. Hart JL, Taylor SP. Family presence for critically ill patients during a pandemic. Chest 2021 Aug;160(2):549-557 [FREE Full text] [doi: 10.1016/j.chest.2021.05.003] [Medline: 33971149]
12. Seeger MW. Best practices in crisis communication: an expert panel process. J Appl Commun Res 2006 Aug;34(3):232-244. [doi: 10.1080/00909880600769944]
13. MacLeod A. The impact of communication on human behaviour in times of crisis. J Bus Contin Emer Plan 2014;8(2):134-140. [Medline: 25416375]
14. Yin F, Wu Z, Xia X, Ji M, Wang Y, Hu Z. Unfolding the determinants of COVID-19 vaccine acceptance in China. J Med Internet Res 2021 Jan 15;23(1):e26089 [FREE Full text] [doi: 10.2196/26089] [Medline: 33400682]

15. Gesser-Edelsburg A, Cohen R, Hijazi R, Abed Elhadi Shahbari N. Analysis of public perception of the Israeli government's early emergency instructions regarding COVID-19: online survey study. J Med Internet Res 2020 May 15;22(5):e19370 [FREE Full text] [doi: 10.2196/19370] [Medline: 32392172]
16. Palenchar MJ, Heath RL. Strategic risk communication: adding value to society. Public Relations Review 2007 Jun;33(2):120-129. [doi: 10.1016/j.pubrev.2006.11.014]
17. Jong W. Evaluating crisis communication. a 30-item checklist for assessing performance during COVID-19 and other pandemics. J Health Commun 2020 Dec 01;25(12):962-970. [doi: 10.1080/10810730.2021.1871791] [Medline: 33455526]
18. Ophir Y. Coverage of epidemics in American newspapers through the lens of the crisis and emergency risk communication framework. Health Secur 2018;16(3):147-157. [doi: 10.1089/hs.2017.0106] [Medline: 29870279]
19. Veil S, Reynolds B, Sellnow TL, Seeger MW. CERC as a theoretical framework for research and practice. Health Promot Pract 2008 Oct;9(4 Suppl):26S-34S. [doi: 10.1177/1524839908322113] [Medline: 18936257]
20. Ngai CSB, Singh RG, Lu W, Koon AC. Grappling with the COVID-19 health crisis: content analysis of communication strategies and their effects on public engagement on social media. J Med Internet Res 2020 Aug 24;22(8):e21360 [FREE Full text] [doi: 10.2196/21360] [Medline: 32750013]
21. Huff C. Families of sick covid-19 patients may develop physical, mental symptoms of their own. The Washington Post. 2020 Oct 25. URL: https://www.washingtonpost.com/health/covid-patient-families-grief-anxiety/2020/10/23/49db919e-028e-11eb-a2db-417cddf4816a_story.html [accessed 2021-08-27]
22. Hafner K. 'A Heart-wrenching thing': Hospital bans on visits devastate families. The New York Times. 2020 Mar 29. URL: https://www.nytimes.com/2020/03/29/health/coronavirus-hospital-visit-ban.html [accessed 2021-08-27]
23. Navathe AS, Liao JM. Barring visitors from hospitals can worsen patient care during COVID. The Philadelphia Inquirer. 2020 Jul 27. URL: https://www.inquirer.com/opinion/commentary/hospital-visits-coronavirus-philadelphia-pennsylvania-regulations-20200727.html [accessed 2021-08-27]
24. Awdish R, Ely E. Keeping loved ones from visiting our coronavirus patients is making them sicker. The Washington Post. 2020 Aug 06. URL: https://www.washingtonpost.com/outlook/2020/08/06/coronavirus-icu-patients-families/ [accessed 2021-08-31]
25. Tashakkori A, Teddlie C. Handbook of Mixed Methods in Social & Behavioral Research. Location: Sage Publications, Inc; 2010.
26. Pennsylvania Department of Health COVID-19 Dashboard. URL: https://www.health.pa.gov/topics/disease/coronavirus/Pages/Cases.aspx [accessed 2021-08-03]
27. Onega T, Weiss JE, Alford‐Teaster J, Goodrich M, Eliassen MS, Kim SJ. Concordance of Rural‐Urban Self‐identity and ZIP Code‐Derived Rural‐Urban Commuting Area (RUCA) designation. J Rural Health 2020 Mar;36(2):274-280. [doi: 10.1111/jrh.12364]
28. Snyder JE, Jensen M, Nguyen NX, Filice CE, Joynt KE. Defining rurality in Medicare administrative data. Med Care 2017 Dec;55(12):e164-e169. [doi: 10.1097/MLR.0000000000000607] [Medline: 29135781]
29. Lenzner T. Are readability formulas valid tools for assessing survey question difficulty? Sociological Methods & Research 2013 Dec 10;43(4):677-698. [doi: 10.1177/0049124113513436]
30. Zhang M, Chow A, Smith H. COVID-19 contact-tracing apps: analysis of the readability of privacy policies. J Med Internet Res 2020 Dec 03;22(12):e21572 [FREE Full text] [doi: 10.2196/21572] [Medline: 33170798]
31. Forman J. Qualitative content analysis. In: Damschroder J, Liva J, Laura AS, editors. Empirical Methods for Bioethics: A Primer, Volume 11 (Advances in Bioethics). Amsterdam, Boston: Elsevier JAI; 2008.
32. Rapley T. Doing Conversation, Discourse and Document Analysis. Thousand Oaks, CA: Sage Publications; 2007.
33. Wood L, Kroger R. Doing Discourse Analysis: Methods for Studying Action in Talk and Text. Thousand Oaks, CA: Sage Publications; May 17, 2000.
34. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
35. Saldana J. The Coding Manual for Qualitative Researchers. Third Edition. Thousand Oaks, CA: Sage Publications; 2015.
36. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. Nurs Health Sci 2013 Sep;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]
37. Guest G, MacQueen KM, Namey EE. Applied Thematic Analysis. Newbury Park, CA: Sage Publications; 2012.
38. Charmaz K. Constructing grounded theory. Thousand Oaks, CA: Sage Publications; 2014.
39. Cotugna N, Vickery CE, Carpenter-Haefele KM. Evaluation of literacy level of patient education pages in health-related journals. J Community Health 2005 Jun;30(3):213-219. [Medline: 15847246]
40. Szmuda T, Özdemir C, Ali S, Singh A, Syed M, Słoniewski P. Readability of online patient education material for the novel coronavirus disease (COVID-19): a cross-sectional health literacy study. Public Health 2020 Aug;185:21-25. [doi: 10.1016/j.puhe.2020.05.041]
41. Chandler J, Rosenzweig C, Moss AJ, Robinson J, Litman L. Online panels in social science research: expanding sampling methods beyond Mechanical Turk. Behav Res Methods 2019 Oct;51(5):2022-2038 [FREE Full text] [doi: 10.3758/s13428-019-01273-7] [Medline: 31512174]

42. Armstrong K, Rose A, Peters N, Long JA, McMurphy S, Shea JA. Distrust of the health care system and self-reported health in the United States. J Gen Intern Med 2006 Apr;21(4):292-297. [doi: 10.1111/j.1525-1497.2006.00396.x]
43. Paas LJ, Morren M. Please do not answer if you are reading this: respondent attention in online panels. Mark Lett 2018 Feb 1;29(1):13-21. [doi: 10.1007/s11002-018-9448-7]
44. Hosmer Jr DW, Lemeshow S, Sturdivant R. Applied Logistic Regression, Third Edition. Hoboken, NJ: John Wiley & Sons, Inc; Mar 2013.
45. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015 Jul;67:1-48 [FREE Full text] [doi: 10.18637/jss.v067.i01]
46. Hosmer D, Lemeshow S, Sturdivant R. Model-building strategies methods for logistic regression. In: Applied Logistic Regression, Third Edition. Hoboken, NJ: John Wiley & Sons, Inc; Mar 2013:89-151.
47. Murray PD, Swanson JR. Visitation restrictions: is it right and how do we support families in the NICU during COVID-19? J Perinatol 2020 Aug 08;40(10):1576-1581. [doi: 10.1038/s41372-020-00781-1]
48. Guan B, Bao G, Liu Q, Raymond RG. Two-way risk communication, public value consensus, and citizens' policy compliance willingness about COVID-19: multilevel analysis based on a nudge view. Administration & Society 2021 Feb 09;53(7):1106-1149. [doi: 10.1177/0095399721990332]
49. Benavides AD, Nukpezah J, Keyes LM, Soujaa I. Adoption of multilingual state emergency management websites: Responsiveness to the risk communication needs of a multilingual society. International Journal of Public Administration 2020 Feb 19;44(5):409-419. [doi: 10.1080/01900692.2020.1728549]
50. Geldsetzer P. Knowledge and perceptions of COVID-19 among the general public in the United States and the United Kingdom: a cross-sectional online survey. Ann Intern Med 2020 Jul 21;173(2):157-160 [FREE Full text] [doi: 10.7326/M20-0912] [Medline: 32196071]
51. Me K, Greenberg E, Jin Y, Paulsen C. The Health Literacy of America's Adults Results From the 2003 National Assessment of Adult Literacy. National Center for Education Statistics. 2006. URL: https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006483 [accessed 2021-08-27]
52. Williams DR, Cooper LA. COVID-19 and health equity-a new kind of "Herd Immunity". JAMA 2020 May 11;323(24):2478-2480. [doi: 10.1001/jama.2020.8051] [Medline: 32391852]
53. Dorn AV, Cooney RE, Sabin ML. COVID-19 exacerbating inequalities in the US. Lancet 2020 Apr 18;395(10232):1243-1244 [FREE Full text] [doi: 10.1016/S0140-6736(20)30893-X] [Medline: 32305087]
54. Wiemers EE, Abrahams S, AlFakhri M, Hotz VJ, Schoeni RF, Seltzer JA. Disparities in vulnerability to severe complications from COVID-19 in the United States. medRxiv. Preprint posted online on June 22, 2020 [FREE Full text] [doi: 10.1101/2020.05.28.20115899] [Medline: 32511522]
55. Miller-Matero LR, Clark KB, Brescacin C, Dubaybo H, Willens DE. Depression and literacy are important factors for missed appointments. Psychol Health Med 2016 Sep;21(6):686-695. [doi: 10.1080/13548506.2015.1120329] [Medline: 26695719]

Abbreviations

CERC: Crisis and Emergency Risk Communication
HSDS: Healthcare System Distrust Scale
RUCA: Rural-Urban Commuting Area

Edited by C Basch; submitted 18.03.21; peer-reviewed by R Shahriarirad, S Molani, K Ludlow; comments to author 03.07.21; revised version received 03.08.21; accepted 03.08.21; published 24.09.21.

Please cite as:
Hart J, Summer A, Yadav KN, Peace S, Hong D, Konu M, Clapp JT
Content and Communication of Inpatient Family Visitation Policies During the COVID-19 Pandemic: Sequential Mixed Methods Study
J Med Internet Res 2021;23(9):e28897
URL: https://www.jmir.org/2021/9/e28897
doi:10.2196/28897
PMID:34406968

Original Paper

Comparing Public Sentiment Toward COVID-19 Vaccines Across Canadian Cities: Analysis of Comments on Reddit

Cathy Yan[1], BSc; Melanie Law[2], BSc; Stephanie Nguyen[3], BASc; Janelle Cheung[4], BSc; Jude Kong[5], BASc, MSc, PhD

[1]Department of Genome Science and Technology, University of British Columbia, Vancouver, BC, Canada
[2]Department of Microbiology and Immunology, University of British Columbia, Vancouver, BC, Canada
[3]Department of Biomedical Engineering, University of British Columbia, Vancouver, BC, Canada
[4]Department of Biochemistry, University of British Columbia, Vancouver, BC, Canada
[5]Department of Mathematics & Statistics, York University, Toronto, ON, Canada

Corresponding Author:
Jude Kong, BASc, MSc, PhD
Department of Mathematics & Statistics
York University
Ross 533N
4700 Keele Street
Toronto, ON, M3J 1P3
Canada
Phone: 1 416 736 2100 ext 66093
Email: jdkong@yorku.ca

Abstract

Background: Social media enables the rapid consumption of news related to COVID-19 and serves as a platform for discussions. Its richness in text-based data in the form of posts and comments allows researchers to identify popular topics and assess public sentiment. Nonetheless, the vast majority of topic extraction and sentiment analysis based on social media is performed on the platform or country level and does not account for local culture and policies.

Objective: The aim of this study is to use location-based subreddits on Reddit to study city-level variations in sentiments toward vaccine-related topics.

Methods: Comments on posts providing regular updates on COVID-19 statistics in the Vancouver (r/vancouver, n=49,291), Toronto (r/toronto, n=20,764), and Calgary (r/calgary, n=21,277) subreddits between July 13, 2020, and June 14, 2021, were extracted. Latent Dirichlet allocation was used to identify frequently discussed topics. Sentiment (joy, sadness, fear, and anger) scores were assigned to comments through random forest regression.

Results: The number of comments on the 250 posts from the Vancouver subreddit positively correlated with the number of new daily COVID-19 cases in British Columbia (R=0.51, 95% CI for slope 0.18-0.29; P<.001). From the comments, 13 topics were identified. Two were related to vaccines, 1 regarding vaccine uptake and the other about vaccine supply. The levels of discussion for both topics were linked to the total number of vaccines administered (Granger test for causality, P<.001). Comments pertaining to either topic displayed higher scores for joy than for other topics (P<.001). Calgary and Toronto also discussed vaccine uptake. Sentiment scores for this topic differed across the 3 cities (P<.001).

Conclusions: Our work demonstrates that data from city-specific subreddits can be used to better understand concerns and sentiments around COVID-19 vaccines at the local level. This can potentially lead to more targeted and publicly acceptable policies based on content on social media.

(J Med Internet Res 2021;23(9):e32685) doi:10.2196/32685

KEYWORDS

COVID-19; public sentiment; social media; Reddit; Canada; communication; sentiment; opinion; emotion; concern; pandemic; vaccine; hesitancy

Introduction

Sixty-five percent of approximately 3.8 billion internet users are currently informed about top news stories from social media platforms such as Twitter, Facebook, and Reddit rather than traditional news outlets [1]. Surveys show that this trend has been especially apparent during the COVID-19 pandemic as more people seek timely updates on the crisis [2]. When used effectively, social media platforms can disseminate relevant health-related information to users such as patients, clinicians, and scientists [1]. However, the unfamiliarity of COVID-19 has led to the frequent transmission of false and conflicting information [3]. While awareness of "fake news" is high among Gen Z individuals and Millennials, less than a quarter of them report posts with false information, and only 8.7% opt to stop receiving updates from the account that produces misleading posts [4]. Propagated misinformation also negatively impacts compliance with public health policies such as social distancing [5].

In addition to what people express on social media, investigation of their underlying attitudes in conjunction with their comments can be key to determine political participation and predict "protester violence" [6,7]. In the context of the pandemic, posts on Twitter and Facebook have been used to examine attitudes toward contact tracing apps in the United Kingdom [8], and a dashboard was built to track emotions in Austria on the basis of the news platform derstandard.at, Twitter, and a chat platform for students [9].

Currently, one of the most critical steps to reducing the spread of COVID-19 is mass vaccination [10]. Unfortunately, social media also provides a platform for growing antivaccination movements and increasing vaccine hesitancy [11-13]. When examining the expression of these opinions that oppose scientific advice, studies capture sentiments across entire platforms or whole countries but fail to capture nuances at a more local level. Thus, we sought to use comments on Reddit to explore discussions surrounding COVID-19 in Toronto (Ontario), Calgary (Alberta), and Vancouver (British Columbia) as sentiment analysis can contribute to improving social management practices within each city.

Reddit is divided into subreddits, which contain posts and discussions relevant to a particular location or topic. As a platform that quickly aggregates content, it effectively disseminates the latest news regarding major events, including the COVID-19 pandemic. For example, daily posts on the Vancouver subreddit (r/vancouver), by the user cyclinginvancouver (u/cyclinginvancouver), provide updated statistics regarding the spread of COVID-19 in British Columbia. Similar posts can be found on the Calgary (r/calgary) and Toronto (r/toronto) subreddits. The posts themselves are unbiased, containing only information such as the number of new cases of COVID-19 infection as determined by a positive test, hospitalizations, vaccinations, and deaths. Thus, people are able to engage in free-form discussion in the comments. Comparatively, other platforms such as Twitter allows a global community to discuss the pandemic; hence, discussions are less specific to regional communities.

The aim of this study is to use location-based subreddits on Reddit to study city-level variations in sentiments toward COVID-19 vaccine–related topics. Here, we first present an analysis of comments from the Vancouver subreddit to understand the topics being discussed and people's attitudes toward local policies. We also specifically explore people's reactions and sentiments toward vaccines and how they align with vaccination rates. Then, we characterize differences in vaccine topics and sentiments among Vancouver, Toronto, and Calgary. Data from similar studies performed using Twitter [14] helped build some of our models, as our study is the first to examine comments from Reddit in this manner. Our novel approach of examining topics of local concern have the potential to inform how public policy can be tailored to specific geographical regions.

Methods

Data Collection

All analyses were performed using Python (version 3.7) and R (version 4.0.2). Our code can be found on GitHub at Mellaw/BDC_Reddit.

Reddit comments were acquired using the Python Reddit API Wrapper (PRAW; version 7.2.0) [15]. A read-only Reddit instance was created and used to obtain a subreddit instance for the Vancouver, Calgary, and Toronto subreddits (r/vancouver, r/calgary, and r/toronto, respectively). Relevant posts were acquired by searching the keywords *Covid-19 Update -*, *Alberta Totals:*, and *COVID-19 in Ontario* in the Vancouver, Calgary, and Toronto subreddits, respectively. Keywords were selected to be specific to posts providing daily updates on COVID-19 statistics. The title, time created, submission ID, and author were extracted and assembled into a data frame using pandas (version 1.2.4) [16]. The submission ID is a string unique to each post. It was used as an input to PRAW for interacting with each post's "CommentForest" or a list of comments and replies. The list method of CommentForest was used to extract the body text, author, and title of the post for all comments.

Twitter data were obtained from the paper "Global Reactions to COVID-19 on Twitter: A Labelled Dataset with Latent Topic, Sentiment and Emotion Attributes" [14] and were made available on OpenICPSR [17]. We used the version with 5000 tweets randomly sampled from the full data set of 132.1 million tweets. The tweets are in English and are from the United States, Singapore, India, and Brazil. Every tweet was scored on how intensely they demonstrated the emotions anger, fear, sadness, and joy on a scale from 0 to 1. Since the data set only provided the IDs of tweets, we used Tweepy (version 3.10.0) to hydrate the tweets [18].

We also used statistics for new cases of COVID-19 and total vaccinations in British Columbia, Alberta, and Ontario. These data, as well as numbers for fatalities, hospitalizations, tests, and recoveries, were provided through a web-based COVID-19 Tracker [19], where real-time data at both the national and provincial level were collected by volunteers. To ensure accuracy, statistics were updated primarily from live press

briefings, with some supplementation from news networks. Source URLs were also provided.

Data Processing

Preprocessing Raw Text

All Reddit comments and Tweets were converted to lowercase, and the Python software package [20] was used to remove nonalphabetical characters, URLs, and references to other users. Texts were then lemmatized using spaCy (version 3.0) [21] and stripped of stop words using Natural Language Toolkit (NLTK; version 3.6.2) [22]. The list of stop words was modified by removing "no" and "not."

Topic Extraction

Topic extraction was performed for Reddit comments by implementing Latent Dirichlet Allocation (LDA) through gensim (version 4.0.1) [23]. After tokenizing, comments were used to create a dictionary, which maps every word to a unique integer ID. This was then converted into a bag-of-words format, essentially counting how many times a word is used. To optimize the number of topics, LDA models were created for 1 to 14 topics. Other nondefault parameters include setting the chunk size to be the number of comments, the number of passes to 10, and α to "auto."

The optimal number of topics was deemed to be the one that minimizes mean Jaccard similarity while maximizing mean coherence across topics. Jaccard similarity is a metric for how many words 2 documents have in common [24]. If 2 topics were identical, their Jaccard similarity would be 1. Conversely, if 2 topics shared no words in common and were thus entirely distinct, their Jaccard similarity would be 0. Coherence for each topic was calculated with gensim using the "c_v" option. It measures the semantic similarity between high-scoring words, which is a function of how often the words co-occur across comments [25].

Each comment was quantitatively assessed to what extent they addressed each topic with a score from 0 to 1. For downstream analyses, a comment was considered to address a topic ("1") if the score was greater than 0.2 and not ("0") otherwise. Word clouds showing the top keywords for each topic were generated using wordcloud (version 1.8.1) and Matplotlib (version 3.4.2). Thus, we were able to identify vaccine-related comments.

Sentiment Analysis

Scores for emotional intensity were assigned to Reddit comments using a model built from the Twitter data set. Using scikit-learn (version 0.24), the Twitter data set was split into training (0.75) and testing (0.25) sets. Prior to modeling, the text of the tweets was processed into a data frame containing term frequency–inverse document frequencies (TF–IDF). The term frequency (TF) is the number of times a word appears in each tweet divided by the total number of words in that tweet. The inverse document frequency (IDF) is the logarithm of the total number of tweets divided by the number of tweets containing the word. TF–IDF is simply the TF multiplied by the IDF.

The TF–IDF data frame was used as the input to a random forest regression model. The model from scikit-learn was used with default parameters. The scores for anger, fear, sadness, and joy were the target variables. A total of 4 models were fitted, 1 for each emotion. To evaluate the models, the data in the TF–IDF data frame for the test data set were used as the predictors, and the root mean square error (RMSE) was used to compare model outputs with true values. The same processing workflow and models were applied to Reddit comments to predict emotional intensity.

Statistics

The correlation between the number of Reddit comments and new cases of COVID-19 was calculated using Pearson correlation in R. Time-series trends for the level of vaccine-related discussion and total vaccinations in British Columbia were compared using the Granger causality test from the lmtest package [26]. Comparisons across topics and cities were performed using the Mann–Whitney *U* test.

Results

Data Overview

From the r/vancouver subreddit, 49,291 comments across 250 daily update posts were obtained between July 13, 2020, and June 14, 2021. After preprocessing, 433 comments were found to be duplicated. To avoid including spam, duplicates from technical glitches, and placeholder text for deleted and removed comments, duplicates were excluded from further analysis, yielding 45,303 usable comments. These comments were contributed by 4261 users ("Redditors"). Summary statistics for r/calgary and r/toronto are also displayed in Table 1.

Table 1. Summary statistics for posts and comments extracted.

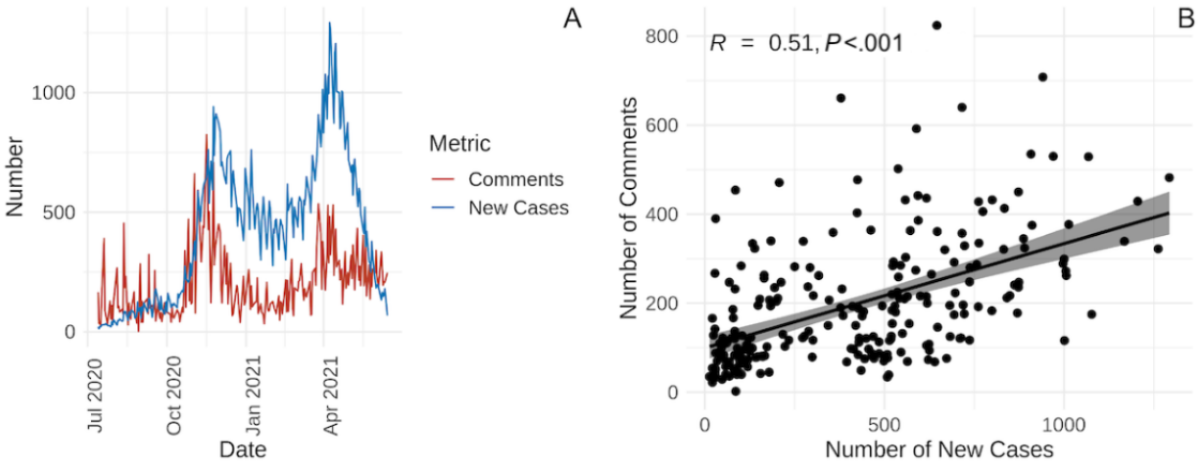
City	Total posts, n	Total comments, n	Usable comments, n
Vancouver	49,291	250	45,303
Toronto	20,764	234	19,105
Calgary	21,277	249	18,886

Vancouver-Specific Analyses

Engagement Level on Reddit Correlates With Daily New COVID-19 Cases

The number of new COVID-19 cases in British Columbia has 2 distinct peaks (Figure 1A). The first is in late November 2020, following a period of low, steady numbers in the summer, with exponential growth beginning in October 2020. The second is in early April 2021, after a local minimum in February 2021, and preceding a steep decline. The number of comments on daily update posts demonstrates a similar trend over time and was found to be significantly correlated with the number of new COVID-19 cases (R=0.51, 95% CI for slope 0.18-0.29; P<.001) (Figure 1B).

Figure 1. (A) Line plot depicting the number of new COVID-19 cases in British Columbia (blue) and the number of comments on each daily update post (red) from July 13, 2020, to June 14, 2021. (B) The number of new COVID-19 cases is significantly positively correlated with the number of comments on daily update posts (R=0.51; P<.001).



Thirteen Main Topics Related to COVID-19 Were Identified From Reddit Comments

By maximizing coherence and minimizing Jaccard similarity, the ideal number of topics was deemed to be 13 (Multimedia Appendix 1). Based on the word clouds (Multimedia Appendix 2) and examples of the first 25 words for the highest scoring comment for each topic (Table 2), the topics are as follows: (1) advocating for restrictions, (2) COVID-19 transmission, (3) impacts of COVID-19 on social spheres, (4) discussion about case numbers, (5) outbreaks in health care facilities, (6) debating how realistic public health orders are, (7) scientific concepts surrounding COVID-19, (8) monitoring travelers and people who have been exposed, (9) violating and enforcing restrictions, (10) vaccine uptake, (11) general speculations, (12) impact on hospitals, and (13) vaccine scarcity.

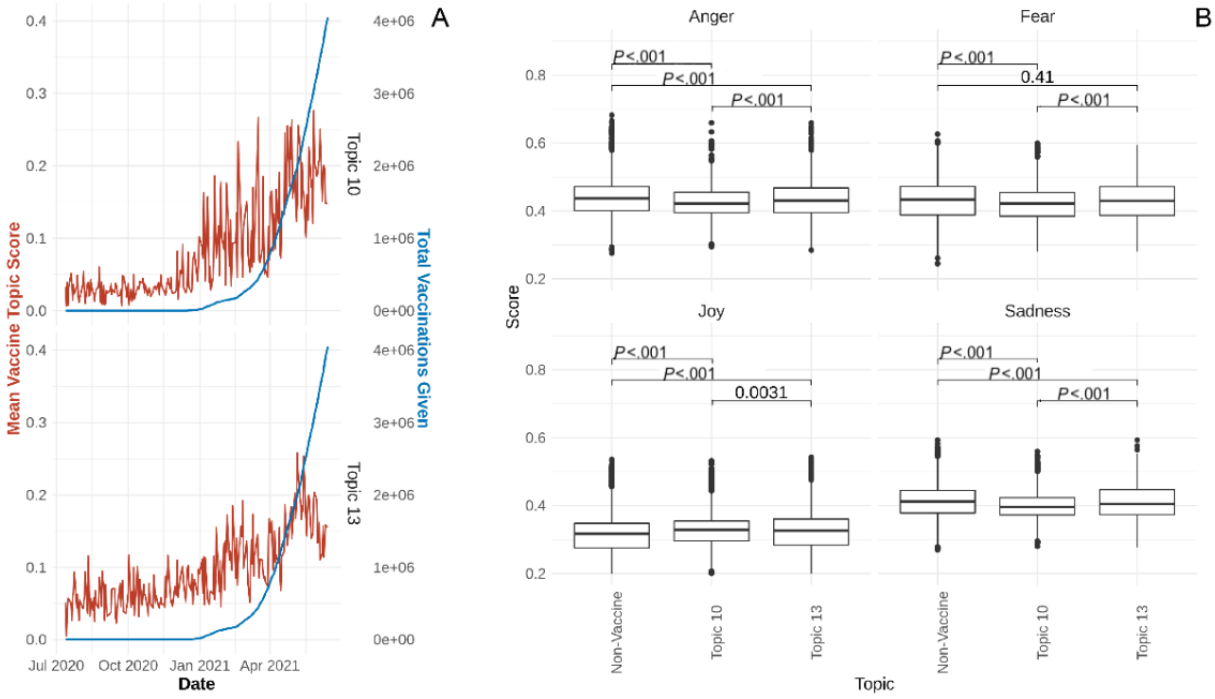
Table 2. Examples of processed comments with the highest score for each topic.

Topic	Processed comment
1	“well nt make zero case still strong restriction truck driver example implement policy nt truck unload instead local worker local worker also test daily one”
2	“not interpretation public health monitoring figure individual direct self isolate day know exposure identify contact tracing question yesterday student return class leave self isolation day”
3	“problem collective reason spread happen exponentially community strong social tie custom obligation ie wedding season lot young adult ca nt live together prior marriage taboo parent”
4	“monday multi day count new case consistent trend flat growth plateaue no acceleration no deceleration seven day trail average basically unchanged per day value near”
5	“six new healthcare facility outbreak braddan private hospital kin village madison care centre royal city manor william lake senior village creekside land outbreak chilliwack“
6	“isolation not fast literally never eliminate virus not canada certainly not globally country temporarily locally eliminate it which coincidentally island nation even island invariably virus introduce”
7	“vaccine currently use new mrna base one reprogram surface marker protein make machinary cell make virus spike protein immune system detect build resistance virus”
8	“people active public health monitoring result identify exposure know case active case recover case vancouver coastal health region case fraser health region case interior health region”
9	“nt see enforcement feasible transit security every bus take last night home bus driver keep tell people bus full pull ahead stop rather let people”
10	“reassurance calculation bc receive vaccine dose percentage first dose adult population date thursday usually shipment dose cumulative march dose march dose march dose april st dose”
11	“selectively ignore scientist scientist science deem trustworthy work right we medium company not part pov actually know people like answer question base see yes no”
12	“nine hospital bc vancouver coastal fraser health move emergency surgery least next two week mean combine elective surgery cancel low mainland fraser health abbotsford burnaby surrey memorial”
13	“parent age range little hesitant well reason azd vaccine cautious approach sure pfizer moderna slightly high effectiveness rating age group still month month half away”

Increase in Vaccine-Related Discussion Correlates With the Number of People Vaccinated

Specifically focusing on discussions surrounding vaccines, we see that the topic scores for topics 10 and 13 began to trend upward starting in January 2021, with topic 13 being slightly more highly discussed before then (Figure 2A). When overlaid with data on the total number of people vaccinated in British Columbia [19], we see that the rise in discussion precedes the rise in the number of vaccinations, although both trend similarly. The Granger test for causality was performed to evaluate the impact of vaccine discussion on total vaccinations. The results were significant for both topics (P<.001).

Figure 2. (A) Line plot displaying the daily average vaccine topic score (red) for topics 10 (top) and 13 (bottom) and the total number of vaccines administered up to that date (blue). (B) Box plots showing the distribution of emotional intensity scores. Comparisons between groups were made using the Mann–Whitney *U* test.



Vaccine-Related Comments Express Significantly Higher Positive Sentiments

Random forest regression models built using tweets labeled with emotional intensity scores were evaluated using RMSE values (Multimedia Appendix 3). Even when applied to Reddit comments, the model appears to retain validity. For each emotion, the processed comment with the highest score is displayed in Table 3. Additionally, the negative emotions (sadness, fear, and anger) are significantly, strongly positively correlated with each other and negatively correlated with joy (Multimedia Appendix 3).

Table 3. Examples of processed comments with the highest intensity scores for their respective emotions.

Emotion	Score	Processed comment
Joy	0.54	“depend happy plateau hit case day long time okay case day stay steady thing look cautiously optimistic definitely call”
Sadness	0.59	“not scientific datum notice friend indocanadian community really enthusiastic vaccine due really sad outlook india covid situation right lot family back india truly suffer take quite seriously”
Anger	0.68	“fucking sick people not give fuck people understand pass around hospital shit show soon”
Fear	0.63	“care home care home interior home town announce outbreak today mom nurse different care home town worried transmission partner lose grandparent already quarantine non covid today two fear come true sick scare sorry”

From July 2020 to April 2021, the emotional intensity scores were steady for all emotions (Multimedia Appendix 3). Negative emotions were expressed more than joy, with anger being the most prominent. From April 2021 onward, the mean scores for joy begin to trend upward while those for negative emotions all trend downward. Investigating the differences in emotional intensity scores between vaccine-related and non–vaccine-related comments, we found that comments for both vaccine topics had significantly higher scores for joy (P<.001) (Figure 2B). However, comments regarding vaccine availability had significantly lower scores for joy and higher scores for negative emotions (P<.001) compared to comments about vaccine uptake.

Comparison Across Cities

Stronger Positive-Sentiment is Exhibited in Vaccine-Related Comments Across Several Canadian Cities

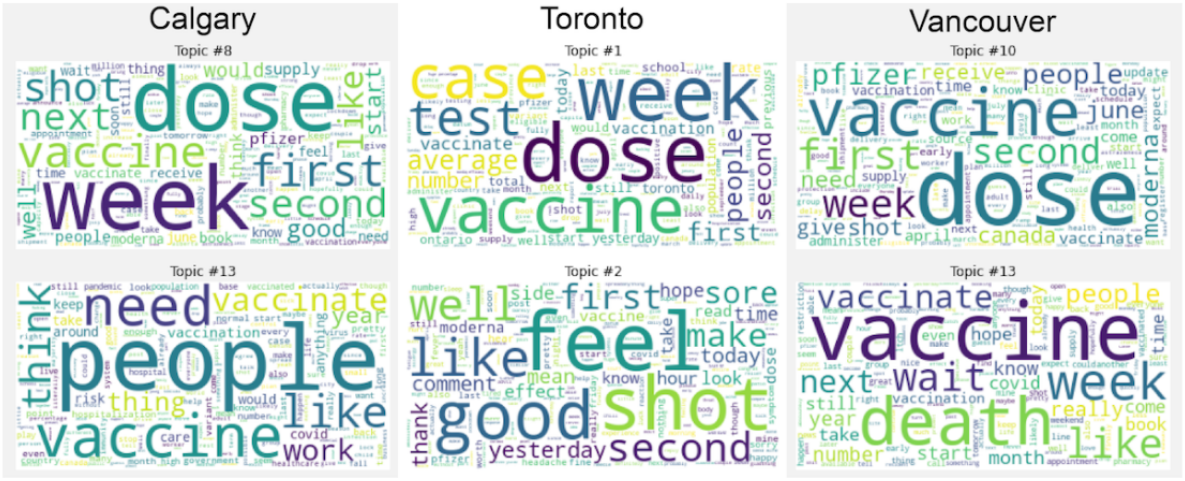
To elucidate the effect of public sentiment on vaccination rates, 2 additional major Canadian cities, Toronto and Calgary, were analyzed using the same approach taken for Vancouver. A total of 13 distinct discussion topics were identified in r/toronto and 14 topics in r/calgary. Both had 2 dominant vaccine-related topics (Figure 3). In Toronto, these discussed vaccine uptake and postvaccine feelings. In r/calgary the 2 vaccine-related topics identified were vaccine uptake and concerns around vaccination rates. Top-scoring comments for each

vaccine-related topic for Toronto and Calgary are shown in Multimedia Appendix 4.

Public sentiments for all emotions in Toronto and Calgary were significantly different between the 2 vaccine-related topics and between vaccine-related and non–vaccine-related comments (P<.001), except for the expression of fear in r/calgary comments (Multimedia Appendix 5). No significance was detected in the degree of fear between non–vaccine-related comments and those that discussed first and second dosages of COVID-19 vaccines.

In the Calgary Reddit community, comments discussing vaccination rates expressed lowest intensity for joy and the highest score for negative emotions (Multimedia Appendix 5). This coincides with Alberta having lower vaccination rates than British Columbia and Ontario (Multimedia Appendix 5). In Toronto, a higher degree of positive sentiment was observed in the vaccine-related comments (P<.001), with the highest median score occurring in comments that discussed vaccine side effects, followed by those pertaining to vaccine uptake (Multimedia Appendix 5).

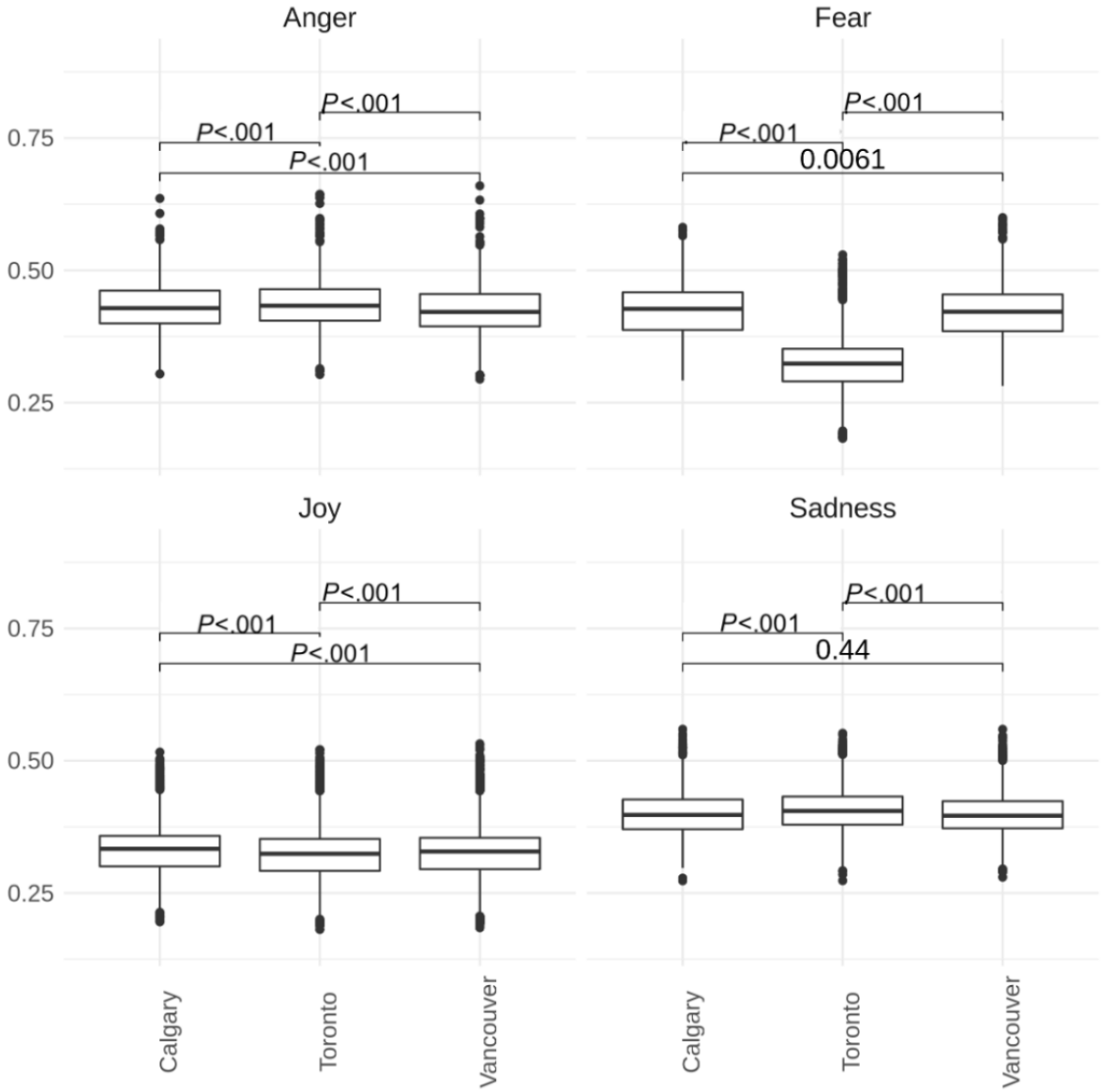
Figure 3. Word clouds for vaccine-related topics across the 3 cities.



Sentiments Toward Vaccine Uptake Significantly Differed Across Cities

Sentiments toward vaccine uptake, the only vaccine-related topic shared by all 3 cities, differed significantly among cities across all emotions (P<.001; Figure 4). Toronto had the highest scores for anger and sadness and the lowest scores for joy and fear. Vancouver and Calgary had statistically the same scores for sadness, but Calgary had significantly higher scores for the other emotions.

Figure 4. Box plots showing the distribution of emotional intensity scores for vaccine uptake. Comparisons between cities were made using the Mann–Whitney *U* test.



Discussion

Principal Findings

In this study, we analyzed comments on the posts in r/vancouver, r/calgary, and r/toronto, which provide daily updates on case numbers, hospitalizations, and other COVID-19–related statistics. We found that the number of comments made on the posts from the Vancouver subreddit positively correlated with the number of new daily COVID-19 cases in British Columbia. From the comments, 13 topics were identified. Two topics were related to vaccines, 1 regarding vaccine uptake and the other about vaccine supply. The levels of discussion for both topics were linked to the total number of vaccines administered. Calgary and Toronto also discussed vaccine uptake, and sentiment scores for this topic differed across the 3 cities (P<.001).

Since July 2020, British Columbia has experienced fluctuations in COVID-19 cases and accompanying restrictions. Cases exponentially increased from 280 cases per day (7-day averages) at the beginning of November 2020 to 833 cases per day at the end of November 2020 [27]. In response, British Columbia issued a 2-week policy on November 7, 2020, prohibiting social gatherings outside of households [28]. This policy was renewed on November 19, 2020, in addition to the implementation of a mask mandate for public spaces [28].

Additionally, in April 2021, cases in British Columbia ranged from 873 to 1130 per day (7-day averages), forming a third peak. The increase in cases caused British Columbia to introduce travel restrictions for nonessential travel between British Columbia health authority boundaries, and to extend the state of emergency twice within the month of April 2021 [29-31].

These peaks in case numbers align with the 2 peaks in engagement on Reddit, and they were found to be significantly positively correlated. This trend suggests that when case numbers increase, more discussion is generated. This finding

builds upon previous literature describing how people bond over the topic of COVID-19 on Twitter [32].

Furthermore, while the number of comments and the number of new cases of COVID-19 increase in tandem from October 2020, the number of comments peaks sooner and decreases at a faster rate than the number of new cases of COVID-19. This suggests that engagement on Reddit is a leading indicator of COVID-19 transmission, as is the case with Google searches, tweets, and Wikipedia page views [33].

However, for the second, higher peak in new cases of COVID-19 in April 2021, the number of Reddit comments increased, albeit at a lower rate, and did not exceed the peak in November 2020. Thus, rather than being a predictor of the number of cases, Reddit engagement patterns may instead be indicative of avoidance behaviors stemming from social media fatigue and the fear of COVID-19 arising from the stresses brought by the November wave [33].

To understand what people discuss on Reddit, we performed topic extraction and identified 13 topics through our model. Compared to studies analyzing themes circulating on Twitter at the beginning of the pandemic, in March 2020, we found that people have continued to compare case numbers, talk about restrictions, and share information on methods of preventing spread [34,35]. In addition, other topics that reflect British Columbia–specific policies and problems were discussed, such as the high transmission of COVID-19 within health care sectors such as long-term care homes during the second wave [36].

Two new, potentially more universal topics surrounding different facets of vaccines have also emerged. The first one addresses vaccine uptake, and the second one is about vaccine availability. Discussion for these topics began trending upward in November 2020, shortly after Pfizer and BioNTech announced the efficacy of their vaccine during phase 3 trials [37]. The level of discussion fluctuated as British Columbia prioritized vaccinating health care workers and residents of long-term care homes [38], but began to rise again in March 2021 as the public was administered vaccines [39].

At first, as the number of people vaccinated continued to grow, vaccine-related discussions increased as well. Comments on vaccine uptake have continued to increase, but have recently decreased for vaccine availability. Topic scores for vaccine availability were also generally higher than that of vaccine uptake until around March 2021. This is likely owing to reports of fluctuations and shortages in vaccine shipments [40,41]. Only recently has British Columbia been able to secure consistent and sufficient vaccine supplies enough to half the wait time between first and second doses [42,43]. Based on the significance of the Granger causality tests, it appears that discussion about either topic may have driven anticipation and demand for the vaccine.

Based on detailed discussions surrounding vaccines, we explored how people felt about them on Reddit by constructing a random forest regression model using tweets labeled with emotional intensity scores. The resulting RMSE values were considered acceptable, and the model was applied to predict emotional intensity scores for Reddit comments. Afterward, we extracted the cleaned comment with the highest score for each emotion. All 4 human authors agreed that the comments demonstrated the emotions they represented. To further validate the scores, a correlation matrix was created. As expected, the negative emotions (anger, fear, and sadness) were strongly positively correlated with each other but strongly negatively correlated with joy.

Observing the trends in emotional intensity of comments over time, we see that comments demonstrate more fear, anger, and sadness than joy [34]. However, comments expressing joy began increasing from April 2021, which coincides with the second peak in case numbers and rise in vaccinations. Since no similar changes were observed in November 2020, we concluded that vaccinations were related to the increase in comments expressing joy. To further confirm this, we compared emotional intensity scores between the 2 types of vaccine-related and non–vaccine-related comments and found that people expressed significantly more positive sentiments about vaccines.

Interestingly, the level of fear expressed about vaccine supply was not significantly different from that in non–vaccine-related comments, and both were significantly higher than vaccine uptake. Nonetheless, both vaccine topics had significantly lower anger and sadness scores than non–vaccine-related topics. This reflects concerns people have about not being able to get the vaccine [44].

Finally, we sought to compare vaccine sentiments across Canadian cities. In addition to discussing vaccine uptake as in Vancouver, Calgary, and Toronto, each had another regionally specific vaccine-related topics. For Toronto, these topics were about vaccine side effects. The word cloud had generally subjective and positive terms including “good” and “feel.” However, it’s emotional scores suggest that the topic may be polarized. Although it has the highest scores for joy, it also has the highest scores for fear. Nonetheless, both vaccine-related topics have significantly lower scores for anger and sadness than non–vaccine-related topics. This suggests that despite feeling generally positive about vaccines, there is still apprehension.

In Calgary, the second topic appears to be about vaccination rates. For this one, the sentiment is clear. The scores for negative emotions significantly exceed those for even non–vaccine-related topics. Since Calgary has the lowest vaccination rates and a lottery exclusively for vaccinated people as an incentive [45], it appears that Redditors are frustrated over the low uptake of vaccines.

When the common topic across all 3 cities—vaccine uptake—was compared, notable differences were observed. Calgary had the highest scores for fear, anger, and joy. Their highest-scoring comment for this topic embodies the first 2 emotions as it seems as though they had inadequate vaccine supply in the midst of high case numbers.

Interestingly, Toronto had the lowest scores for fear, despite also having the lowest scores for joy and the highest scores for sadness and anger. These sentiments could be attributed to frustration over booking vaccines [46] or having strict restrictions still in place despite high vaccination rates and

decreasing case rates [47]. The low scores for fear than those for Calgary and Vancouver, however, are challenging to explain, especially since vaccine-related topics display higher fear than non–vaccine-related topics within Toronto's own subreddit.

Based on our analyses, Reddit comments on posts in city-specific subreddits can be used to assess public sentiment toward COVID-19–related topics. Thus, it is a cost-effective and rapid way for officials to monitor citizens' response to policies. Implementation of sentiment analysis to understand public perceptions of policies in Italy enhanced the accountability and responsiveness of policymakers [48]. Additionally, since engagement on Reddit correlated with COVID-19 cases and vaccination rates, discussion on social media can serve as predictors for real-world statistics.

Finally, our analyses were able to capture variations in sentiment about the same vaccine-related topic across the 3 cities. Accordingly, it is possible for officials to design policies that specifically target populations. For example, reassuring messaging about vaccine side effects and safety may be most useful for Toronto. In contrast, Vancouver could focus more on increasing vaccine supply and Calgary on appealing to vaccine-hesitant groups.

Limitations

Despite the promise in our results, our analyses are not without limitations. First, the data were collected from subreddits, where anyone can comment. However, owing to the community-based nature of the subreddits, we assume that the comments are from commenters located in the cities we are studying. Therefore, the analysis is assumed to be specific to local and provincial policies.

Additionally, emotional intensity scores were assigned to Reddit comments based on tweets because no similarly labeled Reddit data set was available. Common libraries for sentiment analysis, including TextBlob and VADER, lacked specificity for COVID-19–related discussions. Since Twitter is similar to Reddit, in that people post and respond to short, publicly available messages under a username, we assumed that people would use similar language.

Conclusions

Using comments on daily posts containing updates on COVID-19 statistics from a location-specific subreddit, we were able to relate changes in web-based engagement, discussion, and emotional expression to case counts and vaccination rates. Topics relevant to local news and policies were identifiable, as were attitudes toward measures to curb disease spread, such as vaccines. Overall, our study shows that data from social media can be used to better understand concerns and sentiments surrounding the pandemic at the local level, which enables more targeted and publicly acceptable policies.

Acknowledgments

The authors would like to acknowledge the STEM Fellowship for hosting the National Undergraduate Big Data Challenge 2021 and making this study possible.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Line plot comparing Jaccard similarity (red) and coherence (blue) metrics for LDA models created with 1 to 15 topics. The black vertical line denotes the optimal number of topics.
[DOCX File , 70 KB - jmir_v23i9e32685_app1.docx]

Multimedia Appendix 2
Word clouds for topics extracted from r/vancouver.
[DOCX File , 10074 KB - jmir_v23i9e32685_app2.docx]

Multimedia Appendix 3
Sentiment analysis in r/vancouver.
[DOCX File , 318 KB - jmir_v23i9e32685_app3.docx]

Multimedia Appendix 4
Examples of highest-scoring comments for vaccine-related topics for Toronto and Calgary. Approximately the first 25 words of each comment are shown for conciseness.
[DOCX File , 9 KB - jmir_v23i9e32685_app4.docx]

Multimedia Appendix 5
Comparison of sentiment scores for vaccine-related topics in Calgary and Toronto.
[DOCX File , 623 KB - jmir_v23i9e32685_app5.docx]

References

1. Cuello-Garcia C, Pérez-Gaxiola G, van Amelsvoort L. Social media can have an impact on how we manage and investigate the COVID-19 pandemic. J Clin Epidemiol 2020 Nov;127:198-201 [FREE Full text] [doi: 10.1016/j.jclinepi.2020.06.028] [Medline: 32603686]
2. Newman N. Executive Summary and Key Findings of the 2020 Report. Reuters Institute for the Study of Journalism. University of Oxford. URL: https://www.digitalnewsreport.org/survey/2020/overview-key-findings-2020/ [accessed 2021-08-05]
3. Mohammed M, Sha'aban A, Jatau AI, Yunusa I, Isa AM, Wada AS, et al. Assessment of COVID-19 information overload among the general public. J Racial Ethn Health Disparities 2021 Jan 19:1-9 [FREE Full text] [doi: 10.1007/s40615-020-00942-0] [Medline: 33469869]
4. Wunderman Thompson APAC. Wunderman Thompson. URL: https://www.wundermanthompson.com/insight/covid19-infodemic [accessed 2021-06-10]
5. Bridgman A, Merkley E, Loewen PJ, Owen T, Ruths D, Teichmann L, et al. The causes and consequences of COVID-19 misperceptions: Understanding the role of news and social media. HKS Misinfo Review 2020 Jun 18 [FREE Full text] [doi: 10.37016/mr-2020-028]
6. Namkoong K, Fung TKF, Scheufele DA. The politics of emotion: News media attention, emotional responses, and participation during the 2004 U.S. presidential election. Mass Commun Soc 2012 Jan;15(1):25-45. [doi: 10.1080/15205436.2011.563894]
7. Greer C, McLaughlin E. We predict a riot?: Public order policing, new media environments and the rise of the citizen journalist. Br J Criminol 2010 Jul 19;50(6):1041-1059. [doi: 10.1093/bjc/azq039]
8. Cresswell K, Tahir A, Sheikh Z, Hussain Z, Domínguez Hernández A, Harrison E, et al. Understanding public perceptions of COVID-19 contact tracing apps: Artificial intelligence-enabled social media analysis. J Med Internet Res 2021 May 17;23(5):e26618 [FREE Full text] [doi: 10.2196/26618] [Medline: 33939622]
9. Pellert M, Lasser J, Metzler H, Garcia D. Dashboard of sentiment in Austrian social media during COVID-19. Front Big Data 2020;3:32 [FREE Full text] [doi: 10.3389/fdata.2020.00032] [Medline: 33693405]
10. Forni G, Mantovani A, COVID-19 Commission of Accademia Nazionale dei Lincei, Rome. COVID-19 vaccines: Where we stand and challenges ahead. Cell Death Differ 2021 Feb;28(2):626-639 [FREE Full text] [doi: 10.1038/s41418-020-00720-9] [Medline: 33479399]
11. Wilson SL, Wiysonge C. Social media and vaccine hesitancy. BMJ Glob Health 2020 Oct;5(10):e004206 [FREE Full text] [doi: 10.1136/bmjgh-2020-004206] [Medline: 33097547]
12. Burki T. The online anti-vaccine movement in the age of COVID-19. Lancet Digit Health 2020 Oct;2(10):e504-e505 [FREE Full text] [doi: 10.1016/S2589-7500(20)30227-2] [Medline: 32984795]
13. Germani F, Biller-Andorno N. The anti-vaccination infodemic on social media: A behavioral analysis. PLoS One 2021;16(3):e0247642 [FREE Full text] [doi: 10.1371/journal.pone.0247642] [Medline: 33657152]
14. Gupta R, Vishwanath A, Yang Y. Global reactions to COVID-19 on Twitter: A labelled dataset with latent topic, sentiment and emotion attributes. arXiv Preprint posted online July 14, 2020. [FREE Full text]
15. praw-dev / praw. GitHub. URL: https://github.com/praw-dev/praw [accessed 2021-05-20]
16. What's new in 1.2.4 (April 12, 2021). Pandas. URL: https://pandas.pydata.org/pandas-docs/stable/whatsnew/v1.2.4.html [accessed 2021-05-20]
17. Global Reactions to COVID-19 on Twitter: A Labelled Dataset with Latent Topic, Sentiment and Emotion Attributes. openICPSR. URL: https://www.openicpsr.org/openicpsr/project/120321/version/V6/view [accessed 2021-05-21]
18. tweepy / tweepy. GitHub. URL: https://github.com/Tweepy/Tweepy [accessed 2021-05-20]
19. COVID-19 Vaccination Tracker. COVID-19 Tracker Canada. URL: https://covid19tracker.ca/vaccinationtracker.html [accessed 2021-05-21]
20. Python 3.8.2. Python. URL: https://www.python.org/downloads/release/python-382/ [accessed 2021-05-20]
21. Montani I, Honnibal M, Honnibal M, Landeghem SV, Boyd A, Peters H, Roman, murat, GregDubbin, jeannefukumaru, et al. explosion/spaCy: v2.3.6: Bug fixes and base support for Amharic. Zenodo. URL: https://zenodo.org/record/4769120 [accessed 2021-09-15]
22. Wagner W. Steven Bird, Ewan Klein and Edward Loper: Natural language processing with Python, analyzing text with the Natural Language Toolkit. Lang Resour Eval 2010 May 27;44(4):421-424. [doi: 10.1007/s10579-010-9124-x]
23. Řehůřek; R. Fast and faster: A comparison of two streamed matrix decomposition algorithms. arXiv Preprint posted online February 28, 2011. [FREE Full text]
24. Kanani B. Jaccard Similarity – Text Similarity Metric in NLP. Machine Learning Tutorials. 2020 Apr 24. URL: https://studymachinelearning.com/jaccard-similarity-text-similarity-metric-in-nlp/ [accessed 2021-05-25]
25. Evaluate Topic Models: Latent Dirichlet Allocation (LDA). Towards Data Science. URL: https://towardsdatascience.com/evaluate-topic-model-in-python-latent-dirichlet-allocation-lda-7d57484bb5d0 [accessed 2021-05-25]
26. The R Journal. URL: https://cran.r-project.org/doc/Rnews/ [accessed 2021-05-20]
27. B.C. COVID-19 response update. Government of British Columbia. URL: https://news.gov.bc.ca/releases/2020HLTH0102-000540 [accessed 2021-05-23]

28. Kotyk A. Scroll through this timeline of the 1st year of COVID-19 in B.C. CTV News. URL: https://bc.ctvnews.ca/scroll-through-this-timeline-of-the-1st-year-of-covid-19-in-b-c-1.5284929 [accessed 2021-05-23]
29. State of emergency extended to continue B.C.'s COVID-19 response. Government of British Columbia. 2021 Apr 27. URL: https://news.gov.bc.ca/releases/2021PSSG0030-000776 [accessed 2021-05-23]
30. State of emergency extended to continue B.C.'s COVID-19 response. Government of British Columbia. 2021 Apr 13. URL: https://news.gov.bc.ca/releases/2021PSSG0025-000701 [accessed 2021-05-23]
31. Province introduces travel restrictions to curb spread of COVID-19. Government of British Columbia. 2021 Apr 23. URL: https://news.gov.bc.ca/releases/2021PSSG0029-000758 [accessed 2021-05-23]
32. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: Infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
33. O'Leary DE, Storey VC. A Google–Wikipedia–Twitter model as a leading indicator of the numbers of coronavirus deaths. Intell Sys Acc Fin Mgmt 2020 Sep 28;27(3):151-158. [doi: 10.1002/isaf.1482]
34. Boon-Itt S, Skunkan Y. Public perception of the COVID-19 pandemic on Twitter: Sentiment snalysis and topic modeling study. JMIR Public Health Surveill 2020 Nov 11;6(4):e21978 [FREE Full text] [doi: 10.2196/21978] [Medline: 33108310]
35. Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. PLoS One 2020;15(9):e0239441 [FREE Full text] [doi: 10.1371/journal.pone.0239441] [Medline: 32976519]
36. Uguen-Csenge E, Carman T. Here are the B.C. seniors' residences hardest hit by COVID-19 outbreaks. CBC. 2021 Feb 05. URL: https://www.cbc.ca/news/canada/british-columbia/here-are-the-b-c-seniors-residences-hardest-hit-by-covid-19-outbreaks-1.5901929 [accessed 2021-05-23]
37. Pfizer and BioNTech announce vaccine candidate against COVID-19 achieved success in first interim analysis from phase 3 study. Pfizer. 2020 Nov 09. URL: https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-announce-vaccine-candidate-against [accessed 2021-05-25]
38. COVID-19: Guidance for Prioritizing Health Care Workers for COVID-19 Vaccination. Ontario Ministry of Health. 2021 Mar 17. URL: https://www.health.gov.on.ca/en/pro/programs/publichealth/coronavirus/docs/Guidance_for_Prioritizing_HCW_covid19_vaccination_2020-01-08.pdf [accessed 2021-05-25]
39. Ross A. B.C. confirms record high 941 new cases of COVID-19 and 10 more deaths. CBC. 2020 Nov 24. URL: https://www.cbc.ca/news/canada/british-columbia/bc-coronavirus-update-november-24-2020-1.5814873 [accessed 2021-05-25]
40. Bochove D, Bolongaro K. Canada's Vaccine Push Plagued by Confusion, Erratic Supply. Bloomberg. URL: https://www.bloomberg.com/news/articles/2021-05-05/canada-s-vaccine-push-is-plagued-by-confusion-and-erratic-supply [accessed 2021-05-25]
41. Villella S. Concerns about second vaccine dose amid shortage of AstraZeneca supply. CTV News. 2021 May 04. URL: https://kitchener.ctvnews.ca/concerns-about-second-vaccine-dose-amid-shortage-of-astrazeneca-supply-1.5414071 [accessed 2021-05-25]
42. Ghosh T. Canada to get 68M COVID-19 vaccine doses in July thanks to Moderna supply surge. Gobal News. 2021 Jun 18. URL: https://globalnews.ca/news/7962000/canada-covid-vaccine-shipment-7962000/ [accessed 2021-05-25]
43. Azpiri J. 2nd COVID-19 vaccine doses to be sooner than expected, but not 'too soon': B.C.'s top doctor. Global News. 2021 May 25. URL: https://globalnews.ca/news/7892906/bc-covid-19-vaccine-second-dose-timeline/ [accessed 2021-06-25]
44. COVID-19 vaccine supply concerns in British Columbia. Global News. URL: https://globalnews.ca/video/7754241/covid-19-vaccine-supply-concerns-in-british-columbia/ [accessed 2021-05-25]
45. Pearson H. Alberta announces COVID-19 vaccine lottery, 1st prize aimed at 70% 1st-dose goalpost. Global News. 2021 Jun 15. URL: https://globalnews.ca/news/7945464/alberta-covid-19-vaccine-lottery-first-prize/ [accessed 2021-06-25]
46. Thompson N. Ontario's fractured vaccine booking system is complex, but gets job done: experts. CTV News. URL: https://toronto.ctvnews.ca/ontario-s-fractured-vaccine-booking-system-is-complex-but-gets-job-done-experts-1.5480746 [accessed 2021-06-25]
47. Hacisuleyman E, Hale C, Saito Y, Blachere NE, Bergh M, Conlon EG, et al. Vaccine breakthrough infections with SARS-CoV-2 variants. N Engl J Med 2021 Jun 10;384(23):2212-2218 [FREE Full text] [doi: 10.1056/NEJMoa2105000] [Medline: 33882219]
48. Ceron A, Negri F. Public policy and social media: How sentiment analysis can support policy-makers across the policy cycle. Riv Ital Politiche Pubbliche 2015;10(3):309-338. [doi: 10.1483/81600]

Abbreviations

IDF: inverse document frequency
LDA: Latent Dirichlet Allocation
NLTK: Natural Language Toolkit
PRAW: Python Reddit API Wrapper
RCT: randomized controlled trial

RMSE: root mean square error
TF: term frequency
TF–IDF: term frequency–inverse document frequency

Edited by C Basch; submitted 10.08.21; peer-reviewed by I Yunusa, M Mohammed; comments to author 01.09.21; revised version received 08.09.21; accepted 08.09.21; published 24.09.21.

Please cite as:
Yan C, Law M, Nguyen S, Cheung J, Kong J
Comparing Public Sentiment Toward COVID-19 Vaccines Across Canadian Cities: Analysis of Comments on Reddit
J Med Internet Res 2021;23(9):e32685
URL: https://www.jmir.org/2021/9/e32685
doi:10.2196/32685
PMID:34519654

Original Paper

Longitudinal Relationships Among Fear of COVID-19, Smartphone Online Self-Disclosure, Happiness, and Psychological Well-being: Survey Study

Jörg Matthes[1], Prof Dr; Kevin Koban[1], PhD; Ariadne Neureiter[1], MSc, Mag; Anja Stevic[1], PhD

Department of Communication, University of Vienna, Vienna, Austria

Corresponding Author:
Kevin Koban, PhD
Department of Communication
University of Vienna
Währinger Straße 29
R. 7.45
Vienna, 1090
Austria
Phone: 43 1 4277 493 39
Email: kevin.koban@univie.ac.at

Abstract

Background: Given that governmental prevention measures restricted most face-to-face communications, online self-disclosure via smartphones emerged as an alternative coping strategy that aimed at reducing the impact of the COVID-19 pandemic on people's psychological health. Prepandemic research demonstrated that online self-disclosure benefits people's psychological health by establishing meaningful relationships, obtaining social support, and achieving self-acceptance, particularly in times of crisis. However, it is unclear whether these dynamics transition well to lockdown conditions where online self-disclosure must stand almost entirely on its own. Longitudinal investigations are needed to gain insights into the psychological functionalities of online self-disclosure during the COVID-19 pandemic.

Objective: This study aimed to determine the temporal associations between smartphone online self-disclosure (as a communicative behavior) and critical indicators of psychological health (including psychopathological, as well as hedonic and eudaimonic states) during the first COVID-19 lockdown in Austria.

Methods: We conducted a representative 2-wave panel survey between late March/April 2020 and May 2020. A total of 416 participants completed both waves (43.1% attrition rate, given n=731 participants who completed the first wave). A partially metric measurement invariant overtime structural equation model was used to determine the temporal associations among online self-disclosure, fear of COVID-19, happiness, and psychological well-being.

Results: The analysis revealed that fear of COVID-19 significantly predicted online self-disclosure over time (b=0.24, P=.003) and happiness over time (b=−0.14, P=.04), but not psychological well-being (b=0.03, P=.48), that is, stronger COVID-19 fears at T1 prompted more online self-disclosure and less happiness at T2. Online self-disclosure, on the other hand, significantly predicted happiness (b=0.09, P=.02), but neither fear of COVID-19 (b=−0.01, P=.57) nor psychological well-being (b=−0.01, P=.57) over time. Participants who engaged more strongly in online self-disclosure at T1 felt happier at T2, but they did not differ from less-disclosing participants concerning COVID-19 fears and psychological well-being at T2. Importantly, happiness and psychological well-being were significantly related over time (happiness T1 → psychological well-being T2: b=0.11, P<.001; psychological well-being T1 → happiness T2: b=0.42, P<.001).

Conclusions: Our findings suggest that online self-disclosure might play a pivotal role in coping with pandemic stressors. With restrictions on their options, individuals increasingly turn to their smartphones and social media to disclose their feelings, problems, and concerns during lockdown. While online self-disclosure might not alleviate fears or improve psychological well-being, our results demonstrate that it made people experience more happiness during this crisis. This psychological resource may help them withstand the severe psychological consequences of the COVID-19 crisis over longer timeframes.

(J Med Internet Res 2021;23(9):e28700) doi:10.2196/28700

KEYWORDS

COVID-19 pandemic; fear; self-disclosure; happiness, well-being; panel study; smartphones; online platform; social media

Introduction

Background

Across the world, fears (ie, an intense emotion emerging from a perceived imminent threat [1]) related to the COVID-19 pandemic are having a pervasive influence on people's psychological health (understood in a "balanced" way as a continuum that includes both the *absence* of psychopathological and the *presence* of positive psychological states [2]) that extends beyond its immediate threat to physical health [3] and warrants extensive research efforts to face current and future challenges [4]. With respect to these COVID-19 fears, Schimmenti et al [5] used the apocalyptical label of the "four horsemen" to emphasize how significant various states of fear are in explaining people's behavior during the pandemic. According to them, COVID-19 fear involves the following four domains that are mutually connected: fear of/for the body (ie, the body as a potential source of danger and hypervigilance), fear of/for significant others (ie, interpersonal relationships as a potential threat and source for sorrow), fear of (not) knowing (ie, information as a potentially unsettling necessity), and fear of (in)action (ie, decisions that once made or avoided can be potentially harmful). Empirical evidence broadly supports their emphasis showing enhanced levels of fear and other harmful emotional states (such as depressive tendencies and stress) during the COVID-19 pandemic [6-9].

In addition to the fear of getting infected, most governments issued prevention orders that minimized public life and called for physical distance as the appropriate interpersonal norm for private and public gatherings. Consequently, unrestricted face-to-face social environments disappeared almost completely. To deal with this profound change and related adverse psychosocial states, individuals exercised various coping strategies [10,11].

This study focuses on one key coping strategy, *online self-disclosure.* Self-disclosure is defined as the communication of self-related private information to other people [1]. Self-disclosure has long been established as a critical coping behavior in institutionalized therapeutic contexts [12] and informal private settings [13]. In times of the pandemic and particularly during lockdowns, self-disclosure may instead be exercised frequently on social media, most notably via mobile devices such as smartphones. Affordances in such online contexts differ from most offline contexts when it comes to the frequency, breadth (ie, diversity), and depth (ie, intimacy) of self-disclosures [14,15]. Yet, prior research suggests that online self-disclosure can directly contribute to people's psychological health (as an efficient coping behavior) and indirectly by prompting offline self-disclosure as an additional coping resource [16].

Against this background, this study looks at the consequences of online self-disclosures during the first lockdown period of the COVID-19 pandemic in 2020.

This study contributes to our understanding of digital media's roles in times of a pandemic in several ways. First, we, for the first time, shed light on how self-disclosure via smartphones can help individuals cope with pandemic-related stressors, ultimately fostering psychological health. Second, and related to that, we focus on people's online self-disclosure as a communicative *behavior* rather than on *experiential* constructs such as perceived social support and feelings of loneliness. Third, from a methodological standpoint, this study contrasts itself from the great abundance of cross-sectional research by reporting results from a 2-wave longitudinal survey that took place during the first COVID-19 lockdown in Austria. Our sample followed representative quota data, allowing for greater generalizability of the results. Moreover, by employing measurement invariant overtime structural equation models, we can provide the first glimpse into temporal order and directional associations between key constructs. Overall, our study aims at providing the first empirical evidence about how online self-disclosure in times of COVID-19 serves as an effective coping behavior that is provoked by COVID-related fears and leads to improved psychological health.

Fear of COVID-19 as a Predictor of Online Self-Disclosure, Happiness, and Psychological Well-being

In comparison to previous years, overall media use has been on the rise during the COVID-19 pandemic [17,18]. Interestingly, a study by Ohme et al [19] revealed that people tend to prioritize mobile messengers over more public social media platforms when confronted with pandemic-related developments, indicating an immediate need for social support. Large-scale analyses of people's public social media activities exhibited more verbal expressions that would typically be considered as symptomatic of mental health hazards (eg, anxiety, depression, stress, and suicidal ideation), and outreaches for emotional and informational support [20]. Such private and public media activities align with the basic premise of the fever model of disclosure [21]. According to this theoretical approach, self-disclosure is understood as a self-regulative behavior that emerges as a function of experienced distress and aims for maintaining psychological health. Self-disclosing communications are assumed to become more likely under elevated levels of distress to elicit support from close others that may help people cope.

Undoubtedly, the COVID-19 pandemic has been a significant stressor for individuals, especially with respect to fear, as people are "worried, fearful, and uncertain about COVID-19 and the consequences it will have for themselves, their families, communities, and the nation" [6]. Following the fever model of disclosure [21], we theorize that online self-disclosure is one possible coping strategy that individuals use to deal with these fears during lockdown. Following this argumentation, we expected that COVID-19 fears lead to more frequent online self-disclosures. The first hypothesis (H1) was as follows: Fear

of COVID-19 positively predicts online self-disclosure on social media over time.

Except occasionally for playful media-induced feelings [22], fear is generally regarded as a negatively experienced emotion with high valence and arousal [23] that, if not reappraised or extinguished, can damage people's psychological health [24,25]. Protection measures (such as intensified online self-disclosure) may not always suffice to reappraise or eliminate the complex web of fears. In the case of the COVID-19 pandemic (in particular, during the first lockdown), people have been continuously afraid of and hypervigilant about any yet unknown or newly arising threats [5]. It is, therefore, not surprising that findings point toward a negative link between people's COVID-19 fears and their psychological health [26]. Since there were no major alleviating events during the timeframe of this study (ie, March/April to May 2020) that may have mitigated its impact, we hypothesized that people's COVID-19–related fear will have a negative influence on the following two key indicators of psychological health: happiness and psychological well-being. These indicators were selected because each reflects one of the following two dominant traditions in positive psychology: hedonism and eudaimonism [27,28]. More specifically, we follow a previous report [29] in their definition of *happiness* as the presence of pleasant and the absence of unpleasant emotions, and their definition of *psychological well-being* as an umbrella term that includes various eudaimonic experiences, such as meaning and purpose in life, satisfied psychological needs, and self-acceptance. While both indicators contribute individually to psychological health, it is their interaction that can lead people to flourish [30,31]. Due to this mutually reinforcing role in promoting the proverbial good life, we examined both their relationships with fear over time. The second hypothesis (H2) was as follows: Fear of COVID-19 negatively predicts (1) happiness and (2) psychological well-being over time.

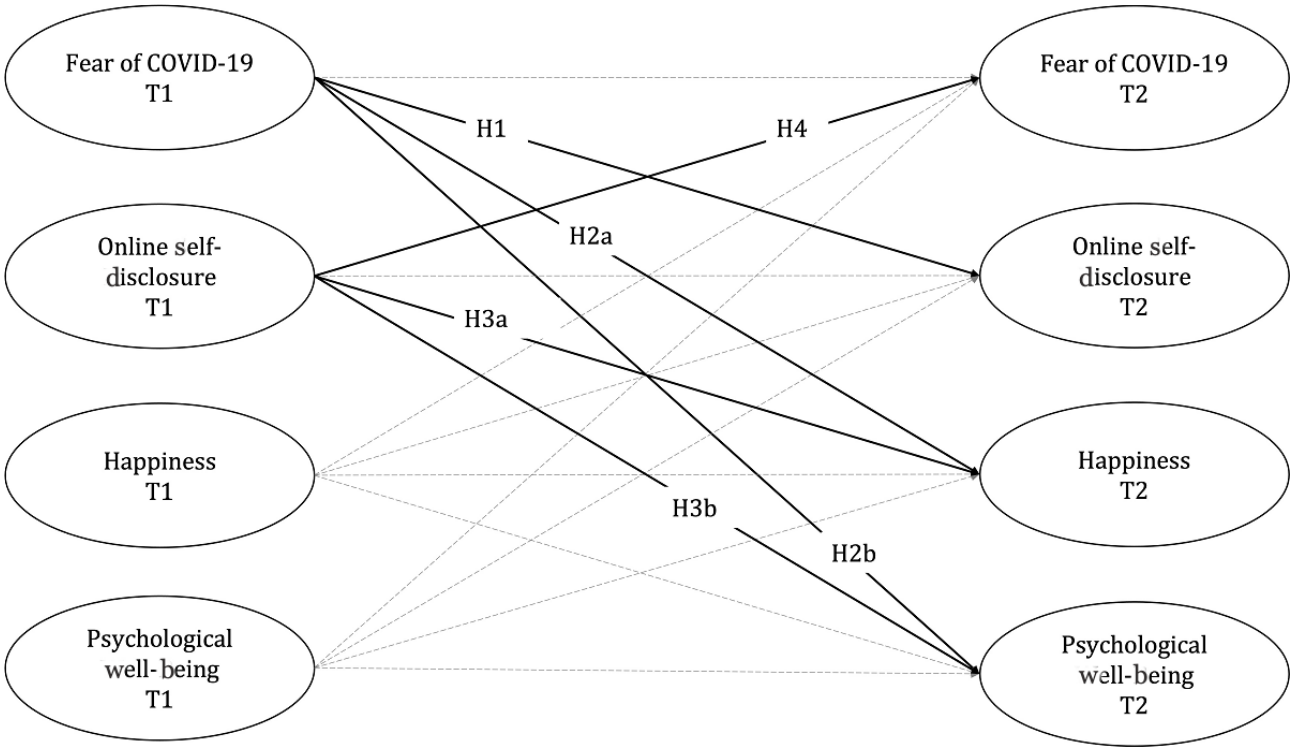
Online Self-Disclosure as a Predictor of Happiness, Psychological Well-being, and Fear of COVID-19

Everything considered, prepandemic research suggests that online self-disclosure can have positive consequences for psychological health. Luo and Hancock [32] detailed different mechanisms through which people may benefit. Most notably, disclosing personal feelings to others in social media constitutes the basis for meaningful online conversations that promote social capital [33] and interpersonal relatedness [34], and help against loneliness [35] and stressful life events [36]. As evidenced in the study by Trepte et al [16], online self-disclosure may also provide a fertile ground for long-term impacts, since it reinforces online self-disclosure in the course of time and carries over to offline (ie, face-to-face) self-disclosures.

This advantageousness might be reinforced during COVID-19 as people may have become more aware and intentional of their online self-disclosures [37]. Within formal therapeutic settings, research has highlighted how online treatments during the pandemic can facilitate self-disclosures and their benefits for psychological health [38,39]. More informally, the unique affordances of social media have been argued to facilitate self-disclosures during COVID-19 that, in turn, may help people to maintain a sense of control [40]. Nevertheless, Canale et al [41] did not find a significant direct correlation between emotional online self-disclosure and psychological health during COVID-19. Indirect effects, however, align with prepandemic links between online self-disclosure and social support. Here, cross-sectional studies have validated the positive effect of social support, very likely following online self-disclosure, on individuals' psychological health [41,42]. In other words, individuals who disclosed their COVID-19–related concerns online were more likely to receive immediate social support, such that they understood or got the impression that they are not alone. Based on these findings, we assumed that online self-disclosure should be positively related to happiness and psychological well-being during the pandemic. The third hypothesis (H3) was as follows: Online self-disclosure on social media positively predicts (1) happiness and (2) psychological well-being over time.

Online self-disclosure during the pandemic may help to not only improve positive psychological states at large but also cope with adverse feelings. The first cross-sectional evidence supports this position. As presented in the study by Canale et al [41], expressing emotions deliberately in online conversations was found to be beneficially associated with coping with COVID-19–related trauma. In this context, emotional online self-disclosures may be understood as a behavioral manifestation linked to adaptive emotion regulation strategies (such as acceptance and perspectivization) that help people cope with their COVID fears [43]. However, cross-sectional investigations cannot determine the directions in which these associations may work. Participants in the study by Canale et al [41] may have overcome COVID-19 trauma due to stronger self-disclosure activities; yet, considering that online conversations are often biased toward displaying positive information [44], it is equally possible that those participants who had scored lower on COVID-19 trauma were merely more motivated to disclose their feelings online. Still, following the report by Luo and Hancock [32], we assumed that people who engage in online self-disclosure during the pandemic might experience less fear about it at a later point in time. The fourth hypothesis (H4) was as follows: Online self-disclosure on social media negatively predicts fear of COVID-19 over time. The full theoretical model is shown in Figure 1.

Figure 1. Full theoretical model of this study. Solid lines indicate the paths that were included in the hypotheses, and dashed lines indicate the paths that were involved in statistical analysis.



Methods

Data Statement

The data of this paper (which is available at Open Science Framework Project Space [45]) is part of a more comprehensive project examining longitudinal relationships between smartphone use in times of COVID-19 and well-being. Importantly, the paper only reports variables that are relevant to the present research interest.

Sampling and Procedure

We conducted a 2-wave panel survey during the first COVID-19 lockdown in Austria. Data collection took place in late March/early April 2020 (T1) and May 2020 (T2) within a 1-month interval. This 1-month (ie, on average) interval was chosen because we were looking at medium-term processes unfolding over several weeks. There was a strong and unpredictable dynamic during the early stages of the COVID-19 pandemic. A larger time span (eg, 6 or 12 months) may have been problematic because several confounding events could have occurred during data collection, making the findings of a larger time-frame meaningless.

Before we started collecting data, we sought ethical clearance from the Institutional Review Board of the Department of Communication at the University of Vienna. Participants were recruited by the professional online polling institute Dynata. Complementarily, another sample was collected simultaneously at the University of Vienna using the same questionnaire, methodology, and quota plan. We used representative quotas for age, gender, and educational level in Austria. For participants to take part in the data collection, they were required to provide their consent, report whether they are using a smartphone, and be at least 16 years of age.

A total of 731 participants (mean age 40.49 years, SD 13.33 years; 394 [53.9%] women; 150 [20.5%] with lower secondary education or less, 340 [46.5%] with vocational school education or secondary education, and 241 [33.0%] with complete university education) completed the survey in the first wave. Among these participants, 416 also completed the second wave (mean age 41.97 years, SD 13.59 years; 226 [54.3%] women; 90 [21.6%] with lower secondary education or less, 187 [45.0%] with vocational school education or secondary education, and 139 [33.4%] with complete university education). The attrition rate between the two waves was 43.1%. There was no significant difference between participants who dropped out after the first wave (T1; n=315) and participants who also completed the second wave (T2; n=416) regarding gender (χ^2_2=1.37, P=.50, Cramér V=0.043), education (χ^2_5=5.33, P=.38, V=0.085), fear of COVID-19 (t_{729}=0.50, P=.62, Cohen d=0.037), online self-disclosure on social media (t_{729}=−1.45, P=.15, d=−0.108), happiness (t_{729}=0.40, P=.69, d=0.030), and psychological well-being ($t_{631.01}$=0.42, P=.68, d=0.032). Participants who dropped out at T2 had a lower age (mean 38.53 years, SD 12.80 years) than participants who participated in both surveys (mean 41.98 years, SD 13.55 years; t_{729}=3.49, P=.001, d=0.261).

Measures

All implemented measures were based on established scales. Given that this study was part of a more comprehensive survey, we opted for short scales that were constructed using high-loading items of these established scales to reduce dropout and prevent respondent fatigue while at the same time minimize reliability and validity losses.

Fear of COVID-19

Similar to Ahorsu et al [8], we employed 2 items measuring fear of COVID-19. We asked participants to indicate their agreement to the following 2 statements: "The coronavirus scares me" and "I am afraid that I could get infected with the coronavirus." Participants answered on a 5-point Likert scale ranging from "strongly disagree" to "strongly agree" (T1: Cronbach α=.76; mean 2.77, SD 1.07; T2: α=.80; mean 2.42, SD 1.07). Principal component analysis resulted in a 1-dimensional scale (eigenvalue=1.61; accounting for 80.03% of the variance).

Online Self-Disclosure on Social Media

We measured online self-disclosure on social media, with 4 items adapted from Schouten et al [35,46]. We asked participants how often they have used their smartphones during the last week to communicate the following content on social media platforms: "personal feelings about the corona crisis," "concerns about the novel coronavirus," "private problems regarding the corona crisis," and "fears of the novel coronavirus." Participants answered on a 7-point Likert scale ranging from "never" to "very often" (T1: α=.91; mean 2.30, SD 1.57; T2: α=.92; mean 1.87, SD 1.32). Principal component analysis indicated a 1-dimensional scale (eigenvalue=3.12; accounting for 78.00% of the variance).

Happiness

We assessed happiness with 3 items derived from Wirth et al [47,48] on a 7-point Likert scale ranging from "never" to "very often." Participants were asked how often they were "in a good mood," "happy," and "joyful" in the last week (T1: α=.92, mean 4.86, SD 1.40; T2: α=.94, mean 5.20, SD 1.36). Principal component analysis showed that happiness is a 1-dimensional scale (eigenvalue=2.59; accounting for 86.44% of the variance).

Psychological Well-being

For psychological well-being, we used 7 items derived from Diener et al [29]. We asked participants to indicate their agreement to the following 7 statements on a 5-point Likert scale ranging from "strongly disagree" to "strongly agree:" "I lead a purposeful and meaningful life," "I am optimistic about my future," "My social relationships are supportive and rewarding," "I am engaged and interested in my daily activities," "I actively contribute to the happiness and well-being of others," "I am competent and capable in the activities that are important to me," and "I am a good person and live a good life" (T1: α=.88, mean 3.97, SD 0.72; T2: α=.90, mean 3.95, SD 0.77). Principal component analysis resulted in a 1-dimensional scale (eigenvalue=4.05; accounting for 57.85% of variance).

Control Variables

As controls, we assessed participants' age, gender (1=male; 2=female), and education (compulsory school levels=low; secondary school levels=moderate; high school and university levels=high). Additionally, we controlled for the sampling method using a dummy variable (0=polling quota sample data; 1=university quota sample data).

Statistical Analysis

Using SPSS Amos [49], we conducted structural equation modeling with full information maximum likelihood estimation. Apart from our control variables, we controlled for autoregressive effects (eg, fear of COVID-19 at T1 as a predictor of fear of COVID-19 at T2). Furthermore, we estimated all reciprocal effects.

Before we conducted the data analysis, we additionally checked for longitudinal measurement invariance of all outcome variables by constraining all factor loadings of our latent variables at T1 and T2 [50]. Based on the comparative fit index (CFI), Tucker-Lewis Index (TLI), Bentler-Bonett Normed Fit Index (NFI), chi-square to degrees of freedom ratio (χ^2/df), and root mean square error of approximation (RMSEA), the goodness of fit of the model can be considered as good (CFI=0.95; TLI=0.93; NFI=0.91; χ^2/df=2.26, P<.001; RMSEA=0.04; 90% CI 0.04-0.05). We established a partial metric invariance model that does not fit the data worse compared to the unconstrained model (χ^2_9=16.54, P=.06). For this partial metric invariance model, 1 factor loading for happiness and 2 for well-being were not forced to be equal over time. Thus, partial metric invariance over time was established, indicating no substantial differences in meaning of all latent variables over time [50].

Results

Zero-order correlations and the main results of the structural model are shown in Table S1 and Table S2 in Multimedia Appendix 1. In line with H1, findings from the autoregressive latent variable model suggests that fear of COVID-19 positively predicts online self-disclosure on social media over time (b=0.24, SE=0.08, P=.003) (Figure 2). Neither happiness nor psychological well-being measured at T1 predicted online self-disclosure on social media at T2. However, we found that higher educated individuals were more likely to self-disclose online over time (b=0.30, SE=0.13, P=.03).

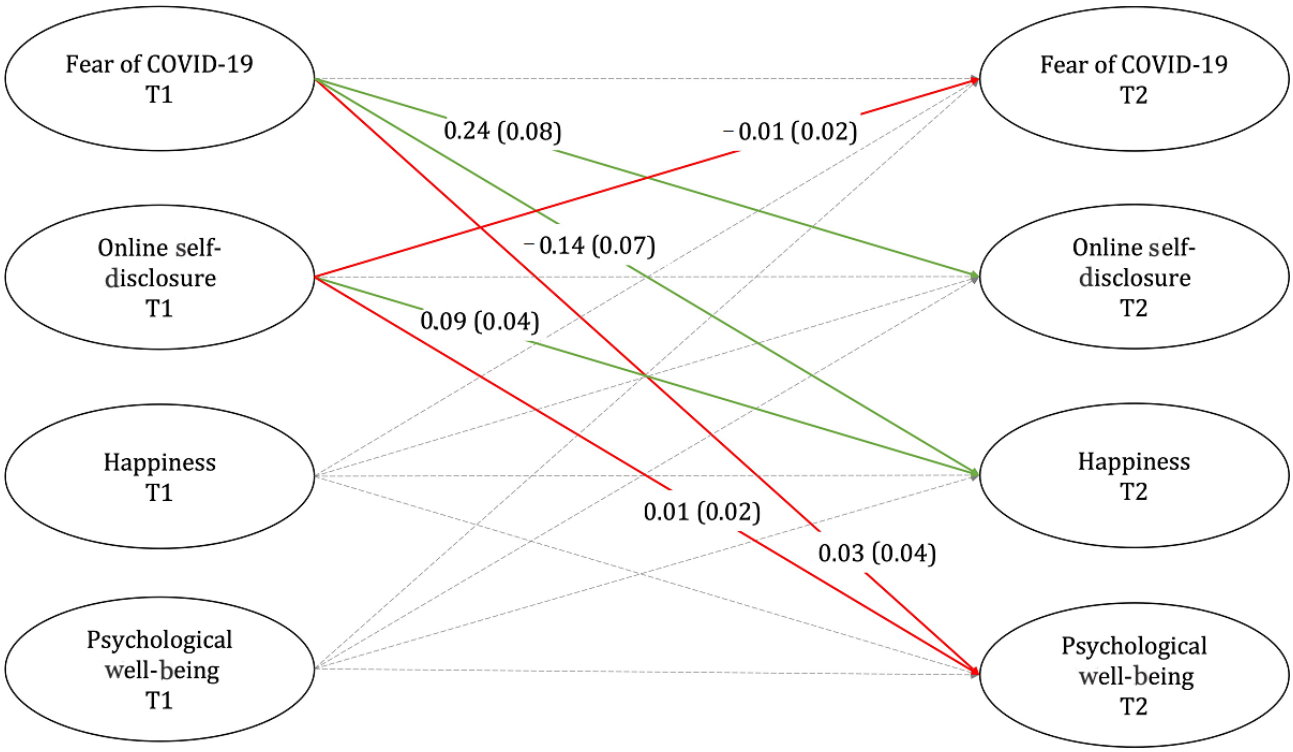
In H2, we assumed that fear of COVID-19 negatively predicts (1) happiness and (2) psychological well-being over time. While we found support for H2a (b=−0.14, SE=0.07, P=.04), there was no support for H2b (b=0.03, SE=0.04, P=.48). Furthermore, there were significant positive relationships between age and psychological well-being (b=0.01, SE=0.00, P=.004), as well as between female gender and psychological well-being (b=0.13, SE=0.06, P=.03).

In line with H3a, we found that online self-disclosure on social media positively predicted happiness (b=0.09, SE=0.04, P=.02). However, against the expectation formulated in H3b, online self-disclosure was found to be unrelated to psychological well-being over time (b=−0.01, SE=0.02, P=.57). Yet, it is important to note that psychological well-being and happiness were significantly related over time (happiness T1 → psychological well-being T2: b=0.11, SE=0.03, P<.001; psychological well-being T1 → happiness T2: b=0.42, SE=0.09, P<.001).

Finally, we found no support for H4, as online self-disclosure was not a significant negative predictor of fear of COVID-19

over time (b=−0.01, SE=0.02, P=.57). Combined with the findings for our first hypothesis, this suggests that fear can prompt self-disclosure, but the reciprocal relationship cannot be observed. Moreover, fear was not significantly predicted by any other variable in our data set, except for psychological well-being, indicating that individuals with high psychological well-being are less likely to experience fear of COVID-19 over time compared to individuals with lower scores (b=−0.15, SE=0.07, P=.04).

Figure 2. Summary of the principal results of the autoregressive structural equation model. Solid lines indicate the paths (presented with unstandardized path coefficients and standard errors in parentheses) that were included in the hypotheses, and dashed lines indicate the paths that were involved in statistical analysis (for detailed results including covariates, see Multimedia Appendix 1). Significant paths are displayed in green, and nonsignificant paths are displayed in red.



Discussion

Principal Findings

This study provides empirical support for the coping strategy of online self-disclosure. More specifically, our main findings reveal that more intense COVID-19 fears at the beginning of the first COVID-19 lockdown (T1) predicted greater online self-disclosure at the end of the lockdown 1 month later (T2), indicating that it has been considered a relevant means of coping in light of governmental restrictions. Further, the study shows that greater engagement in online self-disclosure at T1 coincides with increased feelings of happiness at T2, suggesting coping effectiveness with regard to hedonic emotional states. Notably, psychological well-being at T2 was not linked to COVID-19 fears and online-self disclosure at T1.

In the time of COVID-19, people all over the world have been confronted with significant psychological stressors, such as the likelihood of being infected or the social isolation resulting from lockdown measures. Social media provide a way to cope with these stressors, especially by disclosing self-related private information to other people. Such online self-disclosure has an important psychological function. By disclosing their concerns, fears, or personal problems during the lockdown, individuals connect with others and maintain the feeling of not being alone, which may save them from harmful health-related consequences.

In this study, using partially metric measurement invariant structural equation modeling, we found the first empirical evidence that fear of COVID-19 increased smartphone-based self-disclosure on social media, which, in turn, fostered individuals' happiness. Three aspects of our findings are particularly noteworthy. First, while fear of COVID-19 fostered online self-disclosure over time, we did not observe any evidence for the reverse effect. This means fear leads people to disclose their concerns using digital media, but this does not help to alleviate fear. Interestingly, happiness, which was positively predicted by online self-disclosure, did not reduce fear of COVID-19 over time either. In order to interpret these findings, we need to keep in mind that the study was conducted during the first lockdown in Austria in spring 2020, where the uncertainty regarding the pandemic, the epidemiologic characteristics of the coronavirus, and its consequences for human life were arguably at its peak in many countries. In other words, no matter the amount of online self-disclosure, fears of COVID-19 remained present, most likely due to the ongoing risks and uncertainties associated with it. In fact, the only construct that was found to reduce fear over time was psychological well-being. Individuals who report a high degree of psychological well-being may have more confidence in their ability to master or control the virus or exhibit stronger resilience against adverse psychological states [51]. Clearly, more empirical insight is needed to explain this finding.

Second, it is noteworthy that, compared to happiness, we did not observe a relationship between online self-disclosure and psychological well-being. Happiness can be understood as a relatively transient hedonic emotional state in which individuals experience pleasure and joy [29]. By contrast, psychological well-being includes more stable and long-term eudaimonic indicators such as life satisfaction and quality of life [52]. This refers to the notion that our current life and our life achievements are perceived as being close to our general ideals and goals in life [53]. The inconsistent effectiveness of online self-disclosure for coping may be explained by the contrast in both indicators' scopes. Disclosing to others how one personally feels about COVID-19, even if done repeatedly over the course of 4 weeks, may come as a habitualized temporal relief from pandemic-related stressors; however, it might not by itself be capable of changing the bleak circumstances that created these stressors in the first place (nor stressors unrelated to the pandemic) in such a brief period. Within 4 weeks, online self-disclosure may therefore "only" facilitate ephemeral affective states that got repeatedly threatened by COVID-19 but not more stable long-term self-evaluations that most likely go beyond the circumstances of the pandemic. However, as evident in our data, happiness and psychological well-being are positively and reciprocally related over time. This may suggest that online self-disclosure may impact psychological well-being indirectly, that is, by affecting happiness in the first place, which then influences individuals' psychological well-being. Unfortunately, with only 2 panel waves, we lack the data to test such mediated relationships.

Notwithstanding this limitation, one could argue that happiness serves as an important psychological resource in times of crisis. In states of happiness, individuals are likely to select and process emotionally congruent (ie, positive) information, which may facilitate self-confidence and prevent the emergence of depressive tendencies [54]. One could also assume that happy individuals differ from unhappy individuals in the ways in which they cope with the flow of negative media information surrounding the COVID-19 pandemic [55]. Of course, additional data are needed to firmly test these conjectures.

Third and finally, online self-disclosure was only predicted by fear and not by happiness and psychological well-being, pointing to the very nature of self-disclosure, that is, self-disclosure can be understood as a means to cope with *negative* stressors and not to communicate *positive* experiences. Accordingly, online self-disclosure is not reduced (nor enhanced) in positive life situations. The combination of these findings supports the notion that online self-disclosure comes as a suitable coping strategy in the face of physical distance and government restrictions impeding face-to-face–based coping that appears to withstand the general positivity bias in social media environments.

Limitations

Some limitations need to be acknowledged. Lockdowns and other government measures to fight COVID-19 differ between countries, which may have a significant impact on self-disclosure behaviors. Thus, findings from Austria may not be generalizable to other countries, necessitating additional empirical evidence. Related to this, data collection in this study started at the beginning of the lockdown in March/April 2020, lasting until the end of the lockdown in May 2020 in Austria. The findings may thus not be generalizable to the entire period of the COVID-19 pandemic, such as subsequent COVID-19 lockdowns, that is, individuals may behave in different ways during a second lockdown than they did during the first lockdown. Therefore, familiarity with a lockdown situation and compliance with government measures should be taken into account in future research. Moreover, future studies should use more measurements, ideally spanning over lockdown and nonlockdown periods.

With regard to methodology, we relied on participants' self-reports to measure online self-disclosure. Clearly, self-report data have limitations as they are prone to social desirability and other measurement errors. Participants in this study may have underestimated or overestimated their levels of online self-disclosure. Related to this, our measures and data allow no assumptions about the content of online self-disclosure. Individuals may have enclosed several emotional states, such as fears, practical problems, economic issues, and health-related aspects. In order to shed light on the content of self-disclosure during the pandemic, qualitative methodologies are particularly warranted. Moreover, although we could observe effects over the course of 1 month, our research design did not allow any conclusions about longitudinal mediation mechanisms. Future research should continue this work with more panel waves. Such data would also account for mediation paths among fear of COVID-19, online self-disclosure, and psychological outcomes.

Finally, the relationship between online self-disclosure and psychological well-being may depend on individual predispositions. For instance, chronically distressed individuals may not benefit from online self-disclosure [32]. The reason can be found in their distress-related tendency to share more negative or inauthentic information, which is usually accompanied by more negative received feedback. Future research should investigate these individual-level contingent conditions.

Conclusions

Taken together, our findings suggest that self-disclosure in online digital media can play a major role in restoring people's happiness during COVID-19 and the lockdowns it has brought upon us. With the worldwide spread of COVID-19, fear was a prevalent emotional reaction among many people. To deal with this fear, individuals used digital media, especially social media via their smartphones, to disclose their feelings, problems, and concerns. We showed that such online self-disclosure opportunities contributed to the psychological health of individuals. Although online self-disclosure did not affect psychological well-being, it made people happier over time, which is an important psychological resource in times of uncertainty, crisis, and stress.

Acknowledgments

We would like to thank the group of master's students at the University of Vienna, namely Nadja Daurer, Katharina Feistritzer, Paula Gerhardus, Vera Odorizzi, Aminta Mahmoudi, Raphael Schörkhuber, Stefan Stangassinger, Patrick Stelzl, and Kerstin Widler, for their assistance with the recruitment of participants.

Authors' Contributions

JM and KK drafted the initial manuscript and reviewed and revised the manuscript. JM conceptualized the study and conducted the statistical analysis. AN and AS administered and conceptualized the study, wrote sections of the manuscript, and reviewed and revised the manuscript. All authors approved the final manuscript as submitted and agreed to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Zero-order correlations and results of the structural equation model.
[DOCX File , 25 KB - jmir_v23i9e28700_app1.docx]

References

1. APA Dictionary of Psychology. American Psychological Association. URL: https://dictionary.apa.org/browse [accessed 2021-09-18]
2. Seligman MEP, Parks AC, Steen T. A balanced psychology and a full life. Philos Trans R Soc Lond B Biol Sci 2004 Sep 29;359(1449):1379-1381 [FREE Full text] [doi: 10.1098/rstb.2004.1513] [Medline: 15347529]
3. Mental health and psychosocial considerations during the COVID-19 outbreak. World Health Organization. 2020. URL: https://apps.who.int/iris/bitstream/handle/10665/331490/WHO-2019-nCoV-MentalHealth-2020.1-eng.pdf [accessed 2021-09-18]
4. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. The Lancet Psychiatry 2020 Jun 15;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
5. Schimmenti A, Billieux J, Starcevic V. The four horsemen of fear: An integrated model of understanding fear experiences during the COVID-19 pandemic. Clinical Neuropsychiatry 2020;17(2):41-45. [doi: 10.36131/CN20200202]
6. Fitzpatrick KM, Harris C, Drawve G. Fear of COVID-19 and the mental health consequences in America. Psychol Trauma 2020 Aug;12(S1):S17-S21. [doi: 10.1037/tra0000924] [Medline: 32496100]
7. Salari N, Hosseinian-Far A, Jalali R, Vaisi-Raygani A, Rasoulpoor S, Mohammadi M, et al. Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis. Global Health 2020 Jul 06;16(1):57-11 [FREE Full text] [doi: 10.1186/s12992-020-00589-w] [Medline: 32631403]
8. Ahorsu DK, Lin CY, Imani V, Saffari M, Griffiths MD, Pakpour AH. The Fear of COVID-19 Scale: Development and Initial Validation. Int J Ment Health Addiction 2020 Mar 27:1-9. [doi: 10.1007/s11469-020-00270-8]
9. Coelho CM, Suttiwan P, Arato N, Zsido AN. On the Nature of Fear and Anxiety Triggered by COVID-19. Front Psychol 2020 Nov 9;11:581314-581318 [FREE Full text] [doi: 10.3389/fpsyg.2020.581314] [Medline: 33240172]
10. Figueroa CA, Aguilera A. The Need for a Mental Health Technology Revolution in the COVID-19 Pandemic. Front Psychiatry 2020 Jun 3;11:523-525 [FREE Full text] [doi: 10.3389/fpsyt.2020.00523] [Medline: 32581891]
11. Garfin DR. Technology as a coping tool during the coronavirus disease 2019 (COVID-19) pandemic: Implications and recommendations. Stress Health 2020 Oct 13;36(4):555-559 [FREE Full text] [doi: 10.1002/smi.2975] [Medline: 32762116]
12. Corcoran P. Therapeutic Self-Disclosure: The Talking Cure. In: Corcoran P, Spencer V, editors. Disclosures. London, UK: Routledge; 2000:118-148.
13. Dindia K, Fitzpatrick MA, Kenny DA. Self-Disclosure in Spouse and Stranger Interaction A Social Relations Analysis. Human Comm Res 1997 Mar;23(3):388-412. [doi: 10.1111/j.1468-2958.1997.tb00402.x]
14. Clark-Gordon CV, Bowman ND, Goodboy AK, Wright A. Anonymity and Online Self-Disclosure: A Meta-Analysis. Communication Reports 2019 May 07;32(2):98-111. [doi: 10.1080/08934215.2019.1607516]
15. Nguyen M, Bin YS, Campbell A. Comparing online and offline self-disclosure: a systematic review. Cyberpsychol Behav Soc Netw 2012 Feb;15(2):103-111. [doi: 10.1089/cyber.2011.0277] [Medline: 22032794]
16. Trepte S, Masur PK, Scharkow M. Mutual friends' social support and self-disclosure in face-to-face and instant messenger communication. J Soc Psychol 2018 Dec 26;158(4):430-445. [doi: 10.1080/00224545.2017.1398707] [Medline: 29099670]
17. Mihelj S, Kondor K, Štětka V. Audience Engagement with COVID-19 News: The Impact of Lockdown and Live Coverage, and the Role of Polarization. Journalism Studies 2021 May 25:1-19. [doi: 10.1080/1461670x.2021.1931410]

18. Van Aelst P, Toth F, Castro L, Štětka V, Vreese CD, Aalberg T, et al. Does a Crisis Change News Habits? A Comparative Study of the Effects of COVID-19 on News Media Use in 17 European Countries. Digital Journalism 2021 Jul 21:1-31. [doi: 10.1080/21670811.2021.1943481]
19. Ohme J, Vanden Abeele MMP, Van Gaeveren K, Durnez W, De Marez L. Staying Informed and Bridging "Social Distance": Smartphone News Use and Mobile Messaging Behaviors of Flemish Adults during the First Weeks of the COVID-19 Pandemic. Socius 2020;6:2378023120950190-2378023120950114 [FREE Full text] [doi: 10.1177/2378023120950190] [Medline: 34192139]
20. Saha K, Torous J, Caine ED, De Choudhury M. Psychosocial Effects of the COVID-19 Pandemic: Large-scale Quasi-Experimental Study on Social Media. J Med Internet Res 2020 Nov 24;22(11):e22600 [FREE Full text] [doi: 10.2196/22600] [Medline: 33156805]
21. Stiles W. “I Have to Talk to Somebody”. In: Derlega VJ, Berg JH, editors. Self-Disclosure. Perspectives in Social Psychology. Boston, MA: Springer; 1987:257-282.
22. Martin GN. (Why) Do You Like Scary Movies? A Review of the Empirical Research on Psychological Responses to Horror Films. Front Psychol 2019 Oct 18;10:2298-2222 [FREE Full text] [doi: 10.3389/fpsyg.2019.02298] [Medline: 31681095]
23. Russell JA, Mehrabian A. Evidence for a three-factor theory of emotions. Journal of Research in Personality 1977 Sep;11(3):273-294. [doi: 10.1016/0092-6566(77)90037-x]
24. Reichert A, Tauchmann H. The Causal Impact of Fear of Unemployment on Psychological Health. SSRN Journal 2011 Jul 8:3-23. [doi: 10.2139/ssrn.1880938]
25. Stafford M, Chandola T, Marmot M. Association Between Fear of Crime and Mental Health and Physical Functioning. Am J Public Health 2007 Nov;97(11):2076-2081. [doi: 10.2105/ajph.2006.097154]
26. Sønderskov KM, Dinesen PT, Santini ZI, Østergaard SD. Increased psychological well-being after the apex of the COVID-19 pandemic. Acta Neuropsychiatr 2020 Oct 06;32(5):277-279 [FREE Full text] [doi: 10.1017/neu.2020.26] [Medline: 32624036]
27. Delle Fave A, Brdar I, Freire T, Vella-Brodrick D, Wissing MP. The Eudaimonic and Hedonic Components of Happiness: Qualitative and Quantitative Findings. Soc Indic Res 2010 May 4;100(2):185-207. [doi: 10.1007/s11205-010-9632-5]
28. Huta V, Waterman AS. Eudaimonia and Its Distinction from Hedonia: Developing a Classification and Terminology for Understanding Conceptual and Operational Definitions. J Happiness Stud 2013 Dec 29;15(6):1425-1456. [doi: 10.1007/s10902-013-9485-0]
29. Diener E, Wirtz D, Biswas-Diener R, Tov W, Kim-Prieto C, Choi DW, et al. New Measures of Well-Being. In: Diener E, editor. Assessing Well-Being. Social Indicators Research Series, vol 39. Dordrecht: Springer; 2009:247-266.
30. Keyes CLM, Annas J. Feeling good and functioning well: distinctive concepts in ancient philosophy and contemporary science. The Journal of Positive Psychology 2009 May 14;4(3):197-201. [doi: 10.1080/17439760902844228]
31. Seligman MEP. Flourish: A Visionary New Understanding of Happiness and Well-Being. New York, NY: Free Press; 2011.
32. Luo M, Hancock JT. Self-disclosure and social media: motivations, mechanisms and psychological well-being. Curr Opin Psychol 2020 Feb;31:110-115. [doi: 10.1016/j.copsyc.2019.08.019] [Medline: 31563107]
33. Chen HT, Li X. The contribution of mobile social media to social capital and psychological well-being: Examining the role of communicative use, friending and self-disclosure. Computers in Human Behavior 2017 Oct;75:958-965. [doi: 10.1016/j.chb.2017.06.011]
34. Utz S. The function of self-disclosure on social network sites: Not only intimate, but also positive and entertaining self-disclosures increase the feeling of connection. Computers in Human Behavior 2015 Apr;45:1-10. [doi: 10.1016/j.chb.2014.11.076]
35. Karsay K, Schmuck D, Matthes J, Stevic A. Longitudinal Effects of Excessive Smartphone Use on Stress and Loneliness: The Moderating Role of Self-Disclosure. Cyberpsychol Behav Soc Netw 2019 Nov 01;22(11):706-713. [doi: 10.1089/cyber.2019.0255] [Medline: 31697600]
36. Zhang R. The stress-buffering effect of self-disclosure on Facebook: An examination of stressful life events, social support, and mental health among college students. Computers in Human Behavior 2017 Oct;75:527-537. [doi: 10.1016/j.chb.2017.05.043]
37. Nabity-Grover T, Cheung CMK, Thatcher JB. Inside out and outside in: How the COVID-19 pandemic affects self-disclosure on social media. Int J Inf Manage 2020 Dec;55:102188 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102188] [Medline: 32836645]
38. Feijt M, de Kort Y, Bongers I, Bierbooms J, Westerink J, IJsselsteijn W. Mental Health Care Goes Online: Practitioners' Experiences of Providing Mental Health Care During the COVID-19 Pandemic. Cyberpsychol Behav Soc Netw 2020 Dec;23(12):860-864. [doi: 10.1089/cyber.2020.0370] [Medline: 32815742]
39. Weinberg H. Online group psychotherapy: Challenges and possibilities during COVID-19—A practice review. Group Dynamics: Theory, Research, and Practice 2020 Sep;24(3):201-211. [doi: 10.1037/gdn0000140]
40. Wiederhold BK. Using Social Media to Our Advantage: Alleviating Anxiety During a Pandemic. Cyberpsychol Behav Soc Netw 2020 Apr 01;23(4):197-198. [doi: 10.1089/cyber.2020.29180.bkw] [Medline: 32207630]

41. Canale N, Marino C, Lenzi M, Vieno A, Griffiths MD, Gaboardi M, et al. How Communication Technology Fosters Individual and Social Wellbeing During the Covid-19 Pandemic: Preliminary Support For a Digital Interaction Model. J Happiness Stud 2021 Jun 19:1-19 [FREE Full text] [doi: 10.1007/s10902-021-00421-1] [Medline: 34177363]
42. Qi M, Zhou S, Guo Z, Zhang L, Min H, Li X, et al. The Effect of Social Support on Mental Health in Chinese Adolescents During the Outbreak of COVID-19. J Adolesc Health 2020 Oct 01;67(4):514-518 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.07.001] [Medline: 32753347]
43. Jungmann SM, Witthöft M. Health anxiety, cyberchondria, and coping in the current COVID-19 pandemic: Which factors are related to coronavirus anxiety? J Anxiety Disord 2020 Jun;73:102239 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102239] [Medline: 32502806]
44. Reinecke L, Trepte S. Authenticity and well-being on social network sites: A two-wave longitudinal study on the effects of online authenticity and the positivity bias in SNS communication. Computers in Human Behavior 2014 Jan;30:95-102. [doi: 10.1016/j.chb.2013.07.030]
45. Emotionally Connected: Longitudinal Relationships between Fear of COVID-19, Smartphone Online Self-Disclosure, and Psychological Health. Open Science Framework. URL: https://osf.io/kjrv8/?view_only=3364ca552ecb428c9e19638caec08134 [accessed 2021-09-18]
46. Schouten AP, Valkenburg PM, Peter J. Precursors and Underlying Processes of Adolescents' Online Self-Disclosure: Developing and Testing an "Internet-Attribute-Perception" Model. Media Psychology 2007 Jun 29;10(2):292-315. [doi: 10.1080/15213260701375686]
47. Wirth W, Schemer C, Matthes J. Trivializing the News? Affective Context Effects of Commercials on the Perception of Television News. Mass Communication and Society 2010 Apr 09;13(2):139-156. [doi: 10.1080/15205430903128924]
48. Heiss R, Matthes J. Funny Cats and Politics: Do Humorous Context Posts Impede or Foster the Elaboration of News Posts on Social Media? Communication Research 2019 Feb 09;48(1):100-124. [doi: 10.1177/0093650219826006]
49. SPSS Amos. IBM. URL: https://www.ibm.com/in-en/products/structural-equation-modeling-sem [accessed 2021-09-20]
50. Vandenberg RJ, Lance CE. A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. Organizational Research Methods 2016 Jun 29;3(1):4-70. [doi: 10.1177/109442810031002]
51. Vos L, Habibović M, Nyklíček I, Smeets T, Mertens G. Optimism, mindfulness, and resilience as potential protective factors for the mental health consequences of fear of the coronavirus. Psychiatry Res 2021 Jun;300:113927 [FREE Full text] [doi: 10.1016/j.psychres.2021.113927] [Medline: 33848964]
52. Diener E, Oishi S, Tay L. Advances in subjective well-being research. Nat Hum Behav 2018 Apr 12;2(4):253-260. [doi: 10.1038/s41562-018-0307-6] [Medline: 30936533]
53. Matthes J, Karsay K, Schmuck D, Stevic A. "Too much to handle": Impact of mobile social networking sites on information overload, depressive symptoms, and well-being. Computers in Human Behavior 2020 Apr;105(6):106217. [doi: 10.1016/j.chb.2019.106217]
54. Forgas JP. Mood and judgment: the affect infusion model (AIM). Psychol Bull 1995 Jan;117(1):39-66. [doi: 10.1037/0033-2909.117.1.39] [Medline: 7870863]
55. Fredrickson BL. The broaden-and-build theory of positive emotions. Philos Trans R Soc Lond B Biol Sci 2004 Sep 29;359(1449):1367-1378 [FREE Full text] [doi: 10.1098/rstb.2004.1512] [Medline: 15347528]

Abbreviations

CFI: comparative fit index
NFI: Bentler-Bonett Normed Fit Index
RMSEA: root mean square error of approximation
TLI: Tucker-Lewis Index

Edited by C Basch; submitted 11.03.21; peer-reviewed by C Leung, M Rusll; comments to author 06.07.21; revised version received 23.08.21; accepted 29.08.21; published 27.09.21.

Please cite as:
Matthes J, Koban K, Neureiter A, Stevic A
Longitudinal Relationships Among Fear of COVID-19, Smartphone Online Self-Disclosure, Happiness, and Psychological Well-being: Survey Study
J Med Internet Res 2021;23(9):e28700
URL: https://www.jmir.org/2021/9/e28700
doi:10.2196/28700
PMID:34519657

Original Paper

COVID-19 Mortality Prediction From Deep Learning in a Large Multistate Electronic Health Record and Laboratory Information System Data Set: Algorithm Development and Validation

Saranya Sankaranarayanan[1*], MS; Jagadheshwar Balan[1*], MS; Jesse R Walsh[1], PhD; Yanhong Wu[1], PhD; Sara Minnich[1], BA; Amy Piazza[1], MLS; Collin Osborne[1], MS; Gavin R Oliver[1], MS; Jessica Lesko[1], MA; Kathy L Bates[1], MBA; Kia Khezeli[1], PhD; Darci R Block[1], PhD; Margaret DiGuardo[1], MD; Justin Kreuter[1], MD; John C O'Horo[1], MPH, MD; John Kalantari[1], PhD; Eric W Klee[1], PhD; Mohamed E Salama[1], MD; Benjamin Kipp[1], PhD; William G Morice[1], MD, PhD; Garrett Jenkinson[1], PhD

Mayo Clinic, Rochester, MN, United States

[*]these authors contributed equally

Corresponding Author:
Garrett Jenkinson, PhD
Mayo Clinic
200 1st St SW
Rochester, MN, 55905
United States
Phone: 1 507 293 9457
Email: Jenkinson.William@mayo.edu

Abstract

Background: COVID-19 is caused by the SARS-CoV-2 virus and has strikingly heterogeneous clinical manifestations, with most individuals contracting mild disease but a substantial minority experiencing fulminant cardiopulmonary symptoms or death. The clinical covariates and the laboratory tests performed on a patient provide robust statistics to guide clinical treatment. Deep learning approaches on a data set of this nature enable patient stratification and provide methods to guide clinical treatment.

Objective: Here, we report on the development and prospective validation of a state-of-the-art machine learning model to provide mortality prediction shortly after confirmation of SARS-CoV-2 infection in the Mayo Clinic patient population.

Methods: We retrospectively constructed one of the largest reported and most geographically diverse laboratory information system and electronic health record of COVID-19 data sets in the published literature, which included 11,807 patients residing in 41 states of the United States of America and treated at medical sites across 5 states in 3 time zones. Traditional machine learning models were evaluated independently as well as in a stacked learner approach by using AutoGluon, and various recurrent neural network architectures were considered. The traditional machine learning models were implemented using the AutoGluon-Tabular framework, whereas the recurrent neural networks utilized the TensorFlow Keras framework. We trained these models to operate solely using routine laboratory measurements and clinical covariates available within 72 hours of a patient's first positive COVID-19 nucleic acid test result.

Results: The GRU-D recurrent neural network achieved peak cross-validation performance with 0.938 (SE 0.004) as the area under the receiver operating characteristic (AUROC) curve. This model retained strong performance by reducing the follow-up time to 12 hours (0.916 [SE 0.005] AUROC), and the leave-one-out feature importance analysis indicated that the most independently valuable features were age, Charlson comorbidity index, minimum oxygen saturation, fibrinogen level, and serum iron level. In the prospective testing cohort, this model provided an AUROC of 0.901 and a statistically significant difference in survival (P<.001, hazard ratio for those predicted to survive, 95% CI 0.043-0.106).

Conclusions: Our deep learning approach using GRU-D provides an alert system to flag mortality for COVID-19–positive patients by using clinical covariates and laboratory values within a 72-hour window after the first positive nucleic acid test result.

(J Med Internet Res 2021;23(9):e30157) doi:10.2196/30157

KEYWORDS

COVID-19; mortality; prediction; recurrent neural networks; missing data; time series; deep learning; machine learning; neural network; electronic health record; EHR; algorithm; development; validation

Introduction

COVID-19 is caused by the SARS-CoV-2 virus and is suspected to be of zoonotic origin, with spillover from bats or pangolins into humans in Wuhan, China [1,2]. COVID-19 has become one of the largest public health emergencies of the past century with over 203 million confirmed cases and 4.3 million deaths as of August 2021 according to the World Health Organization [3]. The COVID-19 pandemic has overwhelmed global medical supply chains, hospitals, and economies, which has led governments to respond with varying policies, including mask mandates and travel restrictions [4,5]. At times, hospitals and health care workers have become so overburdened with patients with COVID-19 that they have been forced to ration care, raising logistical and ethical concerns [6].

The clinical course of COVID-19 is diverse with most individuals experiencing mild or asymptomatic disease, but many patients develop life-threatening diseases, including features such as cytokine storms, thrombotic complications, or severe acute respiratory syndrome requiring mechanical ventilation or extracorporeal membrane oxygenation [7]. A major medical challenge is therefore to reliably triage patients according to their risk for severe disease. Age is consistently observed to be a predominant risk factor for severe disease [7], but deaths are not limited to older adults and the majority of older patients survive COVID-19 [7]. Other comorbidities and laboratory test values are expected to be capable of further individualizing and enhancing mortality prediction. Recent studies investigating statistical and machine learning (ML) models for mortality prediction have confirmed that detailed evaluation of medical records can facilitate further stratification of patients [8-12].

A systematic review of 147 published or preprint prediction models found consistent problems with inherent biases in the data sets investigated or created in all such studies, ultimately concluding that “we do not recommend any of these reported prediction models for use in current practice” [12]. Clinical practices differ in the nature of their observational electronic health record (EHR) data set, patient population, clinical practices, and electronic record or laboratory ordering practices. Correspondingly, the literature review conducted at the outset of this study indicated that the existing prediction models were likely unsuited to our clinical setting without essentially starting afresh by retraining, validating, and testing predictions.

We describe Mayo Clinic’s experience assembling, what is to our knowledge, the largest reported COVID-19 database for mortality prediction and using this database to create a system for COVID-19 mortality prediction, tailored to a unique patient population. Despite the biases inherent to it, because this large and growing database represents a health care system spanning 5 states and 3 time zones over a study window greater than 11 months, our model is expected to be the least confounded and the most generalized COVID-19 mortality predictor published to date. We report the successful development and validation of a state-of-the-art ML model to provide mortality prediction shortly after confirmation of SARS-CoV-2 infection in this Mayo Clinic patient population and discuss in detail the various logistical and scientific challenges involved in the early deployment of such a system in a rapidly changing pandemic environment.

Methods

Study Design

This work required the development of a data set and the subsequent modeling of the resultant cohort. After data collection and cleaning, 2 broad classes of algorithms were considered to model this data. The first approach ignores the time series nature of the underlying data and applies traditional ML classifiers. The second approach explicitly models the time series data while dealing with the missing-not-at-random (MNAR) values by using specialized recurrent neural networks (RNNs). Both types of modeling methods were run independently and compared using cross-validation, and a single winning model was selected for prospective performance validation.

EHR and Laboratory Information System Observational Cohort Data Collection

This study adheres to a research protocol approved by the Mayo Clinic Institutional Review Board. Data were retrospectively collected after March 1, 2020 on COVID-19–positive individuals presenting to a Mayo Clinic site or health system, while excluding patients without research consent or from European Union countries covered by the general data protection regulation law. We restricted our focus to 11,807 patients with a positive COVID-19 nucleic acid test result on or before January 27, 2021 and at least one non-COVID test result. Although the data collection system is deployed and ongoing, the January cutoff was selected for this study to provide sufficient cohort size while allowing a minimum of 3 weeks of follow-up to accurately establish survival status. Mayo Clinic’s EHR and laboratory information system (LIS) contain data from each of its 3 campuses (Rochester, Minnesota; Jacksonville, Florida; Scottsdale, Arizona) as well as the surrounding health system sites spanning 5 states (MN, IA, WI, FL, and AZ). Although the EHR contains clinically reportable laboratory results, many of these can only be reported within defined ranges, which can result in qualitative text values rather than the raw numeric measurements. Because many ML algorithms typically work better with quantitative rather than qualitative results, we used the LIS to gather such laboratory testing measurements and the EHR to gather the remaining variables. The EHR data were queried from an underlying Db2 database (IBM Corp), and the LIS data were queried from an SQL database (Microsoft).

Multivariate Time-Series Data With Missingness

The clinical covariates collected were age, sex, height, weight, Charlson comorbidity score, temperature, blood pressure, respiratory rate, oxygen saturation (SpO_2) levels, and diagnoses of chronic kidney disease or diabetes mellitus. Furthermore, we included laboratory test values from a basic metabolic panel, complete blood counts, and some less routine test results of relevance to COVID-19, as determined by scientific literature and physician collaborators. Table 1 details the features collated into our database. In Multimedia Appendix 1, we provide a detailed breakdown of these clinical covariates and laboratory values in our cohort (Table S1 of Multimedia Appendix 1) as well as the cohort's geographic distribution (Figure S1 of Multimedia Appendix 1). Differentiating between missing data and absence of a condition is not possible from EHR diagnostic codes, particularly for patients treated in an outpatient setting. Therefore, we focused mainly on the Charlson comorbidity index [13], which is populated in our EHR when there is a recorded medical history during a "patient encounter" in the EHR. Thus, this variable is available and can be assigned a value corresponding to no comorbidities, which is distinct from missingness in the case of no recorded medical history in the EHR. However, owing to their emphasis within the literature, we also included chronic kidney disease [9] and diabetes mellitus [14] as independent comorbidity variables using their ICD-10 (International Classification of Diseases) codes while acknowledging that these variables conflate missingness with lack of a condition.

Clinical covariates such as pre-existing conditions, height, and weight were sampled infrequently, whereas heart rate and SpO_2 were recorded every 15 minutes for inpatients in our EHR, and other laboratory tests were intermediate in terms of frequency. Therefore, to deal with these multiscale time series measurements, we used the laboratory measurements as the starting point to define our sampling time points. For the variables of sex, age, weight, height, diabetes mellitus, chronic kidney disease, and Charlson comorbidity index, we encoded these variables to exist at the first time point only; in our top performing RNN models, we observed no difference in performance using this strategy when compared to repeating the observations at each time point. For the frequently observed variables of blood pressure systole, blood pressure diastole, temperature, pulse, respiratory rate, and SpO_2, we computed the minimum and maximum measurements for each calendar day and appended these to each laboratory time point during those dates; if no laboratory time point existed on a given day, we created a new one at noon using these minimum and maximum values. We considered time points within ±72 hours of each patient's first positive polymerase chain reaction (PCR) result and performed a sensitivity analysis on the length of the patient follow-up after this positive test result.

Table 1. Feature measurements collected.

Abbreviation	Description (units or levels)
sex	Sex (male or female)
age	Age at time of polymerase chain reaction–positive test result (years)
weight	Weight (kg)
height	Height (cm)
PCR	SARS-CoV-2 nucleic acid test (+ or –)
SERO	SARS-CoV-2 serology antibody test (+ or –)
BASAA	Basophil count test (10^9/L)
EOSAA	Eosinophil count test (10^9/L)
HCT	Hematocrit test (%)
HGB	Hemoglobin test (g/dL)
LYMAA	Lymphocyte count test (10^9/L)
MCV	Mean corpuscular volume test (fL)
MONAA	Monocyte count test (10^9/L)
NEUAA	Neutrophil count test (10^9/L)
PLTC	Platelet count test (10^9/L)
RBC	Red blood cell count test (10^{12}/L)
RDW	Red cell distribution width test (%)
WBC	White blood cell count test (10^9/L)
CRP	C-reactive protein test (mg/L)
D-DIMER	D-dimer test (ng/mL)
FERR	Ferritin test (mg/L)
IL6	Interleukin-6 test (pg/mL)
TRPS	Troponin T test (ng/L)
FIBTP	Fibrinogen test (mg/dL)
LD	Lactate dehydrogenase test (U/L)
IRON	Serum iron test (mg/dL)
TIBC	Total iron binding capacity test (mg/dL)
SAT	Percent iron saturation test (%)
TRSFC	Transferrin test (mg/dL)
BUN	Blood urea nitrogen test (mg/dL)
CHL	Chloride test (mmol/L)
GLU	Glucose test (mg/dL)
CALC	Calcium test (mg/dL)
CREA	Creatinine test (mg/dL)
POTA	Potassium test (mmol/L)
ALB	Albumin test (g/dL)
BICA	Bicarbonate test (mmol/L)
SODI	Sodium test (mmol/L)
BILI	Bilirubin test (mg/dL)
BPsystole	Blood pressure systole (mm Hg)

Abbreviation	Description (units or levels)
BPdiastole	Blood pressure diastole (mm Hg)
Temp	Temperature (°C)
Pulse	Heart rate (1/min)
Resp	Respiratory rate (1/min)
SpO_2	Oxygen saturation (%)
Charlson	Charlson comorbidity index (10-year survival probability)
CKD	Chronic kidney disease (+ or –)
DM	Diabetes mellitus (+ or –)

Time-Flattened ML Models

Time series data were flattened/encoded to a fixed length list of features by carry forward imputation (ie, selection of the most recently observed covariate values), ensuring compatibility with traditional ML models. Specifically, after the data are flattened in this fashion, it forms a tabular prediction task suitable for any canonical supervised classification algorithm. The recently published [15] Python-based automated ML tool AutoGluon-Tabular (v0.2.0) was utilized to enable standardized and reproducible ensemble stacking of many model classes (eg, deep neural networks, LightGBM boosted trees, CatBoost boosted trees, Random Forests, Extremely Randomized Trees, XGBoost, and k-Nearest Neighbors).

AutoGluon-Tabular models were fit to our tabular data frames using the "AutoGluon.TabularPrediction.fit" function using all the default parameters except eval_metric='roc_auc'. After running the fit function, access to each individual model created by AutoGluon was achieved by the "get_model_names" method on the resulting prediction object. This then allowed us to pass the specified model to the "predict_proba" method's optional "model" argument for each of the following model types: KNeighborsUnif, KNeighborsDist, NeuralNetFastAI, LightGBMLarge, NeuralNetMXNet, RandomForestGini, ExtraTreesGini, RandomForestEntr, ExtraTreesEntr, LightGBM, XGBoost, LightGBMXT, CatBoost, WeightedEnsemble_L2. Hereon, we refer to WeightedEnsemble_L2 as the "AutoGluon" model since this was the output of the "predict_proba" method when no single model type was specified.

For relatively static features such as height, weight, or Charlson comorbidity index, we would expect the time-flattened models to be at no disadvantage, whereas the more frequently measured data such as laboratory values or blood pressure will lose information, particularly about trends in the covariates. For instance, 2 individuals with a fever of 39°C recorded in the most recent observation would be treated the same even if one had a sustained high fever and the other had a brief downward trending spike. Of course, there are many potential degrees of freedom to capture more information in the flattened data; one could define a fixed number of the most recent observations or fit a line through the observations over time and pass the slope and intercept as features to the classifier. However, ultimately, the choice to flatten the time series is a choice of convenience and one that attempts to leverage the extensive research efforts devoted to tabular prediction, and therefore, we study here only the last observation carried forward modeling, since proper modeling efforts should account for the time series structure in the EHR data. We next look at models of this form.

RNN Time Series Models

As the second approach, we implemented the modified gated recurrent unit (GRU) binary classification models proposed by Che et al [16] that are capable of accounting for the MNAR patterns within EHR data, and we adopt their notation. For a given patient, we have $D = 54$ variables and a given time series of T time points can be represented as a $T \times D$ matrix X whose rows $x_t \in$  D, $t = 1, \ldots, T$ represent the t-th observation with D variables  , $d = 1, \ldots, D$. Accompanying each observation x_t is a time stamp $s_t \in$  , which starts at time 0, $s_1 = 0$ and a binary masking vector $m_t \in \{0, 1\}^D$ with  taking value 1 when  is observed and 0 otherwise. From these values, we can compute the time intervals.

With these definitions, we can look at various modifications to the standard GRU architecture whose j-th hidden unit has a reset gate  and update gate  with hidden state  at time t and update the equations.

With matrices W_z, W_r, W, U_z, U_r,U and vectors b_z, b_r, b composed of model parameters, $\odot$ is the Hadamard product, and $\sigma(\cdot)$ is the elementwise sigmoid function. Before modifying the architecture, there are 3 methods to use the GRU above to handle missing data: in "GRU-Mean," missing values are imputed by their means in the training data; in "GRU-Forward," missing values are imputed by their last observed value; and in "GRU-Simple," we simply concatenate the x_t, m_t, and δ_t variables into a single observation vector x 't and pass this through the GRU equations above. The GRU-D method uses trainable decay weights.

With W_γ and b_γ being trainable model parameters. The observations are then replaced by the update.

where  is the last observed value of the *d*-th variable and  is the empirical mean of the *d*-th variable in the training data. The modified GRU update equations for GRU-D become the following.

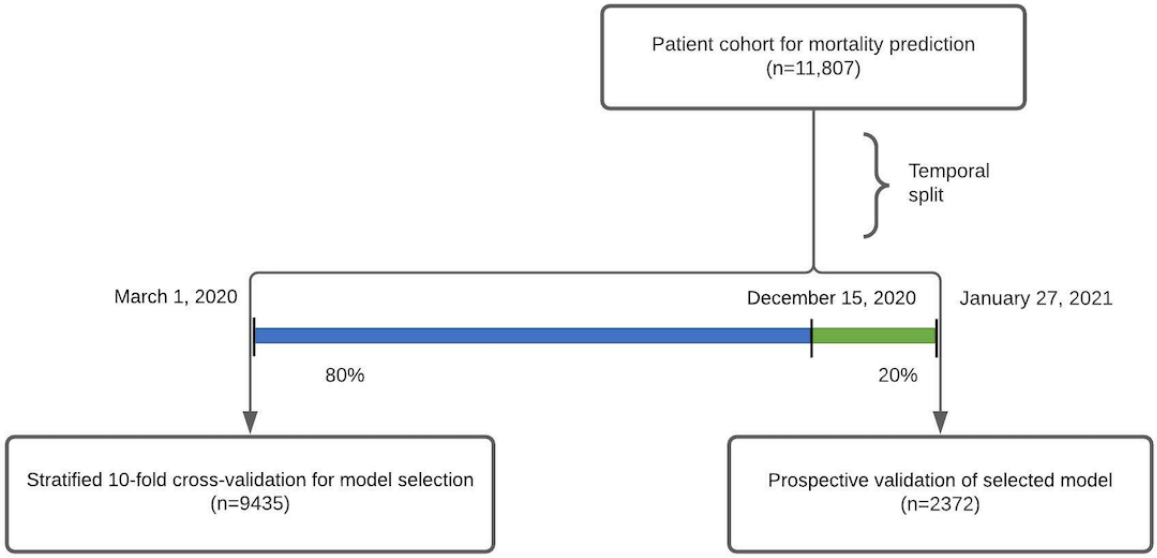
where V_z, V_r, V are new model parameters to directly handle the masking vector m_t in the model.

Our implementation of the above equations in Python is a slightly modified version of the code available on the GRU-D paper's [16] GitHub repository. For the core RNN algorithms, we only edited the original GRU-D code, where required, to be compatible with the more recent versions of tensorflow.keras (version 2.1.0) and numpy (version 1.19.2) used in our high-performance computing cluster environment. We selected the specific RNN algorithm by setting the "--model" argument to be "GRUforward," "GRU0," "GRUsimple," and "GRUD" for GRU-forward, GRU-mean, GRU-simple, and GRU-D, respectively. We utilized the default hyperparameters of the algorithm; however, in our testing, we found that increasing the batch size from 32 to 256 facilitated faster training of the algorithms. Therefore, a batch size of 256 is the only nondefault hyperparameter selection made in our implementation of the RNN algorithms.

Temporal Cohort Split

As depicted in the CONSORT (Consolidated Standards of Reporting Trials) diagram of Figure 1, patients who first tested positive for COVID-19 from March 1, 2020 through December 15, 2020 (9435/11,807, 79.9%) were assigned to a model selection cohort, whereas patients who first tested positive for COVID-19 from December 16, 2020 through January 27, 2021 (2372/11,807, 20.1%) were used as a prospective testing cohort for the final algorithm. All experiments in the model selection cohort were performed using an identical 10-fold stratified cross-validation using binary classification with the positive class defined as death within 21 days of the first positive PCR test result. Only the single best performing model was evaluated on the prospective cohort after being fit against the entire model selection cohort.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram demonstrating the temporal split of our cohort for model selection and prospective validation.



Results

Model Selection

In Figure 2 and Table 2, we compared the results of our various models by using cross-validation area under the receiver operator characteristic (AUROC) curve in the training cohort. Although not in a statistically significant way, we recapitulated the findings of Che et al [16], discovering that the GRU-D model has the highest average cross-validation AUROC curve among all other standard variants of GRU modeling in time series with missing values. In addition, GRU-Simple has higher average cross-validation AUROC curve than the GRU-Forward and GRU-Mean, and the most notable difference underlying these categories is the inclusion of missingness indicators as features to GRU-Simple, which could indicate the value of MNAR patterns in the classification task. GRU-D's biologically inspired architecture attempts to make even more efficient use of this information and exceeds the performance of all the tested RNN methods. AutoGluon, which only had access to the last measurement of each variable, showed strong performance despite this limitation. In Table 2, each individual AutoGluon model was also benchmarked (those with suffix "-AutoGluon"), along with the final ensemble estimate (labeled simply as "AutoGluon"). Although GRU-D ultimately outranked AutoGluon, each method's performance fell within the other's

standard error intervals. AutoGluon's automated hyperparameter tuning and model stacking may indicate that GRU-D could benefit from the addition of hyperparameter search. However, this process may risk overfitting this cross-validation data set, and thus, we selected GRU-D with the default settings rather than attempting to further improve the cross-validation AUROC curve via hyperparameter optimization.

Figure 2. Receiver operating characteristic curves for the 18 models evaluated. AUROC: area under the receiver operating characteristic.

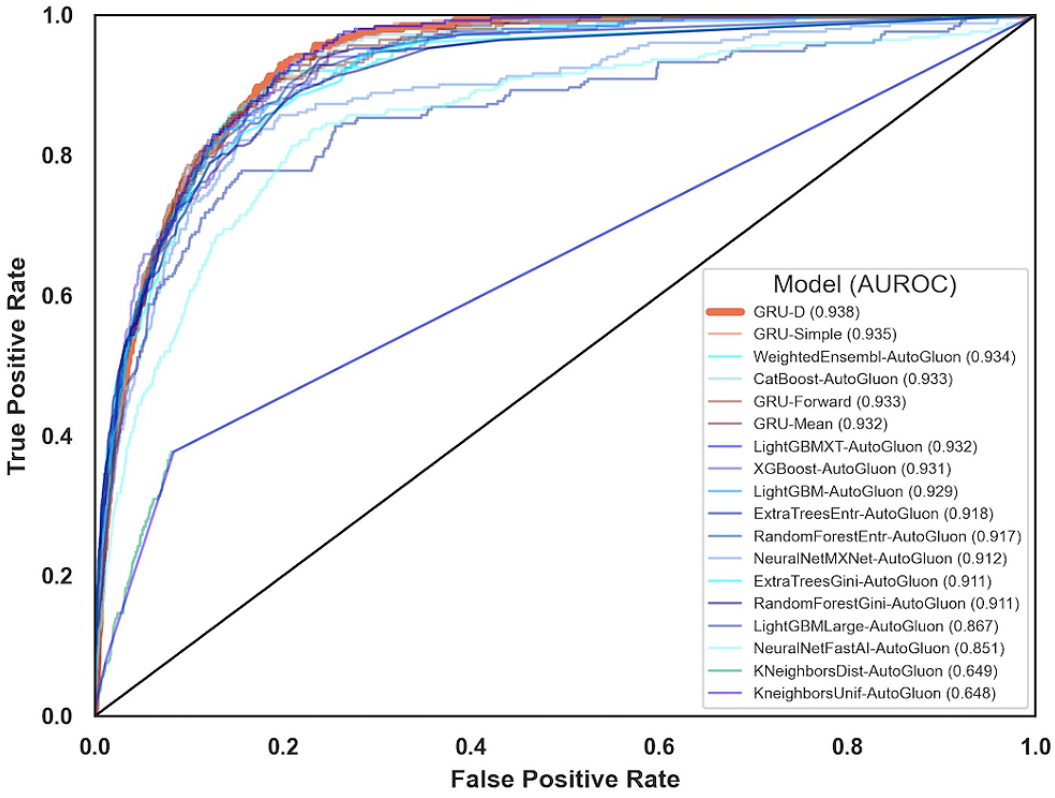


Table 2. Modeling results sorted by performance.

Model	Area under the receiver operator characteristic curve (SE)
KNeighborsUnif-AutoGluon	0.648 (0.011)
KNeighborsDist-AutoGluon	0.649 (0.011)
NeuralNetFastAI-AutoGluon	0.858 (0.013)
LightGBMLarge-AutoGluon	0.867 (0.014)
NeuralNetMXNet-AutoGluon	0.907 (0.008)
RandomForestGini-AutoGluon	0.911 (0.007)
ExtraTreesGini-AutoGluon	0.911 (0.009)
RandomForestEntr-AutoGluon	0.917 (0.008)
ExtraTreesEntr-AutoGluon	0.918 (0.007)
LightGBM-AutoGluon	0.929 (0.007)
XGBoost-AutoGluon	0.931 (0.006)
LightGBMXT-AutoGluon	0.931 (0.005)
GRU-Mean	0.932 (0.005)
GRU-Forward	0.933 (0.006)
CatBoost-AutoGluon	0.933 (0.005)
AutoGluon	0.934 (0.005)
GRU-Simple	0.935 (0.004)
GRU-D	0.938 (0.004)

Length of Time Series

Clearly, we would expect availability of more time series data to result in improved model performance. To determine if predictions could be made utilizing data prior to 72 hours of a patient's first positive PCR test result, we assessed the performance of GRU-D when we restricted the time series to 12, 24, 48, and 72 hours of follow-up after the first positive PCR test result. The results in Table 3 demonstrate that although we lose performance when predicting earlier in the patient's disease, we are still able to provide accurate predictions even using data within the same day (12 hours of follow-up) that a patient tests positive for COVID.

Table 3. GRU-D performance versus length of time series.

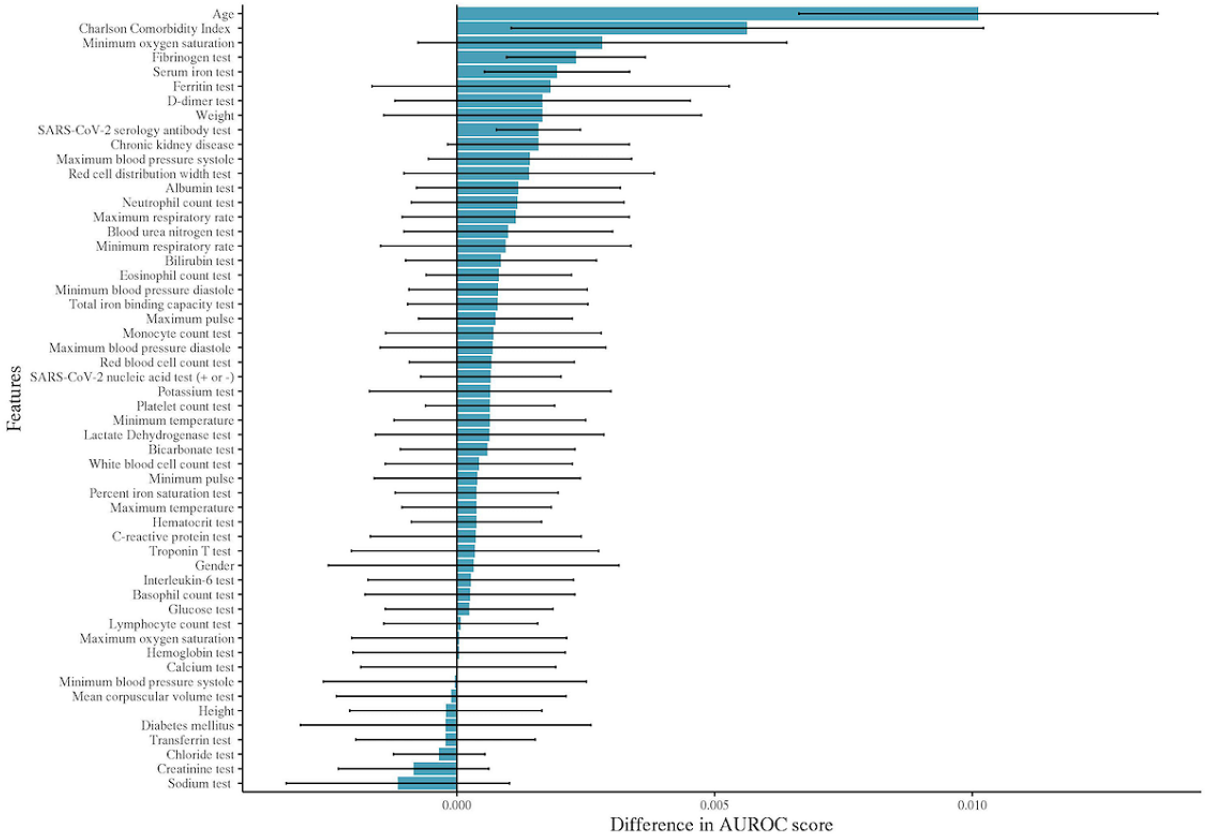
Follow-up after positive finding for polymerase chain reaction	Area under the receiver operator characteristic curve (SE)
12 h	0.916 (0.005)
24 h	0.919 (0.006)
48 h	0.925 (0.005)
72 h	0.938 (0.004)

MNAR as an Asset and Feature Importance

To demonstrate the fact that MNAR data can improve model predictions by GRU-D, we generated a synthetic data set with laboratory test values replaced by Bernoulli coin flips. Therefore, the only valuable information contained within this data set's laboratory values is the missing data patterns that can be viewed as encoding clinical suspicion or concern. For instance, the D-dimer laboratory value is ordered less frequently than other tests, and therefore, its presence alone can be informative of clinical concern for thrombotic events. Our results found that randomizing the laboratory values resulted in an AUROC curve of 0.904 (0.006), which indicates that the laboratory values in aggregate contributed 0.034 to the AUROC score (since this is the drop in performance compared to the model with actual laboratory values). We ran a further experiment omitting the laboratory values entirely, which produced a lower AUROC of 0.890 (0.006). Therefore, the missing patterns alone contributed 0.014 to the AUROC. To contextualize this finding, we dropped each feature individually from the model, assessed the decrease in the AUROC score, and summarized the top 10 features in the decreasing order of the difference in the AUROC score (Figure 3). We note here that the drop due to missing patterns exceeds the drop due to removing any single variable from the analysis, making the MNAR pattern one of the most valuable pieces of information available to GRU-D. In Multimedia Appendix 1, we show a detailed error analysis of our model using these top 10 features. The fact that age and Charlson comorbidity index are the most significant contributors to mortality prediction is consistent with

the well-known risk factors for COVID mortality. The findings of the fibrinogen test, serum iron test, and ferritin test were the 3 most important laboratory values in our models. The presence of chronic kidney disease, weight, serology test, and SpO_2 were the clinical covariates that also ranked in the top 10 variables by importance. Interestingly, height had low importance, indicating that BMI may not be as effective as weight itself in mortality prediction. However, a limitation of this drop-one-feature variable importance is that a low-ranking feature such as height cannot be said to be irrelevant, just that any information it carries is redundant within other features.

Figure 3. Feature importance in the GRU-D recurrent neural network model as defined by the average drop in the area under the receiver operator characteristic curve (with 95% CI) when each feature is individually removed from the analysis. The top 5 features are seen to be age, Charlson comorbidity index, minimum oxygen saturation, fibrinogen levels, and serum iron levels. AUROC: area under the receiver operator characteristic.

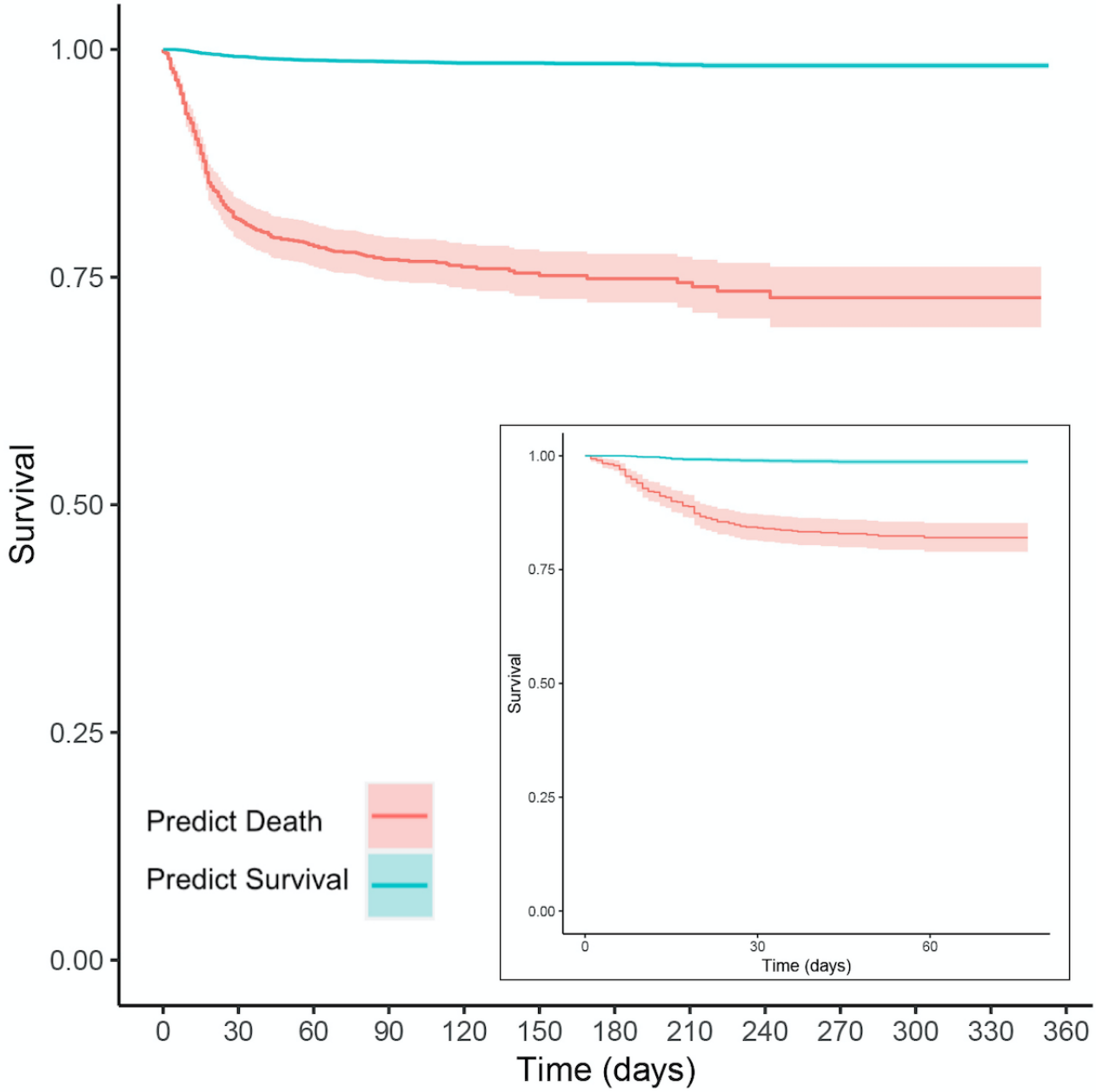


Prospective Validation and Survival Analysis

To demonstrate the efficacy of our proposed mortality prediction, we performed a Kaplan-Meier analysis using the survival R library [17]. Specifically, we chose a decision boundary on the GRU-D model's ROC curve, which provided a specific delineation of high-risk and low-risk groups of patients. In our cross-validation cohort, binary classification provides accuracy of 89% (95% CI 88%-90%), recall of 80% (95% CI 74%-85%), precision of 17% (95% CI 15%-19%), and a negative predictive value of 99% (95% CI 99%-100%). Furthermore, although the precision is somewhat low with numerous false positives, we see among the survivors over twice the rate of mechanical ventilation or extracorporeal membrane oxygenation when they are predicted to die by GRU-D (Fisher exact test $P<.001$, odds ratio 2.1, 95% CI 1.8-2.5). We validated this performance in our prospective testing cohort, finding an AUROC of 0.901, accuracy of 78% (95% CI 76%-79%), recall of 85% (95% CI 77%-91%), precision of 14% (95% CI 12%-17%), and a negative predictive value of 99% (95% CI 99%-100%).

Our Kaplan-Meier analysis results in Figure 4 demonstrate the statistically significant stratification provided by our ML model in both the cross-validation and prospective testing experiments. Building a Cox Proportional Hazards model for our prediction in the cross-validation cohort provides a statistically significant difference in survival between the 2 groups ($P<.001$ for the likelihood ratio, logrank, and Wald tests), with a prediction of survival having a substantially improved hazard ratio of 0.053 (95% CI 0.043-0.066). We validated this finding in the prospective testing cohort with a statistically significant difference in survival between the 2 groups ($P<.001$ for the likelihood ratio, logrank, and Wald tests), with a prediction of survival having a substantially improved hazard ratio of 0.067 (95% CI 0.043-0.106).

Figure 4. Kaplan-Meier survival curves for the GRU-D stratified populations in the cross-validation cohort (main figure) and the prospective test cohort (inset), where teal is the prediction of low risk of death and red is the prediction of high risk. Both figures have 95% CIs visualized for the teal and red curves, although the teal confidence bands are tight due to our large sample sizes.



Discussion

Study Overview

In this study, we collected and processed over 50 laboratory and clinical covariates in a population of nearly 12,000 Mayo Clinic patients who tested positive for SARS-CoV-2 by PCR. In this large and geographically diverse data set, we found that the GRU-D RNN could provide state-of-the-art mortality prediction. This performance remained strong even in a held-out test set that mimics how a deployed system would be trained retrospectively and then prospectively utilized in a clinically evolving pandemic setting.

Principal Results

Our cross-validation experiments summarized in Table 2 indicated that the top performing model to predict mortality in our cohort was the GRU-D RNN. We thus selected the GRU-D method to predict the mortality of patients with COVID-19 and prospectively found an AUROC of 0.901, accuracy of 78% (95% CI 76%-79%), recall of 85% (95% CI 77%-91%), precision of 14% (95% CI 12%-17%), negative predictive value of 99% (95% CI 99%-100%), and a statistically significant difference in survival (P<.001, hazard ratio for those predicted to survive, 95% CI 0.043-0.106). As can be expected in prospective validation, we observed a modest drop in AUROC although most of the performance characteristics were close to their original cross-validation estimates, that is, the negative predictive value was largely unchanged, while precision and

accuracy showed minor decreases with the recall showing modest improvements.

We chose a prospective/retrospective split in time since this is the most realistic way to assess the potential performance of a system if launched clinically, because it would be trained on data up until its go-live date and then run prospectively in a potentially evolving pandemic environment. Notably, the cutoff date for the 80/20 split creating the prospective test set was December 15, 2020, which is the day after the first COVID vaccine received the United States Food and Drug Administration approval, meaning that our prospective cohort represented a distinctly different clinical environment compared to the period in which the model was trained. The relatively minor loss of performance in prospective validation shows the robustness of the modeling herein, but the observed loss of performance also demonstrates the need for continued retraining/validation of such a model during a constantly evolving pandemic.

Limitations

The application and deployment of ML methods in clinical practice require concerted care and diligence. One may be inclined to interpret the high negative predictive value of our prediction algorithm as an indication that the best use of the algorithm in practice is as a screening mechanism to discharge patients who are not at risk in order to conserve resources for higher-risk individuals. However, such a conclusion illustrates a pitfall of using a correlative prediction algorithm to make causal conclusions. The algorithm is highly confident that under the current standards of care at Mayo Clinic, these individuals are not likely to succumb to their illness; this is quite distinct from asserting that it is safe to reduce the care for these patients. Arriving at this latter conclusion would likely require a randomized controlled trial, and given the much lower survival rate published in the New York City data set [11] where medical systems were overcapacity, it seems unlikely that reducing care from those who survived in our cohort would have been a safe measure. Because the Mayo Clinic health systems have not been overcapacity, our mortality predictions should be viewed as representing patient stratification when full clinical support is available.

Therefore, we conclude that the algorithm is better deployed as an alert system that flags only those patients it deems as high risk to provide the treating physician with an additional data point that aims to summarize the many covariates and the laboratory values routinely available. In this context, the algorithm has had abundant experience in the provider's system, effectively "seeing" all patients with COVID-19 that have attended Mayo Clinic and conveying these lessons to physicians who could not have gained such experience personally.

A web interface to this model may allow for widespread usage but given the complexity and error-prone nature of users providing the high dimensional time-series measurements with correct units, the system is better suited for integration within the EHR/LIS infrastructure. We are now exploring the details of deployment of such a GRU-D alert system, which involves discussions with physicians to assess numerous implementation details, for example, deciding whether the alerts would be passive EHR/chart-based flags or a direct page to the frontline clinical provider. Passive chart alerts are less intrusive to existing workflows (ie, a direct page interrupts a physician while tending to other patients) but also provide less-immediate feedback. Additionally, active alerts could also be sent to a triage group to consider if evaluation is needed (for example, from the registered respiratory therapist) rather than interrupting bedside clinicians. Furthermore, for either type of alert, there is the question of prescribing a bedside assessment or leaving it to provider discretion, which is again a matter of balancing disruption of the workflow with the likelihood of missing a critical event. There will not be a universally appropriate implementation for all hospital systems owing to staffing and procedural differences. However, since our algorithm predicts overall COVID-19 mortality and is not tailored to flag imminent events such as cardiopulmonary arrest, it may be appropriate to consider less intrusive chart alerts without prescribed bedside follow-ups.

We have also seen nuances in the challenges and opportunities presented by MNAR data. In the context of traditional statistical inference and imputation, MNAR data is a worst-case scenario so challenging that many practical applications effectively ignore the reality and proceed with algorithms designed for the missing completely at random or missing at random settings. A diligent statistician making this decision may perform a sensitivity analysis under a very limited set of assumed MNAR mechanisms to provide some assurances regarding the robustness of the chosen imputation or analytical strategy [18]. However, here we have demonstrated that classification problems can be quite distinct in this regard. Specifically, if the missing data mechanism is tightly coupled to the ultimate prediction task, it is entirely possible for MNAR data to be an asset rather than an impediment. One can construct a context where the class label is so tightly linked to the missing data mechanism that the patterns of missingness provide more discriminative power than the underlying values themselves (see Multimedia Appendix 2) [19]. In LIS systems, the number of potential laboratory tests that could be ordered at any time is astronomical, and it is unlikely that a practicing physician will ever order a "complete observation" of every test available on a single patient at every point in time. Instead, tests are ordered based on reasonable clinical suspicion that a test might return an abnormal result. From a prognostication point of view, this clinical suspicion is an enormously valuable piece of information that will almost never be captured in a structured data field in the EHR. If an algorithm cannot build off of this clinical suspicion as a starting point, it is also likely that its conclusions may appear to be a "step behind" the ordering clinician. Instead, an algorithm should learn what it can from the MNAR data patterns (here partly encoding clinical suspicion) in addition to the final value returned by the laboratory test.

We also note some of the real-world challenges that are faced when attempting to deploy such an alert system into clinical practice. First, in the retrospective experimental design followed here and by other papers in the literature, the time series data are constructed using the time of sample collection since this is the most biologically accurate way to represent the data and build predictive models. However, in practice, if there can be

delays in the turnaround for certain tests, this will either result in delayed predictions (so that the deployed testing data match its retrospective training counterpart) or result in biased predictions when delayed laboratory test results are treated as missing. Therefore, although 72 hours is early in the course of illness, it is crucial that we have demonstrated reasonable performance even when only considering data collected on the same day as the first positive PCR result, because a real-world delay of 48 hours on certain laboratory test values may occur during a global pandemic, and thus, it is critical that the system can still provide accurate and timely predictions even when laboratory test results are delayed. Additionally, with vaccines now being delivered, the models presented herein should be considered as mortality predictions for an unvaccinated individual, and in practice, a vaccinated individual will be expected to be at low risk for mortality based on the clinical trials data.

Another challenge in dealing with LIS data comes from nonstandardization of test coding prior to reporting to the EHR. In a multisite system, the same laboratory test may have multiple test codes to account for the different ordering facilities or variability in local billing regulations. This creates the potential for discrepancies in the values stored within the underlying database such as differing units of measure. Substantial effort is therefore devoted to linking the LIS results to the EHR to ensure consistency across test codes and complete coverage of results in the EHR. The COVID-19 pandemic has created added complexity due to the rapidly evolving and continuously updating availability of COVID-19 nucleic acid and antibody tests. Therefore, effective data collection and deployment of ML methodologies necessitates extensive team-based laboratory and medical expertise to ensure that data aggregation and modeling efforts can be rapidly modified to suit the changing nature of the underlying data set. Scalability also presents practical challenges. This is illustrated by a scenario in which internal workflows began to fail due to limitations in the number of query results being returned by Tableau, necessitating that SQL queries take place on a high-performance computing cluster using a Python/Pandas toolchain. Although these logistical challenges may be of limited academic interest, they are important to document, as such barriers have been a greater impediment to rapid real world deployment than more traditional topics in the ML literature such as the identification of appropriate classification algorithms.

Comparison With Prior Work

For context, in Table 4, we summarize some of the largest published COVID-19 mortality studies and specifically, the cohorts analyzed and the most relevant features identified. When smaller cohorts see insufficient numbers of deaths for direct mortality prediction, studies tend to focus on the prediction of severe outcomes. For instance, in a cohort of 123 patients with COVID-19 in Vulcan Hill Hospital, China, in the study of Pan et al [20], the mortality classifier based on XGBoost yielded an AUC of 0.86-0.92. Likewise, in a cohort of 372 Chinese cases (99.7% cohort survival rate), Gong et al [9] found that the following variables provided an AUROC of 0.85. Similarly, in a study of 375 patients with COVID-19 conducted by Ko et al [21], the mortality prediction model based on XGBoost had 92% accuracy. In a study of 398 COVID-19–positive patients by Abdulaal et al [22], 86% accuracy was achieved (95% CI 75%-93%). In a large study of 2160 cases over 54 days from 3 hospitals in Wuhan, China with sufficient cases to assess mortality (88% cohort survival rate), Gao et al [8] reported 0.92-0.98 as the AUROC using an ensemble classifier. Furthermore, Vaid et al [11] used 4098 inpatient cases over 68 days in New York City (83% cohort survival rate) to achieve an AUROC of 0.84-0.88 in mortality prediction. Kim et al [23] studied 4787 patients and their XGBoost-based classifier demonstrated an AUC of 0.88-0.89 (95% CI 0.85-0.91) in predicting the need for intensive care, which is distinct from mortality prediction. Bolourani et al [24] studied 11,525 patients to achieve an AUROC of 0.77 in predicting respiratory failure within 48 hours of admission, which is also distinct from mortality prediction, based on data from the emergency department by using an XGBoost model.

The dramatically different cohort mortality rates and the associated predictive accuracies may be in part due to the differing straining of the local health care systems at the time of study (both Wuhan and New York City experienced waves of patients that at different times overwhelmed the health care infrastructure), and the relatively geographically narrow nature of each of these data sets underscores why it is unlikely that these mortality predictions would extend directly to our patient population in a health care system spanning 3 time zones and multiple locales unrepresented in the literature.

Table 4. Summary of the related studies.

Study	Patients (n)	Model/algorithm	Cohort survival	Prediction	Area under the receiver operating characteristic curve	Feature importance
Pan et al [20]	123	XGBoost	52.8%	Mortality	0.86-0.92	Lymphocyte percentage, prothrombin time, lactate dehydrogenase, total bilirubin, eosinophil percentage, creatinine, neutrophil percentage, and albumin level
Gong et al [9]	372	Nomogram	99.7%	Severity	0.85 (95% CI 0.790-0.916)	Higher lactate dehydrogenase, C-reactive protein, red blood cell distribution width, direct bilirubin, blood urea nitrogen, and lower albumin
Ko et al [21]	375	XGBoost	98.1%	Mortality	—,[a] accuracy of 92%	Not assessed
Abdulaal et al [22]	398	Artificial neural network	—	Mortality	—, accuracy of 86% (95% CI 75%-93%)	Altered mentation, dyspnea, age, collapse, gender, and cough
Shi et al [10]	487	Custom risk score calculation	100%	Severity	—	Advanced age, presence of hypertension, and being male
Gao et al [8]	2160	Ensemble model based on logistic regression, gradient-boosted decision tree, neural network, and support vector machine	88%	Mortality	0.92-0.98	Consciousness, chronic kidney disease, lymphocyte counts, sex, sputum, blood urea nitrogen, respiratory rate, oxygen saturation, D-dimer, number of comorbidities, albumin, age, fever, and platelet count
Vaid et al [11]	4098	XGBoost	83%	Mortality	0.84-0.88	Age, anion gap, C-reactive protein, lactate dehydrogenase, oxygen saturation, blood urea nitrogen, ferritin, red cell distribution width, and diastolic blood pressure.
Kim et al [23]	4787	XGBoost	—	Need for intensive care	0.88-0.89	Activities of daily living, age, dyspnea, body temperature, sex, and underlying comorbidities
Bolourani et al [24]	11,525	XGBoost	—	Predicting respiratory failure	0.77	Invasive mode of oxygen delivery being a nonrebreather mask, emergency severity index values of 1 and 3, maximum respiratory rate, maximum, oxygen saturation, Black race, age on admission, eosinophil percentage, serum sodium level, and serum lactate level.
This study	11,807	GRU-D	95.4%	Mortality	0.938 cross-validation; 0.901 prospectively	Figure 2, top 5: age, Charlson comorbidity index, minimum oxygen saturation, fibrinogen level, and serum iron level

[a]Not available.

As indicated in Table 4, this study represents the largest cohort collected for mortality prediction in COVID-19, and the GRU-D algorithm shows state-of-the-art performance. Notably, many papers selected models based on XGBoost, which also showed strong cross-validation performance in our data. However, Table 2 demonstrates that XGBoost was not even in the top 5 algorithms that we assessed. Additionally, in agreement with Gao et al [8], we find that ensemble algorithms such as AutoGluon can provide stronger performance, although as noted previously, the GRU-D algorithm ended up ranked most highly in our cross-validation experiments.

Conclusions

We have aggregated and analyzed one of the largest multistate COVID-19 EHR databases for mortality prediction. Using this database, we have trained and prospectively validated a highly effective ML algorithm using the GRU-D neural network architecture to predict the mortality of patients with COVID-19 shortly after their first positive PCR test result.

Acknowledgments

We thank the Advanced Diagnostics Laboratory, Department of Laboratory Medicine and Pathology, and the Center for Individualized Medicine at Mayo Clinic for funding this research. We thank Dr Nicholas Chia for his insightful discussions and feedback on this manuscript. We are grateful to the thorough and constructive comments of the anonymous reviewers whose feedback greatly improved our paper.

Conflicts of Interest

JCO has equity in Doximity, Inc, and has received small grants from Nference, Inc, and personal fees from Bates College. None of these are related to the current paper.

Multimedia Appendix 1
Error analysis.
[DOCX File , 496 KB - jmir_v23i9e30157_app1.docx]

Multimedia Appendix 2
Monte Carlo missing-not-at-random simulation.
[DOCX File , 592 KB - jmir_v23i9e30157_app2.docx]

References

1. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med 2020 Apr;26(4):450-452 [FREE Full text] [doi: 10.1038/s41591-020-0820-9] [Medline: 32284615]
2. Zhou P, Yang X, Wang X, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020 Mar;579(7798):270-273 [FREE Full text] [doi: 10.1038/s41586-020-2012-7] [Medline: 32015507]
3. WHO Coronavirus disease (COVID-19) dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2021-08-11]
4. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020 Apr 24;368(6489):395-400 [FREE Full text] [doi: 10.1126/science.aba9757] [Medline: 32144116]
5. Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? Lancet 2020 Apr 11;395(10231):1225-1228 [FREE Full text] [doi: 10.1016/S0140-6736(20)30627-9] [Medline: 32178769]
6. Rosenbaum L. Facing Covid-19 in Italy - Ethics, Logistics, and Therapeutics on the Epidemic's Front Line. N Engl J Med 2020 May 14;382(20):1873-1875. [doi: 10.1056/NEJMp2005492] [Medline: 32187459]
7. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. JAMA 2020 Aug 25;324(8):782-793. [doi: 10.1001/jama.2020.12839] [Medline: 32648899]
8. Gao Y, Cai G, Fang W, Li H, Wang S, Chen L, et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. Nat Commun 2020 Oct 06;11(1):5033 [FREE Full text] [doi: 10.1038/s41467-020-18684-2] [Medline: 33024092]
9. Gong J, Ou J, Qiu X, Jie Y, Chen Y, Yuan L, et al. A Tool for Early Prediction of Severe Coronavirus Disease 2019 (COVID-19): A Multicenter Study Using the Risk Nomogram in Wuhan and Guangdong, China. Clin Infect Dis 2020 Jul 28;71(15):833-840 [FREE Full text] [doi: 10.1093/cid/ciaa443] [Medline: 32296824]
10. Shi Y, Yu X, Zhao H, Wang H, Zhao R, Sheng J. Host susceptibility to severe COVID-19 and establishment of a host risk score: findings of 487 cases outside Wuhan. Crit Care 2020 Mar 18;24(1):108 [FREE Full text] [doi: 10.1186/s13054-020-2833-7] [Medline: 32188484]
11. Vaid A, Somani S, Russak AJ, De Freitas JK, Chaudhry FF, Paranjpe I, et al. Machine Learning to Predict Mortality and Critical Events in a Cohort of Patients With COVID-19 in New York City: Model Development and Validation. J Med Internet Res 2020 Nov 06;22(11):e24018 [FREE Full text] [doi: 10.2196/24018] [Medline: 33027032]
12. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ 2020 Apr 07;369:m1328 [FREE Full text] [doi: 10.1136/bmj.m1328] [Medline: 32265220]
13. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [doi: 10.1016/0021-9681(87)90171-8] [Medline: 3558716]
14. Kumar A, Arora A, Sharma P, Anikhindi SA, Bansal N, Singla V, et al. Is diabetes mellitus associated with mortality and severity of COVID-19? A meta-analysis. Diabetes Metab Syndr 2020;14(4):535-545 [FREE Full text] [doi: 10.1016/j.dsx.2020.04.044] [Medline: 32408118]

15. Erickson N, Mueller J, Shirkov A, Zhang H, Larroy P, Li M, et al. AutoGluon-Tabular: robust and accurate AutoML for structured data. 2020 Presented at: 7th ICML Workshop on Automated Machine Learning; July 12-18; Virtual Location URL: https://www.automl.org/wp-content/uploads/2020/07/AutoML_2020_paper_7.pdf
16. Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent Neural Networks for Multivariate Time Series with Missing Values. Sci Rep 2018 Apr 17;8(1):6085 [FREE Full text] [doi: 10.1038/s41598-018-24271-9] [Medline: 29666385]
17. Therneau T, Grambsch P. Modeling Survival Data: Extending the Cox Model. New York, USA: Springer; 2000:7-13.
18. Cro S, Morris TP, Kenward MG, Carpenter JR. Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: A practical guide. Stat Med 2020 Sep 20;39(21):2815-2842. [doi: 10.1002/sim.8569] [Medline: 32419182]
19. Buuren SV, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J Stat Soft 2011;45(3):1-67. [doi: 10.18637/jss.v045.i03]
20. Pan P, Li Y, Xiao Y, Han B, Su L, Su M, et al. Prognostic Assessment of COVID-19 in the Intensive Care Unit by Machine Learning Methods: Model Development and Validation. J Med Internet Res 2020 Nov 11;22(11):e23128 [FREE Full text] [doi: 10.2196/23128] [Medline: 33035175]
21. Ko H, Chung H, Kang WS, Park C, Kim DW, Kim SE, et al. An Artificial Intelligence Model to Predict the Mortality of COVID-19 Patients at Hospital Admission Time Using Routine Blood Samples: Development and Validation of an Ensemble Model. J Med Internet Res 2020 Dec 23;22(12):e25442 [FREE Full text] [doi: 10.2196/25442] [Medline: 33301414]
22. Abdulaal A, Patel A, Charani E, Denny S, Mughal N, Moore L. Prognostic Modeling of COVID-19 Using Artificial Intelligence in the United Kingdom: Model Development and Validation. J Med Internet Res 2020 Aug 25;22(8):e20259 [FREE Full text] [doi: 10.2196/20259] [Medline: 32735549]
23. Kim H, Han D, Kim J, Kim D, Ha B, Seog W, et al. An Easy-to-Use Machine Learning Model to Predict the Prognosis of Patients With COVID-19: Retrospective Cohort Study. J Med Internet Res 2020 Nov 09;22(11):e24225 [FREE Full text] [doi: 10.2196/24225] [Medline: 33108316]
24. Bolourani S, Brenner M, Wang P, McGinn T, Hirsch JS, Barnaby D, Northwell COVID-19 Research Consortium. A Machine Learning Prediction Model of Respiratory Failure Within 48 Hours of Patient Admission for COVID-19: Model Development and Validation. J Med Internet Res 2021 Feb 10;23(2):e24246 [FREE Full text] [doi: 10.2196/24246] [Medline: 33476281]

Abbreviations

AUROC: area under the receiver operator characteristic
EHR: electronic health record
GRU: gated recurrent unit
LIS: laboratory information system
ML: machine learning
MNAR: missing-not-at-random
PCR: polymerase chain reaction
RNN: recurrent neural network

Edited by C Basch; submitted 03.05.21; peer-reviewed by HJ Kim, S Lalmuanawma; comments to author 28.06.21; revised version received 18.07.21; accepted 11.08.21; published 28.09.21.

Please cite as:
Sankaranarayanan S, Balan J, Walsh JR, Wu Y, Minnich S, Piazza A, Osborne C, Oliver GR, Lesko J, Bates KL, Khezeli K, Block DR, DiGuardo M, Kreuter J, O'Horo JC, Kalantari J, Klee EW, Salama ME, Kipp B, Morice WG, Jenkinson G
COVID-19 Mortality Prediction From Deep Learning in a Large Multistate Electronic Health Record and Laboratory Information System Data Set: Algorithm Development and Validation
J Med Internet Res 2021;23(9):e30157
URL: https://www.jmir.org/2021/9/e30157
doi:10.2196/30157
PMID:34449401

Original Paper

Columbia Open Health Data for COVID-19 Research: Database Analysis

Junghwan Lee[1], MA; Jae Hyun Kim[1], PhD; Cong Liu[1], PhD; George Hripcsak[1], MD, MS; Karthik Natarajan[1], PhD; Casey Ta[1], PhD; Chunhua Weng[1], PhD

Columbia University, New York, NY, United States

Corresponding Author:
Chunhua Weng, PhD
Columbia University
Ph-20, 622 W 168 ST
New York, NY
United States
Phone: 1 212 304 7907
Email: cw2384@cumc.columbia.edu

Abstract

Background: COVID-19 has threatened the health of tens of millions of people all over the world. Massive research efforts have been made in response to the COVID-19 pandemic. Utilization of clinical data can accelerate these research efforts to combat the pandemic since important characteristics of the patients are often found by examining the clinical data. Publicly accessible clinical data on COVID-19, however, remain limited despite the immediate need.

Objective: To provide shareable clinical data to catalyze COVID-19 research, we present Columbia Open Health Data for COVID-19 Research (COHD-COVID), a publicly accessible database providing clinical concept prevalence, clinical concept co-occurrence, and clinical symptom prevalence for hospitalized patients with COVID-19. COHD-COVID also provides data on hospitalized patients with influenza and general hospitalized patients as comparator cohorts.

Methods: The data used in COHD-COVID were obtained from NewYork-Presbyterian/Columbia University Irving Medical Center's electronic health records database. Condition, drug, and procedure concepts were obtained from the visits of identified patients from the cohorts. Rare concepts were excluded, and the true concept counts were perturbed using Poisson randomization to protect patient privacy. Concept prevalence, concept prevalence ratio, concept co-occurrence, and symptom prevalence were calculated using the obtained concepts.

Results: Concept prevalence and concept prevalence ratio analyses showed the clinical characteristics of the COVID-19 cohorts, confirming the well-known characteristics of COVID-19 (eg, acute lower respiratory tract infection and cough). The concepts related to the well-known characteristics of COVID-19 recorded high prevalence and high prevalence ratio in the COVID-19 cohort compared to the hospitalized influenza cohort and general hospitalized cohort. Concept co-occurrence analyses showed potential associations between specific concepts. In case of acute lower respiratory tract infection in the COVID-19 cohort, a high co-occurrence ratio was obtained with COVID-19–related concepts and commonly used drugs (eg, disease due to coronavirus and acetaminophen). Symptom prevalence analysis indicated symptom-level characteristics of the cohorts and confirmed that well-known symptoms of COVID-19 (eg, fever, cough, and dyspnea) showed higher prevalence than the hospitalized influenza cohort and the general hospitalized cohort.

Conclusions: We present COHD-COVID, a publicly accessible database providing useful clinical data for hospitalized patients with COVID-19, hospitalized patients with influenza, and general hospitalized patients. We expect COHD-COVID to provide researchers and clinicians quantitative measures of COVID-19–related clinical features to better understand and combat the pandemic.

(J Med Internet Res 2021;23(9):e31122) doi:10.2196/31122

KEYWORDS

COVID-19; open data; electronic health record; data science; research; data; access; database; symptom; cohort; prevalence

Introduction

COVID-19 has threatened the health of tens of millions of people all over the world. The global pandemic caused by COVID-19 has sparked massive research efforts in the fight against the novel disease, including characterizing the disease and clinical progression, identifying risk factors for hospitalization, and finding drugs that can be repurposed to lessen disease severity [1-3]. Utilization of clinical data from different institutions, hospitals, and nations can accelerate these research efforts since important characteristics of the patients are often found by examining the shared clinical data. Although many studies sharing epidemiological data [4,5], public health [6], and social measures [7] for COVID-19 research have been conducted, publicly accessible clinical data on COVID-19 remain limited despite the immediate need [8], mainly owing to the potential risk to patient privacy that can still exist even after deidentification of the data [9].

Recognizing the need for publicly accessible electronic health record (EHR)–derived data in a broad range of clinical and translational research, we previously developed Columbia Open Health Data (COHD). COHD provides open and easy access to prevalence and co-occurrence statistics on conditions, drugs, procedures, and demographics derived from structured EHR data from NewYork-Presbyterian/Columbia University Irving Medical Center (NYP/CUIMC) [10], which serves the large and diverse population of NYC and its surrounding areas. Since its deployment, COHD has accelerated biomedical research by providing 2 informative resources, prevalence and co-occurrence statistics, and their derived association metrics [10-12].

NYC was one of the first epicenters of COVID-19 in the United States with the first confirmed case on February 29, 2020 [13]. As one of the largest academic medical centers in NYC, NYP/CUIMC has admitted more than 4000 patients as of September 1, 2020. We aim to provide shareable clinical data to catalyze future COVID-19 research by presenting Columbia Open Health Data for COVID-19 Research (COHD-COVID), a publicly accessible database providing clinical concept prevalence, the clinical concept prevalence ratio between cohorts, clinical concept co-occurrence, and clinical symptom prevalence for a cohort of hospitalized patients with COVID-19 and comparator cohorts (a cohort of hospitalized patients with influenza and a general hospitalized patient cohort) derived from NYP/CUIMC's electronic health records. In addition to providing publicly accessible data files via the Figshare data repository, we also developed the COHD-COVID web application programming interface (API) for easy access and better usability.

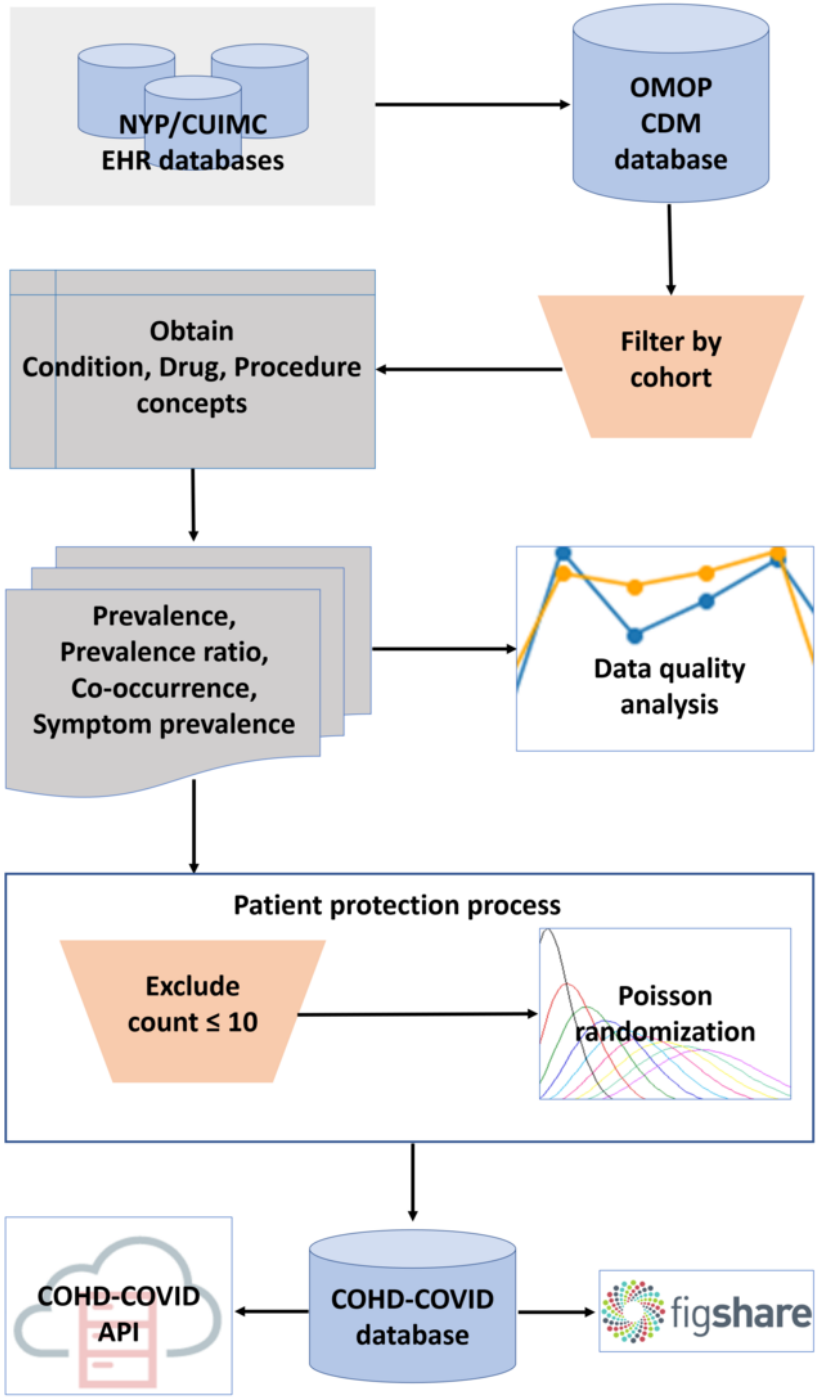
Methods

Methods Overview

We used the term "concept" to refer to clinical entities and events such as conditions (ie, diagnosis), drugs, and procedures. The concepts and their names are defined by the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). When concepts are referenced in this paper, the name of the concept is styled in italics (eg, *Disorder of respiratory system*) to distinguish the formalized concepts from regular text. We also styled entities in the OMOP CDM (eg, *person_id* column in *condition_occurrence* table) in italics.

Figure 1 depicts the overall workflow to create COHD-COVID. Columbia's clinical data warehouse was converted to the OMOP CDM. We first filtered the EHR data in accordance with the cohort definitions, and then EHR for each patient's inpatient visits were identified. Condition, drug, and procedure concepts were obtained from the identified visits. Concept prevalence, concept prevalence ratio, concept co-occurrence, and symptom prevalence analyses followed using the obtained concepts. To protect patient privacy, we excluded rare concepts observed in 10 or fewer visits and perturbed the true counts using Poisson randomization. Perturbation of the true counts and exclusion of rare concepts reduce the uniqueness of individual in the data, which can minimize the risk of reidentification [9]. Perturbed counts generated by the Poisson randomization process do not show a significant difference from the true counts [10]. The resulting data were stored in a MySQL database and made publicly available via the COHD-COVID web API [14]. All analyses were conducted using Python 3.5.2. This study received institutional review board approval with a waiver for informed consent.

Figure 1. Overall workflow of Columbia Open Health Data for COVID-19 Research (COHD-COVID). API: application programming interface; NYP/CUIMC: NewYork-Presbyterian/Columbia University Irving Medical Center; OMOP CDM: Observational Medical Outcomes Partnership Common Data Model.



Data Source

We utilized EHR data from the NYP/CUIMC's clinical data warehouse, where inpatient and outpatient data dating back to 1985 are stored. EHR data were collected during routine clinical care. Patients were notified of potential uses of their data for research at the point of care and data collection. NYP/CUIMC has converted its clinical data warehouse to OMOP CDM on a regular basis. Since NYP/CUIMC covers NYC and the surrounding area, which has a diverse population of 8.2 million people and was an early COVID-19 epicenter, the EHR data from NYP/CUIMC can provide a diverse and large sample of patients with COVID-19.

Three different patient cohorts were used in this study. The COVID-19 cohort was defined as hospitalized patients aged

≥18 years with a COVID-19–related condition diagnosis or a confirmed positive COVID-19 test result during their hospitalization period or within the prior 21 days. Patients identified with the COVID-19 cohort definition from March 1, 2020, to September 1, 2020, were included in the COVID-19 cohort. The influenza cohort was similarly defined as patients aged ≥18 years who had at least 1 occurrence of influenza conditions or precoordinated positive measurements or positive influenza test results during their hospitalization period or within the prior 21 days. The general cohort was defined as all hospitalized patients aged ≥18 years. Patient visits from calendar years 2014 to 2019 were included for the influenza and general cohorts. All cohorts were divided into subcohorts stratified by sex (male vs female) and age (adults aged 18-64 years vs seniors aged >65 years) for further investigation. The COVID-19 and influenza cohort definitions were adapted from the cohort definitions created by the Observational Health Data Science and Informatics' (OHDSI's) international network study for COVID-19 [3]. Summary statistics of the cohorts are provided in Table 1.

Patients belonging to the cohorts based on the 3 cohort definitions above were identified using the unique *person_id* from the *person* table in the OMOP CDM. Condition, drug, and procedure concepts observed in these patients during inpatient visits were extracted from the *condition_occurrence*, *drug_exposure*, and *procedure_occurrence* tables in the OMOP CDM, respectively. Inpatient visits of patients and the concepts in these visits were identified using *person_id* along with *visit_occurrence_id* from the *visit_occurrence* table. We used visit-based counts instead of patient-based counts in the following analyses for robust comparison between cohorts that had different observation windows. For example, if we use patient-based counts, the patients included in cohorts with longer observation windows are likely to be observed with more concepts than the patients in cohorts with shorter observation windows, which could inject bias into the metrics used in the analyses. Thus, we used visit-based counts to mitigate the effect of different observation window lengths to reduce the bias.

Table 1. Basic statistics of 3 cohorts in the Columbia Open Health Data for COVID-19 Research database. The statistics summarized here are based on the data from NewYork-Presbyterian/Columbia University Irving Medical Center as of September 1, 2020.

	COVID-19 cohort	Influenza cohort	General cohort
Time range	March 1, 2020, to August 31, 2020	January 1, 2014, to December 12, 2019	January 1, 2014, to December 12, 2019
Patients, n	4127	3261	175,930
Total inpatient visits, n	4846	3860	314,680
Male patients; inpatient visits, n (%)	2103 (51.0); 2518 (52.0)	1454 (44.6); 1732 (44.9)	67,662 (38.5); 128,908 (41.0)
Female patients; inpatient visits, n (%)	2024 (49.0); 2328 (48.0)	1807 (55.4); 2128 (55.1)	108,268 (61.5); 185,772 (59.0)
Adult patients (aged 18-64 years); inpatient visits, n (%)	2147 (52.0); 2511 (51.8)	1315 (40.3); 1553 (40.2)	104,020 (59.1); 173,843 (55.2)
Senior patients (>65 years); inpatient visits, n (%)	1980 (48.0); 2335 (48.2)	1946 (59.7); 2307 (59.8)	71,910 (40.9); 140,837 (44.8)

Concept Prevalence and Concept Prevalence Ratio Analysis

We calculated the concept prevalence in each cohort as detected from the EHR. The concept prevalence is defined as in equation **(1)**.

where  is the prevalence of concept *C* in cohort *H*,  is the set of unique inpatient visits of patients in cohort *H* observed with concept *C*, and T_H is the set of unique inpatient visits of patients in cohort *H*. We also calculated hierarchical concept prevalence by defining  as the set of unique visits of patients observed with concept *C* or any of concept *C*'s descendant concepts as defined in the *concept_ancestor* table in the OMOP CDM. For example, the hierarchical count for *Ibuprofen* (OMOP concept ID 1177480) not only includes entries where the specific concept *Ibuprofen* was used, but also includes entries using other descendant concepts, such as *Ibuprofen 600 MG Oral Tablet* (OMOP concept ID 19019073). Taking hierarchical relationships into account mitigates some of the issues with coding variations across time and practices as different concepts with minor semantic differences can be aggregated into higher-level concepts.

The concept prevalence ratio indicates how frequently concept *C* occurs in cohort *A* relative to cohort *B*. The natural logarithm of the concept prevalence ratio is defined as in equation **(2)**.

where $LR(C_{A,B})$ is the log ratio of the prevalence of concept *C* for cohort *A* to cohort *B*,  is prevalence of concept *C* in cohort *A*, and  is prevalence of concept *C* in cohort *B*. Hierarchical concept prevalence ratio can be calculated by using hierarchical concept prevalence  and  .

Concept Co-occurrence Analysis

Concept co-occurrence represents how frequently a specific concept pair appears in a cohort. We defined concept co-occurrence prevalence as in equation **(3)**.

where  is the co-occurrence prevalence of concepts C_1 and C_2 in cohort H,  is the set of unique visits of patients observed with concept C_1 and C_2 in cohort H, and T_H is the set of unique visits of patients in the cohort H. We also calculated hierarchical concept co-occurrence using the hierarchy of concepts as described above.

COVID-19 Symptom Prevalence Analysis

Since clinical symptoms often include multiple granular clinical concepts, a set of related concepts for a symptom is required to calculate the prevalence of the symptom. For example, dyspnea, which is one of the major symptoms of COVID-19, can be detected as standard concept *Dyspnea* or *Acute respiratory distress* in different patients. The 2 concepts do not have any hierarchical relationship and thus will not be aggregated by the hierarchical prevalence analyses. Thus, a concept set containing both *Dyspnea* and *Acute respiratory distress* is needed for accurate calculation of the prevalence of dyspnea. We defined symptom prevalence as in equation **(4)**.

where  is prevalence of symptom S in cohort H,  is the set of unique visits of patients observed with concept C_i, T_H is the set of unique visits of patients observed in the cohort H, s is the number of the unique concepts in the concept set for symptom S, and $\cup$ is the union operator. Hierarchy between concepts is not considered in the symptom prevalence analysis since a concept set of a symptom already reflects hierarchy for that symptom.

We used concept sets for 11 major symptoms of COVID-19 (*Cough*, *Chills*, *Abdominal pain*, *Diarrhea*, *Dyspnea*, *Fatigue*, *Fever*, *Myalgia*, *Nausea and vomiting*, *Tachypnea*, and *Throat pain*), which have been defined by OHDSI to calculate symptom prevalence. The concepts included in each symptom are available in Figshare [15].

Data Quality Analysis

Assessing the quality of EHR data is critical for its effects on secondary analysis for research in the health care and medical domains [16]. We calculated the sum of nonrandomized counts of all concepts on a monthly basis for the COVID-19 cohort and on a yearly basis for the general and influenza cohorts. For each of the condition, drug, and procedure domains, we examined the total counts of concepts with the number of visits to detect any issues regarding data quality and temporal plausibility of the EHR data we used in the study [17]. We also calculated the annual mean (SD) values of concept prevalence and concept co-occurrence for the general and influenza cohorts to assess the temporal variance of each concept and concept co-occurrence pair.

Results

Results Overview

In this section, we show a sample of the results of analyses using the data from COHD-COVID. Since COHD-COVID contains massive amounts of data covering several thousand concepts, it is worth noting that only a small sample of the results is shown in this section. Users can obtain the results of interest in addition to the results shown in this section by using the COHD-COVID API [14] or by downloading the flat data files from Figshare [15]. The results of all analyses, concept definitions, and concepts included in each symptom are available as tab-delimited flat files in Figshare, except the concept prevalence ratio analyses, since they can be directly computed from concept prevalence data. COHD-COVID API [14] provides all results.

Concept Prevalence Analysis

Figure 2 shows the prevalence of 10 condition (Figure 2A) and drug concepts (Figure 2B) in the COVID-19 cohort, influenza cohort, and general cohort. We chose the 10 most prevalent condition and drug concepts in the COVID-19 cohort for this use case. For drug concepts, we chose the most prevalent drug ingredient concepts using hierarchical analysis to count multiple drugs that have different brand names, dosages, and formulations but are based on the same ingredients together. The condition concepts were chosen without hierarchical analysis to identify the top 10 specific conditions. *Fever* showed the highest prevalence (0.2619) among all condition concepts for the COVID-19 cohort, followed by *Cough* (0.2491) and *Dyspnea* (0.2594). *Acetaminophen* was the most prevalent (0.7912) drug ingredient used for patients with COVID-19, followed by *Enoxaparin* (0.5803) and *Glucose* (0.4424). Figures 3 and 4 show the prevalence of 10 condition and drug concepts, respectively, in the COVID-19 cohort stratified by age (Figures 3A and 4A) and gender (Figure 4A and 4B). The 10 condition and drug concepts were the top 10 most prevalent concepts in the full COVID-19 cohort without stratification.

Figure 2. (A) Condition and (B) drug concept prevalence in the COVID-19 cohort, influenza cohort, and general cohort.

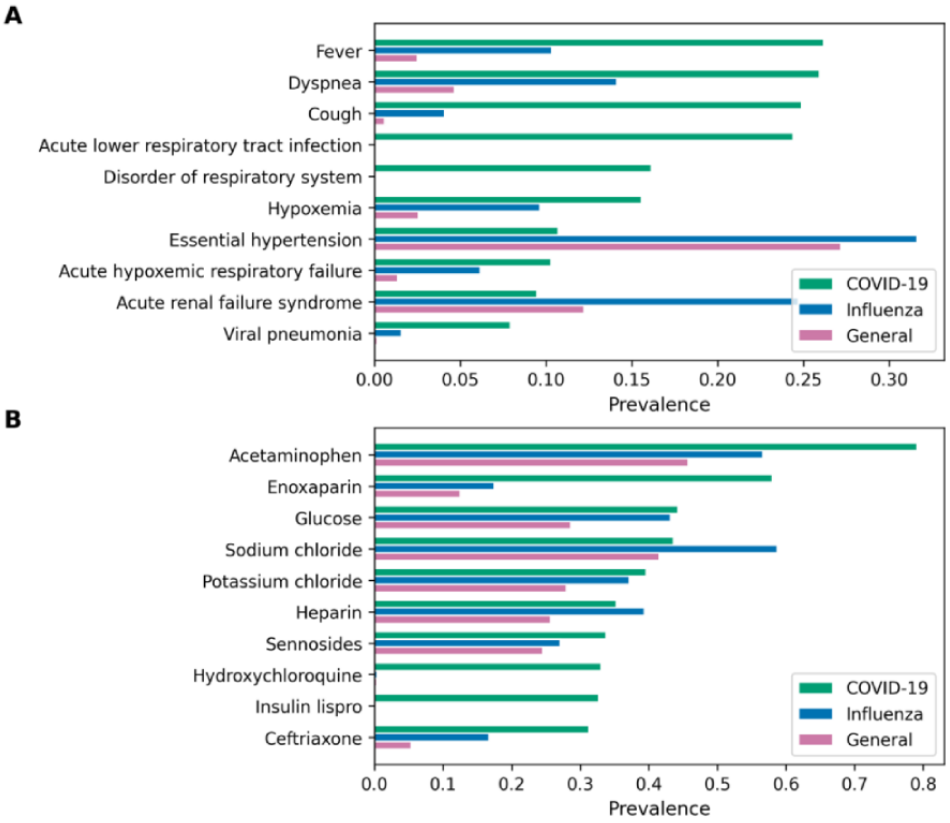


Figure 3. Condition concept prevalence in (A) age and (B) sex sub-cohorts of the COVID-19 cohort. The full COVID-19 cohort indicates original COVID-19 cohort without stratification.

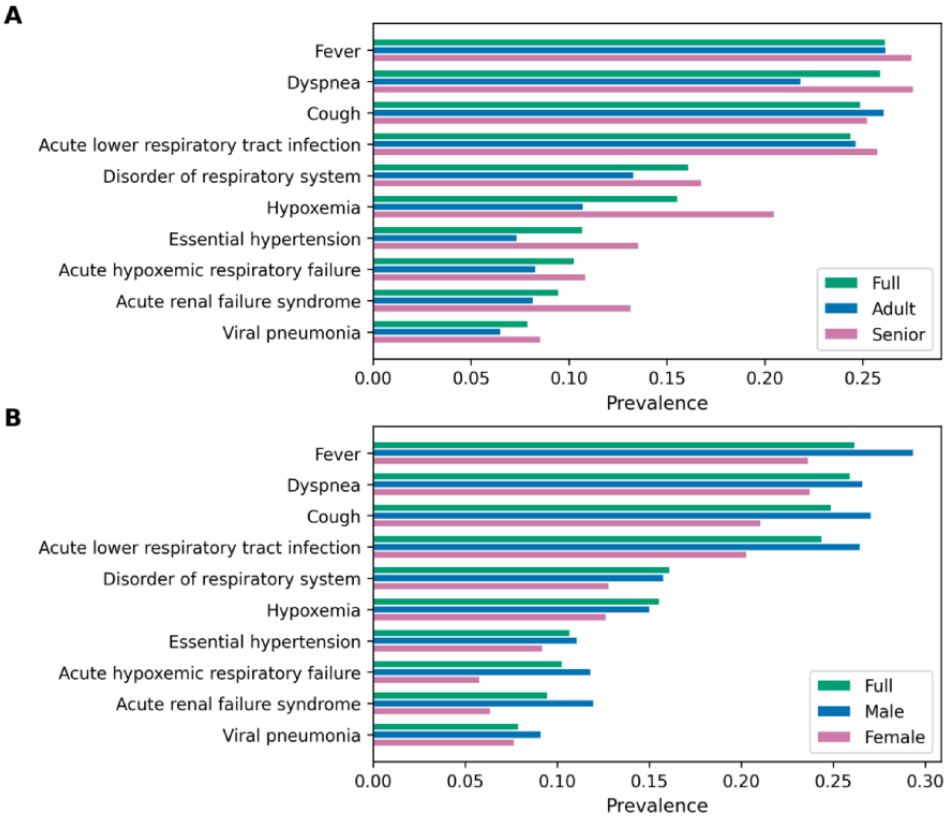
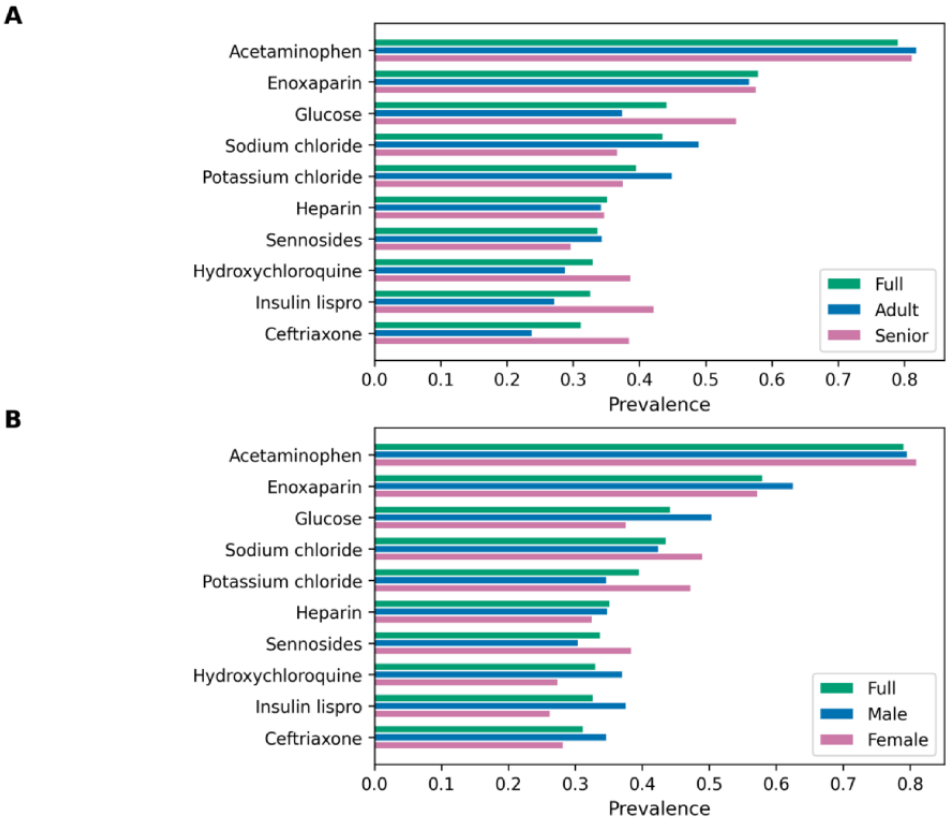


Figure 4. Drug concept prevalence in (A) age and (B) sex subcohorts of the COVID-19 cohort. The full COVID-19 cohort indicates original COVID-19 cohort without stratification.



Concept Prevalence Ratio Analysis

Table 2 shows the top 10 condition concepts that showed the highest concept prevalence ratio for the COVID-19 cohort relative to the comparator cohorts (general cohort and influenza cohort). *Disease due to Coronavirus* and *Acute lower respiratory tract infection* showed the highest concept prevalence ratio for the COVID-19 cohort relative to the influenza cohort and the general cohort, respectively.

Table 2. Top 10 condition concepts that show the highest concept prevalence ratio for the COVID-19 cohort relative to the general and influenza cohorts.

Influenza cohort (prevalence ratio)	General cohort (prevalence ratio)
Disease due to Coronavirus (3.28)	Acute lower respiratory tract infection (7.30)
Disorientated (2.44)	Disorder of respiratory system (5.56)
Blood chemistry abnormal (2.25)	General finding of observation of patient (5.17)
Acute respiratory distress syndrome (2.16)	Outcome of delivery – finding (5.13)
Cerebral infarction (1.84)	Chest pain on breathing (5.09)
Cough (1.81)	Disease due to Coronavirus (4.95)
Viral pneumonia (1.61)	Patient status finding (4.60)
Acute respiratory distress (1.52)	Unplanned pregnancy (4.23)
Heart disease (1.32)	Acute respiratory distress syndrome (4.17)
Delivery normal (1.28)	Deliveries by cesarean (3.98)

Concept Co-occurrence Analysis

Table 3 shows the top 10 most frequently co-occurring concepts with *Acute lower respiratory tract infection* in the COVID-19 cohort. *Acute lower respiratory tract infection* was chosen among the 10 most prevalent condition concepts in the COVID-19 cohort. We used nonhierarchical concept co-occurrence to obtain the result.

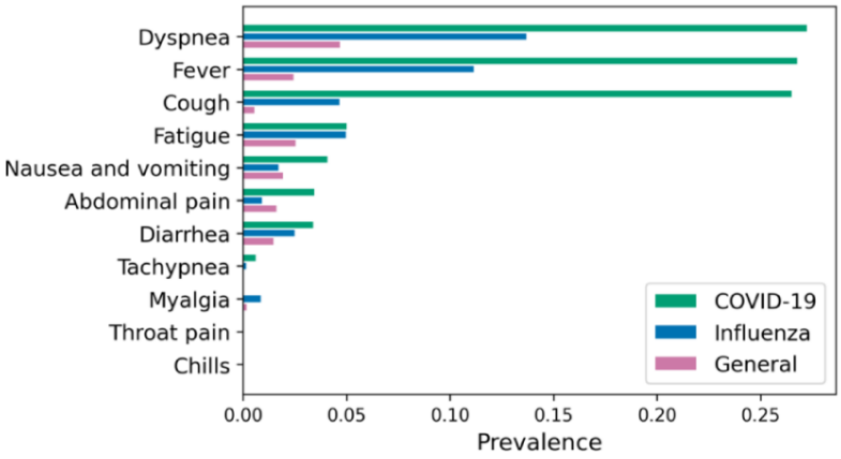
Table 3. The 10 concepts that most frequently co-occurred with Acute lower respiratory tract infection in the full COVID-19 cohort.

Concept name	Co-occurrence prevalence
Disease due to Coronavirus	0.2329
Radiologic examination, chest; single view	0.2214
Infectious agent detection by nucleic acid (DNA or RNA); severe acute respiratory syndrome coronavirus 2	0.2205
Acetaminophen 325 MG Oral Tablet	0.2037
0.4 ML Enoxaparin sodium 100 MG/ML Prefilled Syringe	0.1504
Hydroxychloroquine Sulfate 200 MG Oral Tablet	0.1246
50 ML Glucose 500 MG/ML Prefilled Syringe	0.1085
Ceftriaxone 1000 MG Injection	0.1073
Blood typing, serologic; ABO	0.1044
Glucose 0.4 MG/MG Oral Gel	0.0970

COVID-19 Symptom Prevalence Analysis

Figure 5 shows the prevalence of the 11 major COVID-19 symptoms for all 3 cohorts. In the COVID-19 cohort, *Dyspnea* showed the highest prevalence among the 11 symptoms followed by *Fever* and *Cough*.

Figure 5. Symptom prevalence of 11 major symptoms in patients with COVID-19 for all 3 cohorts.



Data Quality Analysis

Figure 6 shows the total counts across all concepts in the condition domain (Figure 6A), drug domain (Figure 6B), procedure domain (Figure 6C), and the total visits per month between March 2020 and August 2020 (Figure 6D) for the COVID-19 cohort. The total counts of conditions, drugs, procedures, and total visits all show steep increases in March and April 2020, when the number of COVID-19 cases surged in NYC (Figure 6E) [18]. The total counts of all domains and the total visits decreased starting May 2020 as the number of patients with COVID-19 in NYC decreased.

Figure 7 shows the total counts per year across all concepts in the condition domain (Figure 7A), drug domain (Figure 7B), procedure domain (Figure 7C), and the total visits per year between 2014 and 2019 (Figure 7D) for the general cohort and influenza cohort. The total counts of conditions, drugs, procedures, and total visits per year for the 2 cohorts show consistent trends.

The annual mean (SD) value of concept prevalence and concept co-occurrence for the general and influenza cohorts are available in Figshare [15] to assess the temporal variance of each concept and concept co-occurrence pair. The mean and standard deviation of annual concept prevalence and co-occurrence should only be compared to each other to assess the stability of the concept over the given time period of the data set.

A change in the EHR system can affect the quality and characteristics of EHR data collected for secondary research. NYP/CUIMC changed its EHR system as of February 1, 2020, from Allscripts to Epic, which might affect the COVID-19 cohort data and its characteristics as opposed to the influenza and general cohorts, which were collected prior to the EHR change. To detect the impact of the change, we performed a *t* test for all concepts reported in the COVID-19 cohort between the counts from a pre-Epic period (January 1, 2020, to January 31, 2020) and from a post-Epic period (February 1, 2020, to February 29, 2020). The post-Epic period was chosen to minimize the inclusion of dates when COVID-19 would have impacted clinical practices in NYC. The counts of concepts were recorded on a daily basis. We considered that there would be no difference between the 2 periods if the counts for the concept from the 2 periods are the same (ie, all 0 counts for the

2 periods). Of all 1066 unique concepts reported in the COVID-19 cohort, 119 (11.2%) concepts showed a significant difference ($P \leq .01$) between the pre- and post-Epic periods. The *P* values of the *t* tests for all individual concepts are available in Figshare [15] to allow users to factor in these data quality considerations for each concept.

Figure 6. Total counts across all concepts in the (A) condition domain, (B) drug domain, (C) procedure domain, and (D) the total visits per month between March 2020 and August 2020 for the COVID-19 cohort. (E) Total COVID-19–positive cases per day in New York City from March 2020 to August 2020.

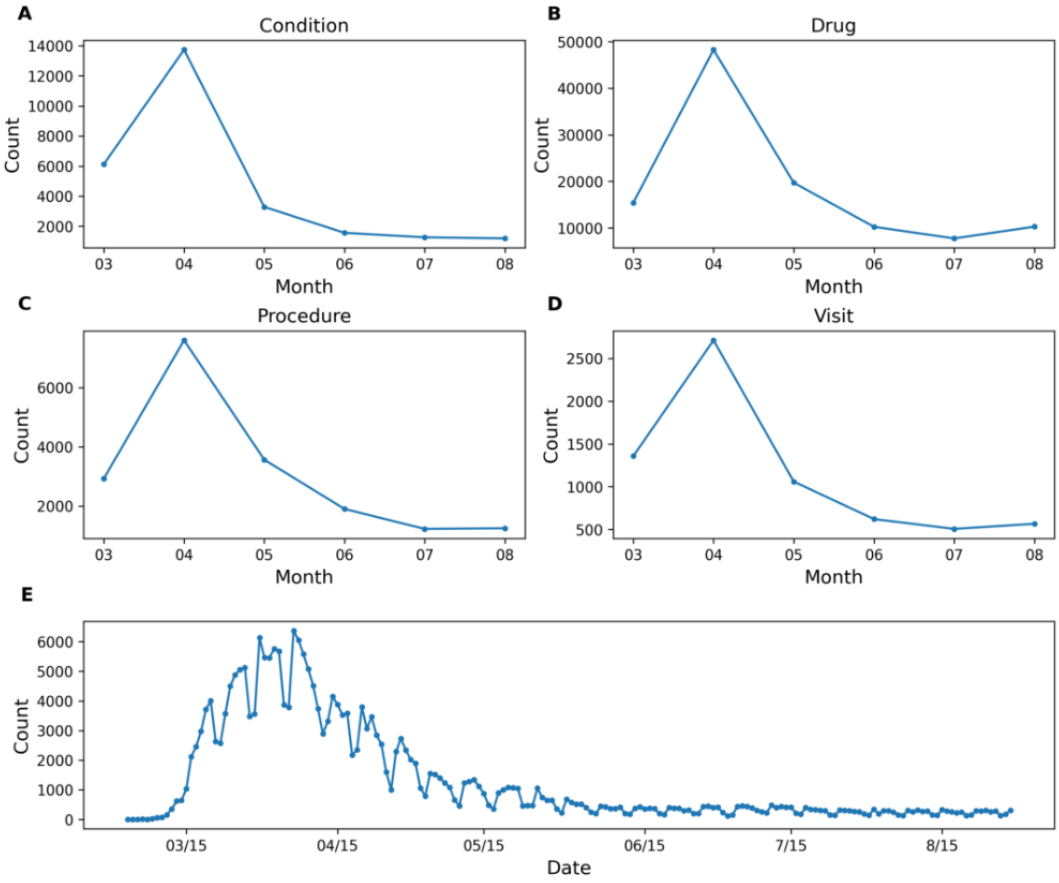
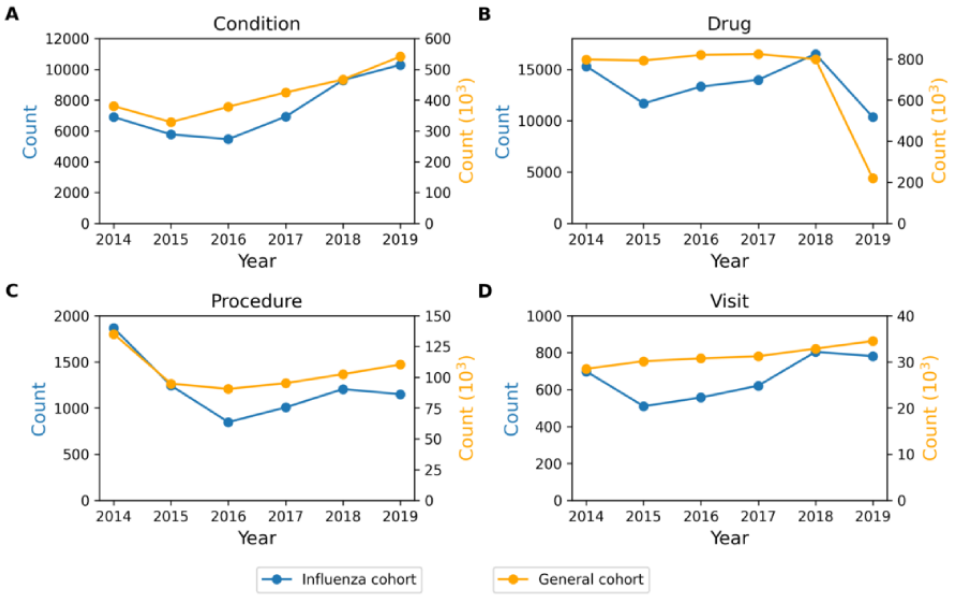


Figure 7. Total counts across all concepts in the (A) condition domain, (B) drug domain, (C) procedure domain, and (D) the total visits per year between 2014 and 2019 for the general and influenza cohort.



Discussion

Principal Findings

In this study, we present a publicly accessible database providing clinical concept prevalence, clinical concept co-occurrence, and clinical symptom prevalence for hospitalized patients with COVID-19, hospitalized patients with influenza, and general hospitalized patients. We showed the utility of the data for future data-driven studies by exploring the results of the concept prevalence analysis, concept prevalence ratio analysis, concept co-occurrence analysis, and symptom prevalence analysis. Results from the analyses aligned with published findings and can be used to find novel hypotheses as described in the following discussion. The results of all analyses were provided as flat files in Figshare [15] and also easily accessible through web API [14].

Concept prevalence analysis on the 3 cohorts can be used to determine the clinical characteristics of the COVID-19 cohort. A comparison of the prevalence of these concepts between the COVID-19 cohort and the influenza or general cohorts provides contextual evidence regarding whether the concept is associated with COVID-19 or if it is common among hospitalized patients. Figure 2A shows that *Fever*, *Cough*, *Dyspnea*, and *Acute lower respiratory syndrome* were more highly prevalent in the COVID-19 cohort than in the 2 comparator cohorts. The concepts highly prevalent in the COVID-19 cohort were well-known symptoms of COVID-19, concurrent with existing studies. As shown in Figure 2B, *Hydroxychloriquine*, which was one of the drugs widely administered to patients with COVID-19 [19] during the early months of the pandemic, showed high prevalence in the COVID-19 cohort.

Concept prevalence analysis on the COVID-19 cohort and its subcohorts can be used to determine the clinical characteristics of the subcohorts and determine if these characters differ between the sexes or between adult and older patients. Figure 3A shows that the senior cohort had higher prevalences in all 10 condition concepts than the full and adult cohorts and also shows that *Essential hypertension* and *Hypoxemia* were particularly more prevalent in senior patients with COVID-19, which indicates that senior patients are more likely to have these apparent symptoms of COVID-19 than adult patients. Figure 4A shows that drug ingredients related to type 2 diabetes (eg, *Glucose* and *Insulin Lispro*) showed higher prevalence in senior patients than in adult patients, which indicates that type 2 diabetes is a more common comorbidity among senior patients with COVID-19.

Concept prevalence ratio analysis can be used to unveil how often specific concepts appeared in the COVID-19 cohort compared to the comparator cohorts. Table 2 shows that common and general symptoms of COVID-19 (eg, *Chest pain on breathing*, *Cough*, and *Disease due to coronavirus*) showed a high prevalence ratio in the COVID-19 cohort relative to the general cohort. In contrast, more severe symptoms of COVID-19 (eg, *Disoriented*, *Acute respiratory distress syndrome*, *Viral pneumonia*, and *Blood chemistry abnormal*) showed a higher prevalence ratio in the COVID-19 cohort than in the influenza cohort. Since both cohorts include hospitalized patients, these results may indicate that the aforementioned conditions are more strongly associated with the SARS-CoV-2 infection than with influenza. We also observed high prevalence ratios in concepts related to delivery in the COVID-19 cohort relative to the general cohort. Regrading *Outcome of delivery – finding*; for example, we observed a visit prevalence rate of 1.18% in the COVID-19 cohort, but only 0.0070% in the general cohort, yielding a prevalence ratio of 5.13. During normal times, deliveries only account for a small fraction of all inpatient visits. However, during the pandemic, there was a dramatic decrease in elective surgeries and hospitalizations besides those related to COVID-19 since hospital capacity was diverted their focus toward patients with COVID-19. In contrast, patients going into labor cannot be rescheduled and were regularly tested for SARS-CoV-2, which increased their representation within the COVID-19 cohort.

Potential associations between specific concepts of interest can be found through co-occurrence analysis. For instance, Table 3 confirms that the concepts associated with SARS-CoV-2 tests (ie, *Disease due to Coronavirus*, *Radiologic examination, chest; single view*, and *Infectious agent detection by nucleic acid [DNA or RNA]; severe acute respiratory syndrome coronavirus 2*) showed high co-occurrence prevalence with *Acute lower respiratory tract infection*, which shows natural strong associations between testing concepts for COVID-19 and one of the most prevalent concepts in the COVID-19 cohort.

Symptom prevalence analysis can be used to examine symptom-level characteristics of the cohorts. The COVID-19 cohort showed higher prevalence in dyspnea, fever, and cough symptoms than the other 2 comparator cohorts, which is concurrent with the known characteristics and symptoms of COVID-19 [20]. In contrast, a few of the known COVID-19 symptoms (eg, myalgia, throat pain, and chills) did now show high prevalence in the COVID-19 cohort, which might indicate that those symptoms are not clinically distinctively prevalent in patients with COVID-19.

Most of the results from the analyses align with those of existing studies, thus empirically validating the utility of COHD-COVID. COHD-COVID also can be used to find novel hypotheses related COVID-19. COHD-COVID can be used as cross-institutional data to validate or support other COVID-19 studies. For instance, the high prevalence ratio of *Cerebral infarction* in the COVID-19 cohort compared to that in the influenza cohort (Table 2) corroborates with reports from a few studies that SARS-CoV-2 might be more likely to cause thrombotic vascular events, including stroke, than other coronavirus and seasonal infectious diseases [21,22].

While we admit that the aggregated concept-level analyses may not be suitable to answer some clinical research questions, COHD-COVID will be useful for hypothesis generation and for validating emerging newly published findings on COVID-19 using real-world data. COHD, the precedent study of COHD-COVID, has been integrated into the National Center for Advancing Translational Sciences Biomedical Data Translator program, where the EHR-based data associations from COHD are linked to other sources of knowledge via knowledge graphs, allowing automated algorithms to perform

reasoning on these knowledge graphs to answer biomedical questions and suggest novel hypotheses [23,24].

Limitations

There are several limitations to this study, which should be noted. One of the limitations is that analyses performed in this study can be affected by the factors included in the data acquisition process (eg, change in the EHR system, human biases, and errors during entry). For example, coding trends and patterns (ie, the trend and pattern of frequently used concepts) can be changed through a shift in system. Another limitation is that multiple visits from the same patient can be used to calculate the metrics in the analyses since we used visit-based counts instead of patient-based counts for the analyses. This could affect the results of the specific concepts appearing in patients who are hospitalized more frequently. We also admit that some of results from the analyses cannot be validated by existing or up-to-date findings, considering rapidly growing and changing knowledge related to COVID-19. Thus, it is desirable to conduct a literature search before utilizing the data and results. We will update the data and results on a regular basis to further alleviate this limitation.

The EHR data used in this study were obtained from a single site: NYP/CUIMC. Even though NYP/CUIMC is a large academic medical center whose services cover the city and its surrounding areas, we admit that performing the analyses on the basis of the EHR data across multiple institutions and nations will be beneficial since multiple sites can diversify the population, improve accuracy, increase power and sensitivity to rare conditions, validate results by comparing across sites, and reduce variance that might exist in specific locations. Since the OMOP CDM provides the fundamentals to perform the same analyses on clinical data across different sites, we hope to collaborate with future studies sharing clinical characteristics of patients with COVID-19 and to generate a larger, richer, and more robust database that can be leveraged in translational research on COVID-19.

Conclusions

In this study, we present COHD-COVID, a publicly accessible database providing useful clinical data for hospitalized patients with COVID-19, hospitalized patients with influenza, and general hospitalized patients. The analyses using the data from COHD-COVID confirmed the well-known clinical characteristics of patients with COVID-19 and can also be used to find novel hypotheses related to COVID-19. We expect COHD-COVID will provide researchers and clinicians quantitative measures of COVID-19–related clinical features to better understand and combat the pandemic.

Acknowledgments

This study was supported by National Library of Medicine grant R01LM012895-03S1 and National Center for Advancing Translational Science grant 1OT2TR003434-01.

Conflicts of Interest

None declared.

References

1. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar 28;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]
2. Shah B, Modi P, Sagar SR. In silico studies on therapeutic agents for COVID-19: Drug repurposing approach. Life Sci 2020 Jul 01;252:117652 [FREE Full text] [doi: 10.1016/j.lfs.2020.117652] [Medline: 32278693]
3. Burn E, You SC, Sena AG, Kostka K, Abedtash H, Abrahão MTF, et al. Deep phenotyping of 34,128 adult patients hospitalised with COVID-19 in an international network study. Nat Commun 2020 Oct 06;11(1):5009 [FREE Full text] [doi: 10.1038/s41467-020-18849-z] [Medline: 33024121]
4. Xu B, Gutierrez B, Mekaru S, Sewalk K, Goodwin L, Loskill A, et al. Epidemiological data from the COVID-19 outbreak, real-time case information. Sci Data 2020 Mar 24;7(1):106 [FREE Full text] [doi: 10.1038/s41597-020-0448-0] [Medline: 32210236]
5. Hasell J, Mathieu E, Beltekian D, Macdonald B, Giattino C, Ortiz-Ospina E, et al. A cross-country database of COVID-19 testing. Sci Data 2020 Oct 08;7(1):345 [FREE Full text] [doi: 10.1038/s41597-020-00688-8] [Medline: 33033256]
6. Zheng Q, Jones FK, Leavitt SV, Ung L, Labrique AB, Peters DH, HIT-COVID Collaboration. HIT-COVID, a global database tracking public health interventions to COVID-19. Sci Data 2020 Aug 27;7(1):286 [FREE Full text] [doi: 10.1038/s41597-020-00610-2] [Medline: 32855428]
7. Desvars-Larrive A, Dervic E, Haug N, Niederkrotenthaler T, Chen J, Di Natale A, et al. A structured open dataset of government interventions in response to COVID-19. Sci Data 2020 Aug 27;7(1):285 [FREE Full text] [doi: 10.1038/s41597-020-00609-9] [Medline: 32855430]
8. Argenziano MG, Bruce SL, Slater CL, Tiao JR, Baldwin MR, Barr RG, et al. Characterization and clinical course of 1000 patients with coronavirus disease 2019 in New York: retrospective case series. BMJ 2020 May 29;369:m1996 [FREE Full text] [doi: 10.1136/bmj.m1996] [Medline: 32471884]

9. Benitez K, Malin B. Evaluating re-identification risks with respect to the HIPAA privacy rule. J Am Med Inform Assoc 2010;17(2):169-177 [FREE Full text] [doi: 10.1136/jamia.2009.000026] [Medline: 20190059]
10. Ta CN, Dumontier M, Hripcsak G, Tatonetti NP, Weng C. Columbia Open Health Data, clinical concept prevalence and co-occurrence from electronic health records. Sci Data 2018 Nov 27;5:180273 [FREE Full text] [doi: 10.1038/sdata.2018.273] [Medline: 30480666]
11. Ahalt SC, Chute CG, Fecho K, Glusman G, Hadlock J, Taylor CO, Biomedical Data Translator Consortium. Clinical Data: Sources and Types, Regulatory Constraints, Applications. Clin Transl Sci 2019 Jul;12(4):329-333 [FREE Full text] [doi: 10.1111/cts.12638] [Medline: 31074176]
12. Fecho K, Ahalt SC, Arunachalam S, Champion J, Chute CG, Davis S, Biomedical Data Translator Consortium. Sex, obesity, diabetes, and exposure to particulate matter among patients with severe asthma: Scientific insights from a comparative analysis of open clinical data sources during a five-day hackathon. J Biomed Inform 2019 Dec;100:103325 [FREE Full text] [doi: 10.1016/j.jbi.2019.103325] [Medline: 31676459]
13. Dyal NP. First Case of COVID-19 in NYC, First Death Reported in Washington State. Infectious Disease Advisor. 2020 Mar 03. URL: https://www.infectiousdiseaseadvisor.com/home/topics/respiratory/first-case-of-covid-19-in-nyc-first-death-reported-in-washington-state/ [accessed 2020-11-01]
14. Columbia Open Health Data - COVID-19. URL: https://covid.cohd.io/ [accessed 2021-09-20]
15. Lee J, Kim J, Liu C, Hripcsak G, Ta C, Weng C. COHD-COVID: Columbia Open Health Data for COVID-19 Research. Figshare. URL: https://doi.org/10.6084/m9.figshare.14739426.v1 [accessed 2021-09-20]
16. Ta CN, Weng C. Detecting Systemic Data Quality Issues in Electronic Health Records. Stud Health Technol Inform 2019 Aug 21;264:383-387 [FREE Full text] [doi: 10.3233/SHTI190248] [Medline: 31437950]
17. Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Wash DC) 2016;4(1):1244 [FREE Full text] [doi: 10.13063/2327-9214.1244] [Medline: 27713905]
18. COVID-19 Daily Counts of Cases, Hospitalizations, and Deaths. NYC Open Data. URL: https://data.cityofnewyork.us/resource/rc75-m7u3 [accessed 2021-09-20]
19. Geleris J, Sun Y, Platt J, Zucker J, Baldwin M, Hripcsak G, et al. Observational Study of Hydroxychloroquine in Hospitalized Patients with Covid-19. N Engl J Med 2020 Jun 18;382(25):2411-2418 [FREE Full text] [doi: 10.1056/NEJMoa2012410] [Medline: 32379955]
20. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
21. Fifi JT, Mocco J. COVID-19 related stroke in young individuals. Lancet Neurol 2020 Sep;19(9):713-715 [FREE Full text] [doi: 10.1016/S1474-4422(20)30272-6] [Medline: 32822622]
22. Oxley TJ, Mocco J, Majidi S, Kellner CP, Shoirah H, Singh IP, et al. Large-Vessel Stroke as a Presenting Feature of Covid-19 in the Young. N Engl J Med 2020 May 14;382(20):e60. [doi: 10.1056/NEJMc2009787] [Medline: 32343504]
23. Biomedical Data Translator Consortium. Toward A Universal Biomedical Data Translator. Clin Transl Sci 2019 Mar;12(2):86-90 [FREE Full text] [doi: 10.1111/cts.12591] [Medline: 30412337]
24. Biomedical Data Translator Consortium. The Biomedical Data Translator Program: Conception, Culture, and Community. Clin Transl Sci 2019 Mar;12(2):91-94 [FREE Full text] [doi: 10.1111/cts.12592] [Medline: 30412340]

Abbreviations

API: application programming interface
COHD: Columbia Open Health Data
COHD-COVID: Columbia Open Health Data for COVID-19 Research
EHR: Electronic health records
NYP/CUIMC: NewYork-Presbyterian/Columbia University Irving Medical Center
OHDSI: Observational Health Data Science and Informatics
OMOP CDM: Observational Medical Outcomes Partnership Common Data Model

Edited by C Basch; submitted 11.06.21; peer-reviewed by H Estiri, L Kruesi, K Fecho; comments to author 21.07.21; revised version received 03.08.21; accepted 03.08.21; published 30.09.21.

Please cite as:
Lee J, Kim JH, Liu C, Hripcsak G, Natarajan K, Ta C, Weng C
Columbia Open Health Data for COVID-19 Research: Database Analysis
J Med Internet Res 2021;23(9):e31122
URL: https://www.jmir.org/2021/9/e31122
doi:10.2196/31122
PMID:34543225

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org